This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Technical Correspondence

Mining Skier Transportation Patterns From Ski Resort Lift Usage Data

Boris Delibašić, Petar Marković, Pavlos Delias, and Zoran Obradović

***Abstract*—Descriptive data analysis is used for mining spatial and temporal patterns from ski lift entrance data. The data, collected through radio-frequency identification scanners, cover one skiing season with approximately 1.2 million recorded ski lift transportations. Cluster analysis was performed on ten subsamples, and the obtained clusters were cross-validated. Several types of skier behavior were found. Temporal clustering revealed that skier patterns differ according to time of maximal performance and length of stay in the ski lift transportation system. Spatial clustering revealed that it is reasonable to have as many clusters as ski lifts in a ski resort, since skiers tend to choose a dominant ski lift during a skier-day. The detected patterns reveal valuable information that can be used for potential improvement of products and services offered by ski resorts.**

***Index Terms*—Kopaonik ski resort, pattern clustering, radio-frequency identification (RFID), ski lift transportation.**

I. Introduction

Ski resorts constantly tend to conform to skier needs, achieve cost-effectiveness, and generate stable revenues. The goal of this paper is to reveal skier transportation patterns that can help ski resorts improve the services and products they offer to skiers and plan their operations more efficiently.

Skiing is a multibillion dollar business, with 7.1 billion USD spent worldwide and an average of 57.1 million visits in U.S. ski resorts since 2002/2003 (https://www.nsaa.org/) and an average of 400 million visits to ski resorts worldwide [18]. The behavior of skiers has changed rapidly during the past two decades. Skiers now have access to immediate information about ski resort weather conditions (e.g., http://www.snow-forecast.com/), ski ticket prices (e.g., http://www.liftopia.com/), accommodation and transportation prices. Skiers are able to plan their ski resort visits in a few hours, reach a ski resort the same day, and stay as long as the price/performance ratio suits them. Due to both unstable weather conditions in the past decade and better informed skiers, the management at ski resorts must adapt to the new circumstances by planning resort capacities and operations more efficiently.

Ski resort management uses data analysis quite commonly, as it can help them to better plan operations and budgets. In [10], a model is proposed for a number of skier-days predictions. Ski lift ticket pricing was also studied (see, e.g., [6] and [20]), as it directly influences ski resort profits. A lot of effort is being invested in enhancing the ski experience in ski resorts [14]. On August 12 2015, EpicMix (http://www.epicmix.com/) announced a new service for crowd-sourced lift line wait time optimization introduced for the season 2015/2016 at the Vail Res, which provides skiers with real-time information on congestion on ski lift entrances.

Manuscript received August 6, 2015; revised February 18, 2016 and June 8, 2016; accepted November 12, 2016. This work was supported in part by a U.S. Department of State CIES Fulbright Visiting Program grant, conducted at the Center for Data Analysis and Biomedical Informatics, Temple University. This paper was recommended by Associate Editor L. Lin.

B. Delibašić and P. Marković are with the Faculty of Organizational Sciences, University of Belgrade, Belgrade 11000, Serbia (e-mail: boris.delibasic@fon.bg.ac.rs; peleplay@gmail.com).

P. Delias is with the Eastern Macedonia and Thrace Institute of Technology, Kavala 654 04, Greece (e-mail: pdelias@teikav.edu.gr).

Z. Obradović is with the Center for Data Analytics and Biomedical Informatics, Temple University, Philadelphia, PA 19122 USA (e-mail: zoran.obradovic@temple.edu).

Digital Object Identifier 10.1109/THMS.2016.2633438

Most ski resorts are equipped with radio-frequency identification (RFID) technology that keeps track of all ski lift gate entrances. These massive data are mostly used for basic reporting by ski resort management (e.g., number of ski lift transportations per lift), as well as owners of ski passes (e.g., number of visited ski lifts, total vertical meters distance achieved, etc.). The hypothesis explored in this study is that using data mining on large ski lift transportation data can help ski resorts detect skier transportation patterns and exploit the massive data potential.

The analysis of ski lift transportation is an emerging field, as there are currently only few papers that analyze RFID ski lift transportation data (e.g., [5]). On the other hand, there are many papers that examine transportation patterns from RFID-collected transportation data, e.g., bicycle rentals in Barcelona [7], the London underground system [11]; therefore, analysis methods for this kind of data are well known.

II. Methods

A. Participants

The term skier will be used for all the persons using ski lifts, i.e., skiers, snowboarders, etc. Our study considers 21 121 skiers observed at Kopaonik ski resort in Serbia with six-day (i.e., weekly) ski tickets during the period 2009/01/01–2009/03/31. The above time period was chosen since it represents the season peak, with 89.27% of the overall season ski lift transportations and all ski lifts were operating during this period of the season. The ski lift transportation dataset contains 1 248 755 ski lift transportation records. The dataset was previously filtered to remove infeasible ski lift transitions, records with travel time shorter than officially declared, and records where exit time is earlier than entry time. After removing such records (a total of 27 635 records, or 0.49% of the original dataset), the final dataset was derived.

B. Materials

Kopaonik is the largest ski resort in Serbia. Its peak rises to 2017 m. On average, there are 160 days with natural snow coverage and 200 sunny days.

The skiing season lasts from December 1st until early May, depending on weather conditions. During the season 2009, a total of 17 ski lifts were operating on the ski resort. All ski lift gates were equipped with RFID scanners collecting data about ski lift transportations. The 17 ski lifts provide access to 32 ski slopes. There are 15 blue-easy, ten red-medium, and seven black- advanced slopes. Basic characteristics of each ski lift are given in Table I. For each ski lift, we present the lift abbreviation that will be used further in the text, difficulty of reachable slopes (blue-easy, red-medium, black-advanced), type of ski lift, peak of the ski lift exit, the declared hourly ski lift capacity (the share of total ski lift transportations is shown in brackets), transportation time of lifts in minutes, the observed percentage of ski lift transportations,

2168-2291 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
SKI LIFTS AND THEIR BASIC CHARACTERISTICS

Ski-lift	Code	Type	Peak	Difficulty of slopes	Hourly ski lift transp. capacity (skier/hour)	Transp. time (min)	% of ski lift transp.	% of start
Suncana dolina	SUN	2-seat chairlift	Krst	Blue	800 (3.62%)	4.5	1.82%	0.47%
Malo jezero	JEZ	Rope-tow	Krst	Blue	880 (3.98%)	2.5	4.14%	8.27%
Masinac	MAS	Rope-tow	Masinac	Blue	450 (2.04%)	2	4.58%	10.74%
Centar	CEN	2-seat chairlift	Suvo rudiste	Blue-Red	800 (3.62%)	10	0.9%	6.73%
Suvo rudiste	SUV	T-bar	Suvo rudiste	Blue-Red	1200 (5.43%)	3	1.18%	0%
Pancicev vrh	PAN	4-seat chairlift	Suvo rudiste	Red	2400 (10.86%)	5	17.18%	16.61%
Duboka 1	D1	4-seat chairlift	Suvo rudiste	Red-Black	1800 (8.15%)	6.5	7.2%	0%
Krcmar	KRC	2-seat chairlift	Pancicev vrh	Red	800 (3.62%)	13	0.88%	0%
Duboka 2	D2	4-seat chairlift	Karaman greben	Blue	1800 (8.15%)	7.5	6.53%	0%
Karaman greben	KGB	4-seat chairlift	Karaman greben	Blue	3000 (13.58%)	8	21.49%	46,58%
Gvozdac	GVO	T-bar	Mali Karaman	Black	1200 (5.43%)	4.5	2.43%	0.37%
Knezevske bare	KNE	Rope-tow	Mali Karaman	Blue	900 (4.07%)	4.5	2.41%	0%
Mali Karaman	MAK	4-seat chairlift	Mali Karaman	Blue	2400 (10.86%)	5	17.97%	7.62%
Marine vode	MAR	Rope-tow	Karaman	Blue	900 (4.07%)	4.5	4.18%	0.58%
Karaman	KAR	Rope-tow	Karaman	Blue	900 (4.07%)	4.5	3.51%	0.87%
Jaram	JAR	Rope-tow	Jaram	Blue	900 (4.07%)	3	0.9%	0.55%
Gobelja relej	GOR	Rope-tow	Gobelja relej	Red-Black	900 (4.07%)	3.5	2.43%	0.59%

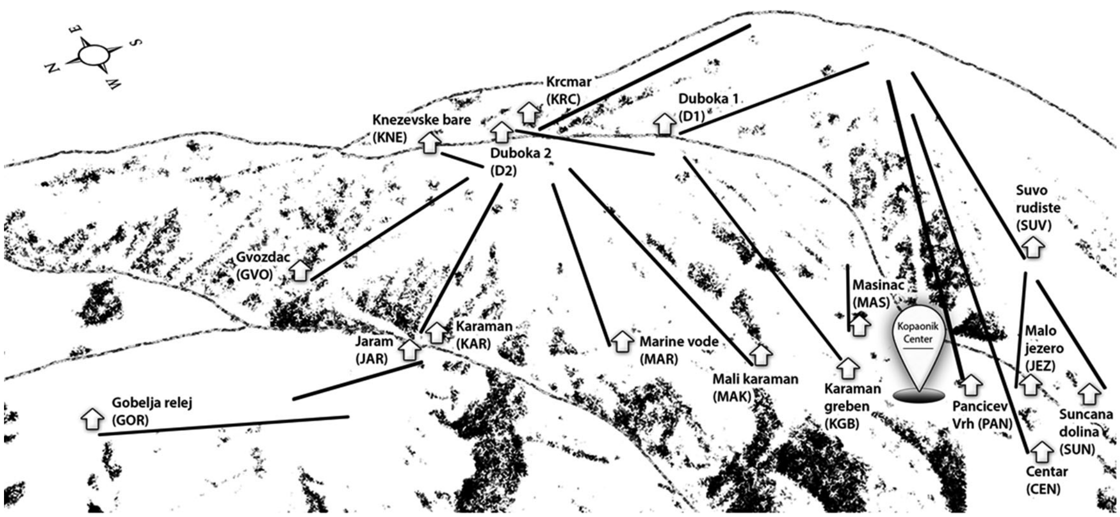

Fig. 1. Mt. Kopaonik ski resort ski lifts in season 2009.

and the observed percentage of the first ski lift transportations (the first transport that skiers make during a day). It should be noted that not all ski lifts can be used as starting points for skiers, as they are not near accommodation facilities. All but one ski lift has only one exit point—"D1" ski lift has an intermediate exit and entrance point.

The ski lift transportation system in the ski resort is presented in Fig. 1. "SUN" is the most southerly positioned ski lift, while "GOR" is the most northerly positioned one. The ski resort center, where most of the accommodation facilities are positioned, has access to the ski lifts "CEN," "JEZ," "PAN," "MAS," and "KGB."

C. Analysis Methods

RFID data systems are commonly used to control the entrance of passengers to transportation systems (e.g., metro systems [11], ski lift systems [5]) and for rentals (e.g., bicycles [7], cars—http://www.zipcar.com/).

Clustering is frequently used for analyzing RFID-collected transportation data. In a previous study, three techniques for spatial clustering of London Cycle Hire Scheme were used [1]. A spatiotemporal cluster analysis of bicycle station usage from Barcelona shared bicycling system was also considered [7]. Hierarchical clustering was applied to identify common behavior across stations and show how these behaviors relate to location, neighborhood, and time of day. Temporal clusters were identified with hierarchical clustering on London metro transportation data [11]. Four fuzzy k-medoids algorithm modifications were used to mine clusters from a single skier-day in the Val Gardena ski resort, Italy [5]. Two types of skiers have been identified—variety seekers and loyal skiers. Loyal skiers tend to stick to one or few ski lifts, while variety seekers tend to change ski lifts often during a skiing day. Besides clustering, association rule mining was proposed to be used for injury patterns extraction on ski lift transportation data [3].

In this paper, k-means was used to cluster spatial and temporal patterns of skier transportation.

K-means clustering is probably the most popular clustering algorithm, and it is identified as one of the top ten data mining algorithms [21]. It has been successfully used for behavior analysis from sensor network systems [15], unsupervised activity detection [17], clustering of user activities [12], and clustering points of interest in daily physical activity data [2].

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

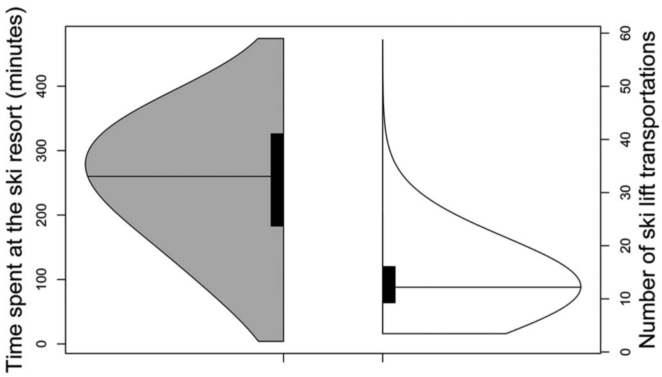

Fig. 2. Distribution of daily time spent at the ski resort (left-hand side) and daily number of ski lift transportations (right-hand side).

D. Descriptive Statistics

Two rough, yet integral measurements used to describe skier transportation are the time skiers spend daily in a ski resort (calculated as the sum of all the minutes elapsed between the first and the last lift transportation per day/per skier) and the number of lift transportations they make. The measurements distributions are shown on violin plots in Fig. 2. A violin plot represents a combination of a box plot and a kernel density plot [9]. A violin plot places a box plot in the middle of a bulge. The bulge is formed by the kernel density plots, i.e., a kind of smooth histogram.

The left-hand side violin of Fig. 2 illustrates the daily distribution of the time that skiers spent in the ski resort. The horizontal line around the middle of the bulge indicates the median value (4.3 h), while the top of the black box is the 75th percentile (5.4 h) and its bottom is the 25th percentile (3 h). It can be noted that the distribution is quite normally balanced around the median value. Since the ski resort is operating from 9 A.M. to 4 P.M., the majority of skiers are actually not using the full potential of their ski tickets, which is a rather surprising fact.

The right-hand side of Fig. 2 depicts the distribution of ski lift daily transportations. On average, there are 11.77 (S.D. 5.73) ski lift transportations per skier per day. The horizontal line around the middle of the bulge indicates the median value (11 runs) while the top of the black box is the 75th percentile (15 runs) and its bottom is the 25th percentile (eight runs).

The research from [5] reveals that there are potentially two types of skiers—those who stick to a ski lift and those who change lifts often. In the Mt. Kopaonik dataset, 50% of ski lift transportations are repeated at least once. Considering the transitions between lifts, common patterns for the skier pathways can be discovered. In Fig. 3, a Markov chain of transition probabilities between ski lifts is shown. We applied a process discovery method proposed by Wolff [19]. They proposed a model-based clustering that exploits first-order Markov chain to partition process behavior. In our case, ski lift transportations play the role of process activities, and the output is a Markov chain that fits the actual sequences of ski lift visits. We used the sequence clustering plugin implementation provided with ProM 5.2 (http://www.promtools.org/doku.php?id = prom52) to generate the chain, where transition probabilities between ski lifts can be observed.

It can be noticed that the probability of repeating a ride on ski lifts is usually about 50%. There are, however, some exceptions, i.e., lifts with very small repetition probabilities, which can serve as an indicator that a ski lift is inefficient (e.g., "CEN," "SUV" (both dismantled after season 2009), and "KRC" [dismantled after season 2014)] or that a lift is mostly used for transition (e.g., "JAR," used for access to "GOR").

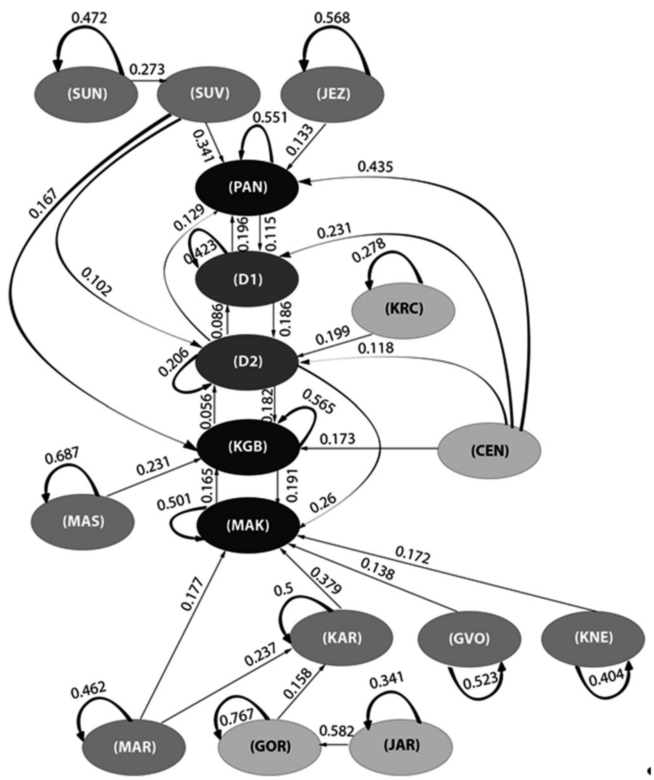

Fig. 3. Markov chain of Mt. Kopaonik ski lifts. Darker nodes have more ski lift transportations. Only transitions with probabilities larger or equal to 5% are displayed.

TABLE II
AVERAGE NUMBER OF LIFTS USED DURING A DAY, AND CUMULATIVE NUMBER OF LIFTS USED DURING A WEEK

Day of skiing	Average distinct locations visited (S.D.)	Cumulative distinct locations visited (S.D.)
1	3.59 (2.13)	3.59 (2.13)
2	3.98 (2.22)	5.20 (2.67)
3	4.01 (2.18)	6.21 (2.92)
4	4.04 (2.17)	6.98 (3.07)
5	4.24 (2.21)	7.61 (3.20)
6	4.40 (2.23)	8.06 (3.31)

On the other hand, certain lifts have very high repetition probabilities (e.g., "GOR") due to the fact that some lifts do not offer possibilities to change ski lifts, so skiers repeat skiing using the same lift. The darker nodes represent lifts that were visited more frequently. The most occupied ski lifts are "PAN," "D1," "D2," "KGB," and "MAK." They are the hubs of the ski resort with more than 70% of ski lift transportations (see Table I), and they are used as transfer lifts to other ski lifts, as well.

Table II shows the average number of distinct lifts used, and the cumulative number of lifts used during a week, for each skiing day (1 through 6 for six-day tickets). It can be noticed that skiers tend to use more different lifts with an increase in skiing days. Skiers usually visit only four different ski lifts during a day, and during the whole week, on average, they visit a total of eight ski lifts (note that there were 17 ski lifts operating).

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

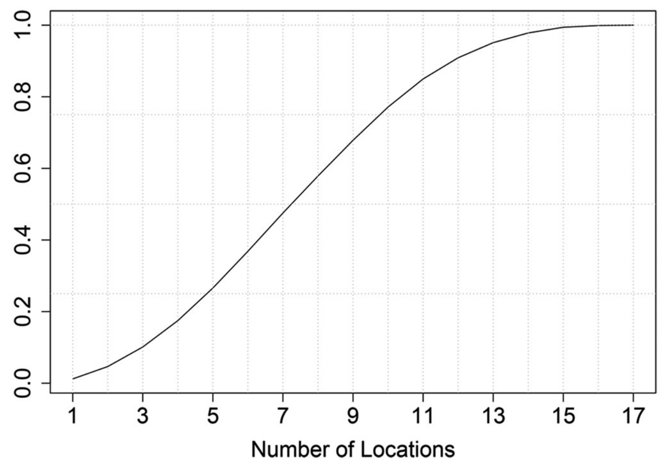

Fig. 4. Cumulative distribution of skiers visiting the 17 Mt. Kopaonik ski lifts.

In Fig. 4, the distribution of skiers visiting 1 through 17 lifts is shown. On average, skiers actually visit only half of the ski resort. Ninety percent of the skier population visits no more than 12 lifts. Ninety-nine of the population visits no more than 15 lifts.

E. Temporal Clustering

The original ski lift transportation dataset (<skier>, <day>, <time>, <ski lift>) was rearranged for clustering. Each record presented a skier-day, and columns were normalized frequencies of runs on temporal one-hour bands [7] (eight attributes in total, from 9 A.M. to 4 P.M.).

A clustering procedure from [11] was adapted to produce robust clusters.

1) The dataset was split into ten folds.
2) *K*-means [13] implementation in Orange [4] was applied on all bins. *K* (number of clusters) was iterated from 2 to 20. *K*-means was initialized so the starting centroids were the points farthest apart. The optimal number of clusters was selected by using the silhouette index [16] validation measure.
3) Since the optimal *K* was quite stable on all folds, with an average value of 9.3 (S.D. 0.67), *K*-means was rerun with $K = 9$ on all folds. The silhouette index had an average value of 0.41 (0.04). Ten *K*-means models were produced, each having nine centroids (a cluster centroid is the average point of all objects belonging to that cluster).
4) The nine centroids produced on the ten folds (90 centroids in total) were cross-joined using hierarchical agglomerative clustering with weighted average linkage, where the number of clusters was set to 9.
5) The final nine centroids were obtained thereafter.

F. Spatial Clustering

For spatial clustering, the original dataset was transformed so each row represents a skier-day and each column the frequency of each ski lift uses. The same procedure used for temporal clustering was also applied for spatial clustering. The optimal clusters number was on average 17.3 (S.D. 0.44). Then, we set *K*-means to search for 17 clusters in each fold. Each cluster actually was matched to its most frequent ski lift. The average silhouette value for the clustering models, averaged over the ten folds, was 0.42 (0.01).

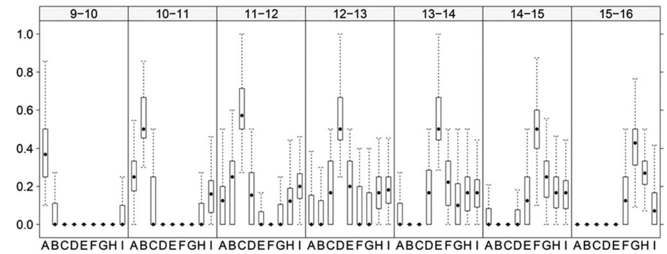

Fig. 5. Distribution of cluster members per hour band. Clusters are labeled A through I.

III. RESULTS

A. Temporal Clustering

The revealed centroids and their population percentage are the following.

1) *9 A.M. centroid (2.62%)*: skiers active a couple of hours that achieve maximal performance from 9 to 10 A.M.
2) *10 A.M. centroid (4.26%):* skiers active a couple of hours that achieve maximal performance from 10 to 11 A.M.
3) *11 A.M. centroid (4.84%):* skiers active a couple of hours that achieve maximal performance from 11 A.M. to 12 P.M.
4) *12 P.M. centroid (5.56%):* skiers active a couple of hours that achieve maximal performance from 12 to 1 P.M.
5) *1 P.M. centroid (5.06%):* skiers active a couple of hours that achieve maximal performance from 1 to 2 P.M.
6) *2 P.M. centroid (3.75%):* skiers active a couple of hours that achieve maximal performance from 2 to 3 P.M.
7) *3 P.M. centroid (2.16%):* skiers active a couple of hours that achieve maximal performance from 3 to 4 P.M.
8) *Amplifying centroid (3.63%):* skiers active most of the day (yet avoiding early morning and slightly preferring late afternoon). These skiers have a trend of increasing their activities during the day. Maximal performance is achieved from 2 to 3 P.M.
9) *Baseline centroid (68.02%):* skiers active most of the day and achieve maximal performance from 11 A.M. to 12 P.M. This cluster includes the largest share of skiers and represents the most common transportation pattern.

The distribution of cluster members' frequencies of runs per hour band is shown in Fig. 5. The big black dots inside the boxes present median values. The edges of the boxes represent the first and third quartile (the top of the box is the 75th percentile and the bottom the 25th percentile), while the dotted lines extend out from the top and the bottom of the box to 1.5 times the interquartile range (assuming there are data points out that far). Finally, outliers are not illustrated to avoid cluttering the diagram. The *y*-axis of Fig. 5 shows the relative frequencies of population cluster runs for every time band. The *x*-axis is repeating the cluster labels for every band.

The baseline skier population temporal cluster has a share of approximately ⅔ of the six-day ticket skier population and achieves maximal performance at 11 A.M. However, nearly 30% of skiers with a valid whole day ski pass tend to ski only a couple of hours and achieve maximal performances in various time periods from 9 A.M. to 4 P.M.

B. Spatial Clustering

The results of spatial clustering are shown in Fig. 6. Here, the *x*-axis represents clusters, while the *y*-axis represents ski lifts. The cell values show the distribution of ski lift transportations of clusters on different lifts (values were normalized for every cluster). The row names, besides

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

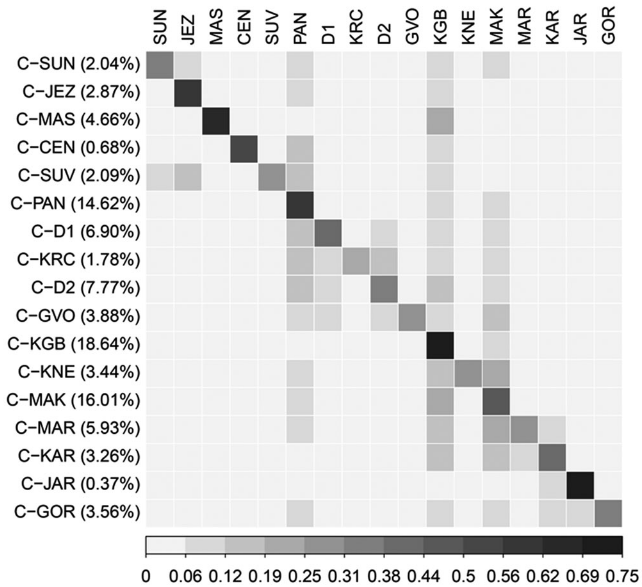

Fig. 6. Seventeen spatial clusters. Rows represent clusters, and columns ski lifts. The cells represent the intensity of a cluster using a lift.

the labels of the clusters, also present the percentage of the population that belongs to the corresponding cluster.

Several observations can be made.

1) A strong diagonal shows that skiers tend to choose one lift as the most occupied on a specific day. This is in line with the conclusions in Fig. 3, where most ski lift transportations have a high probability to be repeated.
2) The three most occupied ski lifts used by all clusters are the major hubs of the ski resort, i.e., “PAN,” “KGB,” and “MAK,” which can be also noticed in Fig. 3, as these lifts provide the most ski lift transportations on the resort.
3) Skiing clusters tend to contain, besides the major lift, between 1 and 5 additional ski lifts (average = 3, S.D. = 1.33).
4) Skiers tend to use ski lifts which give them access to slopes of equal difficulty.
5) There is spatial proximity among ski lifts in every cluster.
6) Skiers tend to group around ski lifts, i.e., ski lifts are natural cluster representatives on a ski resort.

IV. Discussion

Ski resorts use RFID technologies to control skier transportation through lifts in the mountain. This leads to massive data collection since every single lift usage is recorded. However, this massive collection of data is still underexploited.

The analysis presented in this paper provides valuable insight for understanding temporal and spatial ski lift transportation patterns which could help in recognizing popular ski lifts. Ski lifts may periodically become either candidates for dismantlement, or for replacement with lifts of a more appropriate capacity, or even for technological upgrades. This kind of decision not only supports control for congestion, but can improve user experience as well. Our analysis can be seen as an initial exploratory step of a full technical analysis aimed to provide specific recommendations; however, such a discussion is out of the scope of this paper.

This research was conducted on a part of the total population and for a certain time period. The major limitation of this paper is the fact that the ski resort does not track demographic data of the ski population, so such data were not available. Although the sample represents the biggest group of skier population, complementing the dataset with additional records could yield richer insights.

V. Conclusion

The objective of this study was not to develop a new method, but to discover useful knowledge in skier transportation data, and this is achieved by finding several types of skier behaviors, which benefits to evidence-based ski resort decision support and management. Skier transportation patterns have been analyzed on the weekly ski tickets holder population. Only 70% of the population fully exploit the daily time capacity of the ski tickets, and 90% of the population visit maximum 70% of the ski lifts at least once during the whole week. Skier transportation patterns differ by length of skiing, hours of maximal performance, and the set of ski lifts used. There is a gap between ski lift transportation patterns and temporally and spatially uniformly distributed usage of ski resort facilities, which could motivate new products and services to be offered to skiers. To conclude, this work is a step toward exploiting the existing, underused resources for evidence-based ski resort decision support and management.

Acknowledgment

The authors are grateful to the ski resorts of Serbia for providing data for this research and for their support throughout the research.

References

[1] R. Beecham, J. Wood, and A. Bowerman, “Studying commuting behaviours using collaborative visual analytics,” *Comput., Environ. Urban Syst.*, vol. 47, pp. 5–15, 2014, doi: 10.1016/j.compenvurbsys.2013.10.007.

[2] J. H. Chiang, P. C. Yang, and H. Tu, “Pattern analysis in daily physical activity data for personal health management,” *Pervasive Mobile Comput.*, vol. 13, pp. 13–25, 2013, doi: 10.1016/j.pmcj.2013.12.003.

[3] B. Delibasic and Z. Obradovic, “Towards a DGSS prototype for early warning for ski injuries,” in *Proc. 28th IEEE Int. Conf. Data Eng. Workshops*, Apr. 2013, pp. 68–72, doi: 10.1109/ICDEW.2012.73.

[4] J. Demšar *et al.*, “Orange: Data mining toolbox in Python,” *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 2349–2353, 2013.

[5] P. D’Urso and R. Massari, “Fuzzy clustering of human activity patterns,” *Fuzzy Sets Syst.*, vol. 215, pp. 29–54, 2013.

[6] M. Falk, “A hedonic price model for ski lift tickets,” *Tourism Manage.*, vol. 29, no. 6, pp. 1172–1184, 2008.

[7] J. Froehlich, J. Neumann, and N. Oliver, “Sensing and predicting the pulse of the city through shared bicycling,” in *Proc. 21st Int. Joint Conf. Artif. Intell.*, Jul. 2009, pp. 1420–1426.

[8] J. L. Hintze and R. D. Nelson, “Violin plots: A box plot-density trace synergism,” *Amer. Statist.*, vol. 52, no. 2, pp. 181–184, 1998, doi: 10.1080/00031305.1998.10480559.

[9] M. King, A. Abrahams, and C. Ragsdale, “Ensemble methods for advanced skier days prediction,” *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1176–1188, 2014.

[10] N. Lathia, C. Smith, J. Froehlich, and L. Capra, “Individuals among commuters: Building personalised transport information services from fare collection systems,” *Pervasive Mobile Comput.*, vol. 9, no. 5, pp. 643–664, 2013, doi: 10.1016/j.pmcj.2012.10.007.

[11] M. Lv, L. Chen, and G. Chen, “Mining user similarity based on routine activities,” *Inf. Sci.*, vol. 236, pp. 17–32, 2013, doi: 10.1016/j.ins.2013.02.050.

[12] G. W. Milligan and M. C. Cooper, “Methodology review: Clustering methods,” *Appl. Psychol. Meas.*, vol. 11, no. 4, pp. 329–354, 1987, doi: 10.1177/014662168701100401.

[13] B. Pfleging, A. Schmidt, and F. Michahelles, “Ubiquitous connectivity in the mountains: Enhancing the ski experience,” *IEEE Pervasive Comput.*, vol. 12, no. 2, pp. 5–9, Apr.–Jun. 2013.

[14] K. Rieping, G. Englebienne, and B. Kröse, “Behaviour analysis of elderly using topic models,” *Pervasive Mobile Comput.*, vol. 15, pp. 181–199, 2014, doi: 10.1016/j.pmcj.2014.07.001.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[15] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

[16] J. Seiter, O. Amft, M. Rossi, and G. Tröster, "Discovery of activity composites using topic models: An analysis of unsupervised methods," *Pervasive Mobile Comput.*, vol. 15, pp. 215–227, 2014, doi: 10.1016/j.pmcj.2014.05.007.

[17] L. Vanat, International Report on Snow & Mountain Tourism: Overview of the key industry figures for ski resorts, 2014. [Online]. Available: http://www.isiaski.org/download/20140517_ISIA_Vuokatti_1b_presentation_vanat.pdf

[18] G. M. Veiga and D. R Ferreira, "Understanding spaghetti models with sequence clustering for prom," in *Proc. Bus. Process Manage. Workshops*, Jan. 2010, pp. 92–103, doi: 10.1007/978-3-642-12186-9_10.

[19] F. C. Wolff, "Lift ticket prices and quality in French ski resorts: Insights from a non-parametric analysis," *Eur. J. Oper. Res.*, vol. 237, no. 3, pp. 1155–1164, 2014, doi: 10.1016/j.ejor.2014.02.062.

[20] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008, doi: 10.1007/s10115-007-0114-2.

[21] H. Yang, C. Li, and J. Hu, "RFID intrusion detection with possibilistic fuzzy c-means clustering," *J. Comput. Inf. Syst.*, vol. 6, no. 8, pp. 2623–2632, 2010.