

Sequential Pattern Mining of Tourist Spatiotemporal Movement

Maria Isabel R. Abucejo¹, Jovelyn C. Cuizon^{2*}

¹Holy Name University, Tagbilaran City, Philippines,

<https://orcid.org/0000-0001-8217-9494>

²University of San Jose - Recoletos, Cebu City, Philippines,

<https://orcid.org/0000-0002-1990-7943>

*Email Correspondence: jvlync@gmail.com

Abstract

The study aimed to develop a software application to capture tourist activity information, extract movement patterns from the dataset through sequential pattern mining (SPM), and visualize spatiotemporal movement. Tourist activity information was captured through crowdsourced trajectory movements by scanning unique QR (Quick Response) codes for each visited tourist spots. The AprioriAll algorithm was used to find frequent trajectory patterns on tourist visits. The resulting maximal k-sequences and their subsequences represent the recommended trip itinerary. The spatial and temporal movements were visualized through a flow map and a heat map, respectively. The directed edges in the flow map show the recommended sequence of tourist sites to visit. The heat map shows the density of tourist visits in different areas at time intervals. The application was validated with selected tour planning experts to verify functional suitability, usability, and acceptability. Experimental results show positive indicators that the application met the users' expectations.

Keywords: data mining, sequence pattern mining, AprioriAll, spatiotemporal, flow maps

1.0 Introduction

Sequential pattern mining (SPM) of tourists' spatiotemporal data helps in understanding tourists' mobility for successful and efficient destination management, product and image development, and attraction marketing (Xia et al., 2005); (Lew & McKercher, 2006). SPM is a data mining technique that finds interesting sequential patterns, sequences of item sets that frequently occurred in a specific order (Kumar & Srinivas, 2011). Tourist spatiotemporal data are sequences of location visits in longitude and latitude and the time visited (Li, 2014). The sequential transitions between tourist destinations could reveal popular

paths shared by tourists, which could help design offers to promote low travel frequency areas and provide itinerary recommendations (Thimm & Seepold, 2016).

Existing studies on tourist tracking reveal different methods in extracting tourist footprints ranging from the traditional way (Caldeira & Kastenholz, 2015) such as surveys and interviews to technology-based approaches such as GIS tools (Grinberger et al., 2014), digital travel journals and mobile tagging (Raun et al., 2016). Surveillance cameras to track volunteer trajectories in an indoor activity were also used to detect points of interest and provide ambient tourist guidance (Basiri et al.,

2018). Location-based social network (LBSN) data provide spatiotemporal metadata useful to capture tourist activity information. Regions-of-interest (RoI) visitation sequences identified from Flickr geotagged photos through SPM helped discover the relationship between hotspots (Birmingham & Lee, 2014). Another study used the Pattern-Growth method, which uses the Prefix-Span algorithm to find tourist sequential activity patterns from Foursquare check-ins (Talpur & Zhang, 2018). Flow maps were used to visualize spatiotemporal movements and contact patterns of humans from telco data to track infectious disease transmission (Ni et al., 2017).

While existing positioning technologies were proven to effectively track tourist movements, recording them through GIS tools and location-aware applications, poses data privacy issues as the tourists might think they are on surveillance. QR (Quick Response) code to capture tourist activity may solve these concerns since data is limited only to the visited tourist spots. Tourists are assigned a unique QR code that retrieves only relevant attributes such as nationality, age, and gender. Captured data are also anonymized to avoid exposure to sensitive personal details.

This study aims to develop Bohol-Tour-Trace, a software application to track the tourists' movements through QR code, applies sequence pattern mining to analyze the tourists' spatial and temporal patterns to recommend point-to-point routes visualized through flow and heat maps. The application will help the local tourism office analyze trends relating to travel and create a databank of tourist activity movement in Bohol. Moreover, this study will support the local government unit and private tour operators' initiatives in implementing the mandates of Republic Act No. 9593, otherwise known as the Tourism Act of 2009. One of the

objectives of this Act is to enhance the collection, analysis, and dissemination of data that accurately measure the impact of tourism to facilitate planning in the public and private sectors. Responsibilities and roles of the different tourism sectors include undertaking research and data gathering on local tourism trends and other relevant tourism information and provide them to the Tourism Department for use in the implementation of the Act (The Tourism Act, 2009).

2.0 Methodology

The project consists of two (2) main components: a mobile application for the tourist and a web-based dashboard for data mining and visualization for tourist spot operators, and the government agencies to track tourist activity while on tour.

The application runs on mobile and web platforms. The mobile component was developed using React Native, a cross-platform mobile development framework, while the web application was developed using Laravel, a PHP MVC (Model-View-Controller) framework. This study used Axios to manage process requests between mobile and web client applications through RESTful web service calls; while OpenStreet map was used for creating the editable map and leaflet libraries for the generation of the flow map and heat map.

The mobile application includes a registration form, trip history and page on facts about Bohol. A unique QR code is generated for each user which serves as unique identifier when visiting a tourist site. User also adds information of companions who are minors and/or person with disability (PWD) for them to get counted as well. The tourist spot operator uses the web application as kiosk to scan QR codes of visiting tourists. It also provides an interface which displays demographic

summary of visits. The concerned government agencies monitor tourist foot traffic through a web dashboard. A back-end process runs sequence pattern mining, an algorithm to generate tour itinerary, display flow map of tourist movement and heatmap of tourist density.

Tourist Activity Information Capture

Tourist spatiotemporal data are captured through crowdsourcing. Crowdsourcing is a simple yet powerful concept for collecting data and generating information from many contributors (Basiri et al., 2016). Tourists visiting Bohol will be required to install and register an account with Bohol-Tour-Trace, a mobile application that tracks tourist spots visited in the itinerary. It automatically assigns a unique QR code containing tourist information that includes tourist ID, age, gender, and nationality. Figure 1 presents screenshots of the registration and tourist information capture form.

Tourist spots and accommodation providers require all visitors to scan the QR code through designated QR code scanners upon entry to their premises, as shown in Figure 2. Spatial and temporal data (tourist information, date of visit,

time of visit, and spot ID) gathered are saved in a centralized database accessible by both tour operators, and the tourism office through the web dashboard. The mobile application also maintains records of details of current and previous trips.

The web dashboard shows the real-time statistics of tourist visits available to the tourism office for monitoring. The authorized user can also generate statistical reports about the tourists and tourist sites/stops, as shown in Figure 3.

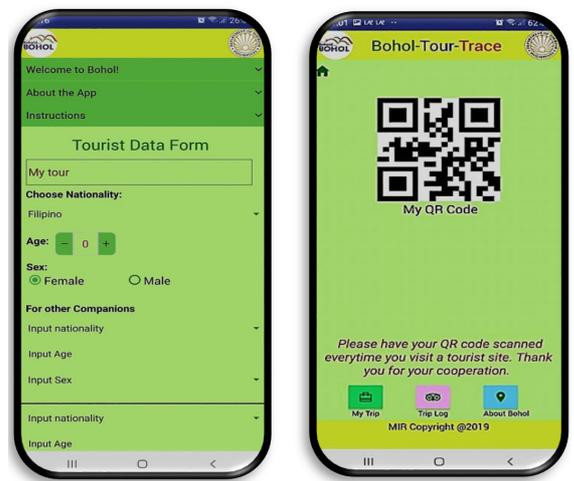


Figure 1. Screenshot of Mobile Registration and Tourist Information Capture

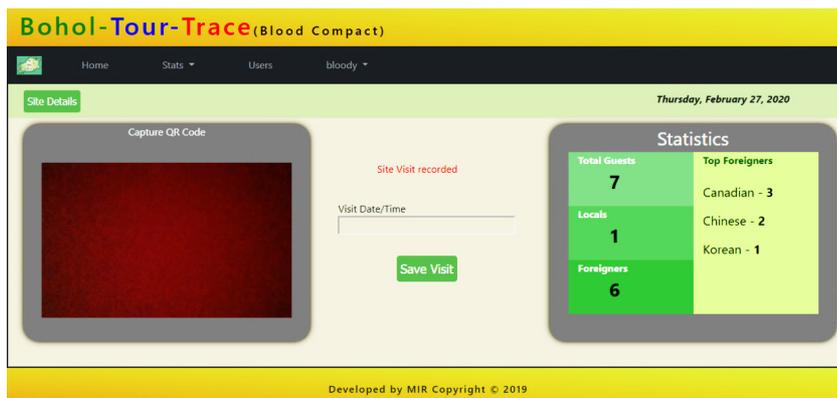


Figure 2. Screenshot of Tour Provider QR Scanner



Figure 3. Screenshot of the Web Dashboard

Sequential Pattern Mining

Records of past tourist visits from tour agencies that consist of anonymized tourist information such as age brackets, nationality, and tour itinerary

serve as the initial dataset composed of 200 tourists' visits, 15 active tourist spots, and about 900 site visit transactions. Table 1 presents a list of tourist spots uniquely identified by a tourist site ID.

Table 1. Active Tourist Spots

SID	Tourist Site Name	Latitude	Longitude
29	Blood Compact	9.628949	123.879588
30	Prony Python	9.6165	123.933
31	Loboc River Cruise	9.636443	124.031897
32	Man-Made Forest	9.665446	124.069528
33	Chocolate hills	9.811278	124.164377
34	Hanging Bridge	9.679177	124.043888
35	Bilar Eco Park	9.667926	124.102885
36	Necitas Jardin	9.866471	124.35433
37	Balicasag Island	9.5156	123.6841
38	Isola de San Francisco	9.5597	123.7206
40	Dumaluan Beach Resort	9.568476	123.788431
42	Hinagdanan Cave	9.620753	123.805704
43	Baclayon Church	9.622661	123.91373
44	Corella Tarsier Sanctuary	9.702052	123.936748
45	Billar Butterfly Sanctuary	9.703871	124.101575

Each visit transaction is an ordered set of tourist site IDs. Table 2 shows a snapshot of the sequence database.

Table 2. Sample tour site visit sequences

TID	Visit sequence
001	{29, 43, 30, 44, 31, 32, 33, 45, 34}
070	{37, 38, 40, 33, 34}
121	{37, 38, 40}
151	{42, 36}

Litem Identification

Litem (large item sequences) were extracted from the database representing a list of visit sequences that satisfies the minimum support, a threshold used to prune the list. The minimum support is derived using the exponential decay function adopted from the proposed solution (Fournier-Viger, 2010). The formula used to calculate the minimum threshold is:

$$minsum(x) = e^{(-ax-b)} + c$$

where:

- x is the number of transactions
- a and b are constants that influence how quickly the curve will decrease when x increases
- c the smallest value that the function will produce

The function adjusts the minimum support threshold automatically according to the database size, in the case of this research, the number of tourists' visits in. Given $a = 0.005$ and $b = 2$, and $c = 0.20$, the minimum support threshold for 200 tourist visits is 25%. Thus, the minimum support count is 50.

L1, large 1-sequences, is formed from tourist site IDs with more than 50 visits. Figure 4 shows the list of unique tourist sites and their support count.

Hinagdanan Cave, Bilar Eco Park, and Necitas Jardin were pruned from the list since the support count does not reach the threshold.

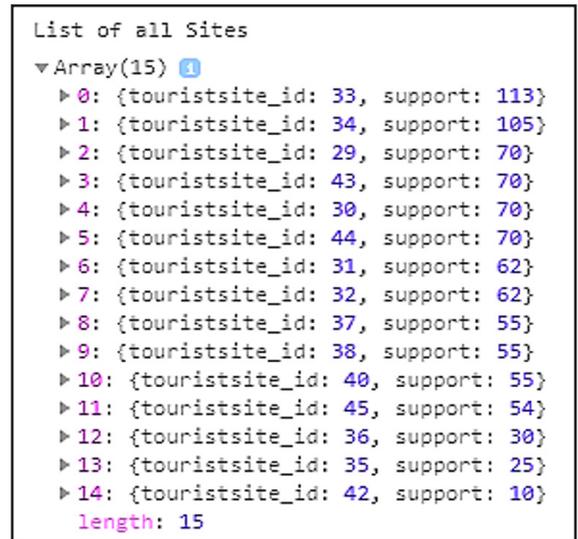


Figure 4. 1-Sequences Support Count

Maximal Sequences Generation

The maximal sequence generation starts with candidate generation of 2-sequences up to k-sequences from Litem or large 1-sequences using the Apriori-generate algorithm. Any (k-1)-subsequences of s (length k) that is not frequent cannot be a subsequence of a frequent k-sequence. Sequences that fall below the minimum support are trimmed from the candidate list. The resulting k-sequences are the input to generating the next k+1-sequences candidates. Table 3 shows k-sequences generated for $k = 2, 3$, and 4.

These steps are repeated until no more sequences are generated, or the generated sequences do not satisfy the minimum support. The last sequences that meet the minimum support are considered the maximal sequences. Table 4 shows k-sequences generated for $k = 5$ to 8.

Table 3. *k*-sequences for *k*=2,3,4

2-sequences	3-sequences	4-sequences
{33, 34}	{32, 33, 45}	{32, 33, 45, 34}
{32, 33}	{40, 33, 34}	{29, 43, 30, 44}
{40, 33}	{33, 45, 34}	{40, 33, 44, 31}
{33, 45}	{30, 44, 31}	{30, 44, 31, 32}
{45, 34}	{44, 31, 32}	{44, 31, 32, 33}
{29, 43}	{29, 43, 30}	{31, 32, 33, 45}
{43, 30}	{43, 30, 44}	{37, 38, 40, 33}
{30, 44}	{31, 32, 33}	{38, 40, 33, 34}
{44, 31}	{37, 38, 40}	
{31, 32}	{38, 40, 33}	
{37, 38}		
{38, 40}		

Table 4. *k*-sequences for *k*=5,6,7,8

5-sequences	{29, 43, 30, 44, 31} {43, 30, 44, 31, 32} {30, 44, 31, 32, 33} {44, 31, 32, 33, 45} {31, 32, 33, 45, 34} {37, 38, 40, 33, 34}
6-sequences	{29, 43, 30, 44, 31, 32} {43, 30, 44, 31, 32, 33} {30, 44, 31, 32, 33, 45} {44, 31, 32, 33, 45, 34}
7-sequences	{29, 43, 30, 44, 31, 32, 33} {43, 30, 44, 31, 32, 33, 45} {30, 44, 31, 32, 33, 45, 34}
8-sequences	{29, 43, 30, 44, 31, 32, 33, 45} {43, 30, 44, 31, 32, 33, 45, 34}

Candidate generation stops at 9-sequence {29, 43, 30, 44, 31, 32, 33, 45, 34} since there is only one sequence left and there are no other sequences that can be generated out of a single sequence. Pruning in large sequences is done by eliminating candidates that are subsequences of the maximal sequence. The only candidate which is not contained in {29, 43, 30, 44, 31, 32, 33, 45, 34} is the 5-sequence {37, 38, 40, 33, 34}. Thus, the maximal sequences with minimum support of 25% are {29, 43, 30, 44, 31, 32, 33, 45, 34} and {37, 38, 40, 33, 34}. The maximal sequences represent a frequent tourist activity pattern. Therefore, the most popular routes are tourist movement patterns {29, 43, 30, 44, 31, 32, 33, 45, 34} and {37, 38, 40, 33, 34}.

Visualization

Spatiotemporal data representing tourist mobility are visually presented through flow maps and heat maps. The flow map shows a directed graph depicting the maximal sequences with tourist sites as nodes and directed arrows representing the visit's sequence. Markers for each tourist site are plotted in an OpenStreet map layer using its location coordinates in longitude and latitude with a directed polyline to illustrate the movements from one place to another. Figure 5 shows the flow map of spatial movement.

The density of tourists visiting a site at a given period is shown in the red gradient in the heat map. It also displays the temporal details of the visits. The higher the concentration of points in an area, the more heat is applied, thus a darker intensity. Figure 6 shows a sample heatmap showing the density of visits from 9:00 AM to 12:00 PM based on the populated dataset.

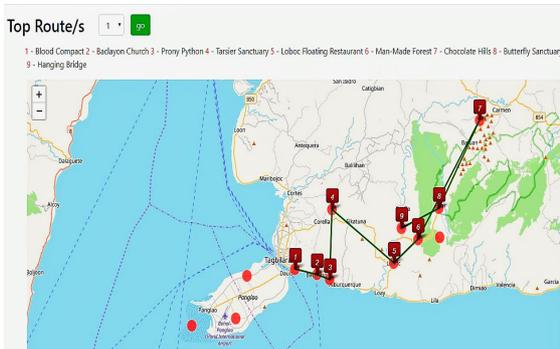


Figure 5. Spatial Flow Map of Tourist Movement

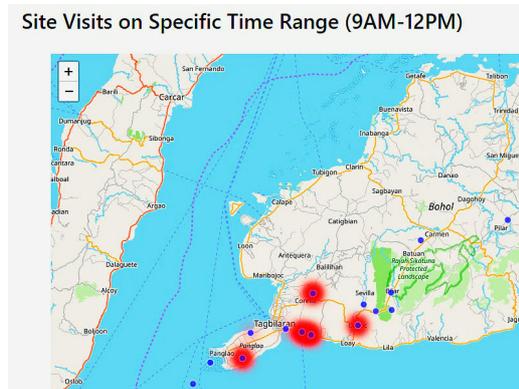


Figure 6. Temporal Pattern Heatmap

3.0 Results and Discussion

Spatial and Temporal Pattern

The maximal sequences representing frequent tourist mobility pattern are {29, 43, 30, 44, 31, 32, 33, 45, 34} and {37, 38, 40, 33, 34}. This provides tourists two recommended itineraries as follows:

Option 1: Blood Compact -> Baclayon Church -> Prony Python -> Corella Tarsier Sanctuary -> Loboc River Cruise -> Man-Made Forest -> Chocolate Hills -> Bilar Butterfly Sanctuary -> Hanging Bridge

Option 2: Balicasag Island -> Isola de San Francisco -> Dumaluan Beach Resort -> Chocolate hills -> Hanging Bridge

For temporal recommendations, most tourists go to the following places: beaches of Balicasag Island, and Isola de San Francisco; Blood Compact early in the morning; visit the historic Baclayon Church; take a look at the tarsiers in Corella Tarsier Sanctuary; take on a floating restaurant cruise at Loboc River at mid-morning; then experience nature with the Man-made forest; Bilar Butterfly Sanctuary; the Chocolate Hills in the afternoon; and end the day with a walk on the Hanging Bridge.

The subsequences of the maximal sequences may also be used to recommend a shorter itinerary. One of the AprioriAll algorithm's limitations is the absence of time constraints to specify the recommended minimum or maximum time between tourist sites in the itinerary.

Software Testing

To ensure that the software meets all the functional and non-functional requirements, Bohol-Tour-Trace was tested for functional suitability, usability, and performance.

Functional Suitability

The application was tested to determine whether expected features are available and working correctly. Testers are selected tourism office staff for each municipality where active tourist sites are situated. They were given test case scenarios as guide in performing the task. Testing results reveal functional suitability at an acceptable level. Suggestions for modifications and improvement were noted and were integrated into the application.

Usability Testing

A usability test was undertaken to determine whether the application is intuitive, user-friendly,

and visually appealing. Testers were given a survey questionnaire to check on the design, logic, navigation, and learnability.

Table 5. Summary of the Result of Usability Test

Usability Test Areas	Rating
Design (Layout of the webpage, Color Combination, Clarity of text (label and instructions), Tab, buttons, other elements used on the page, Tables, Graphs, Maps)	4.6
Logic and Navigation (shifting to another page, choosing an option on a menu, locating buttons, tabs, and other elements on the page)	4.43
Learnability (understanding instructions provided, remembering buttons, tabs, other elements used on the page, reading labels, tabs generating tables, graphs, and maps)	4.55

The test showed positive results in all usability test areas with an overall rating of 4.52, as shown in Table 5.

Acceptability Testing

To validate the spatial and temporal behavior results, an acceptability test with human experts in tourists' itinerary planning was done. Ten (10) tour agency specialists whose task is to prepare the itinerary for prospective clients were selected to validate system results. The evaluators were asked on a 5-Likert scale the likelihood of the tourists to take the tour itinerary generated by the system as an entire trip and the suitability of the suggested time (e.g., morning, mid-day, afternoon) to visit selected tourist sites. They were presented with the generated route and heat maps and were asked to comment on them.

Table 6. Summary of the Result of Acceptability Test

Usability Test Areas	Rating
The likelihood that tourist take the system-generated itinerary	4.7
Suitability of the suggested time for selected tourist sites	
Early Morning (04:00 AM – 08:00 AM)	4.4
Mid Morning (09:00 AM – 12:00 AM)	4.4
Afternoon (01:00 PM – 04:00 PM)	4.6

The result showed a positive acceptance rating among trusted and recognized tour agency staff/owners in Bohol with an overall rating of 4.5, as shown in Table 6.

4.0 Conclusion and Future Works

The study attempted to apply sequential pattern mining (SPM) on tourists' activity data to generate sequential movement patterns and visualize movement trails from spatiotemporal tourist data. It turned out AprioriAll algorithm, which is often associated with market basket analysis, also works well in an itinerary generation. Despite the limitation in the absence of time associated with each tourist spot in the recommended itinerary, the application has met user acceptance standards, as shown in the results in both usability and functional suitability tests from tour planning experts.

This method presented in this study may also be applied for contact tracing of people to monitor the transmission of disease and predict future location hot spots based on movement patterns. Future researchers may also explore other available SPM algorithms to add time constraints and user similarity to generate a personalized itinerary based on tourists' demographics.

References

- Basiri, A., Amirian, P., & Mooney, P. (2016). Using crowdsourced trajectories for automated OSM data entry approach. *Sensors, 16*(9), 1510. [10.3390/s16091510](https://doi.org/10.3390/s16091510)
- Basiri, A., Amirian, P., Winstanley, A., & Moore, T. (2018). Making tourist guidance systems more intelligent, adaptive and personalised using crowd sourced movement data. *Journal of Ambient Intelligence and Humanized Computing, 9*, 413–427. <https://doi.org/10.1007/s12652-017-0550-0>
- Bermingham, L., & Lee, I. (2014). Spatio-temporal sequential pattern mining for tourism sciences. *Procedia Computer Science, 29*, 379–389. <https://doi.org/10.1016/j.procs.2014.05.034>
- Caldeira, A.M., & Kastenzholz, E. (2015). Spatiotemporal behaviour of the urban multi-attraction tourist: Does distance travelled from country of origin make a difference? *Tourism & Management Studies, 11*(1), 91-97. https://www.researchgate.net/publication/277863124_Spatiotemporal_behaviour_of_the_urban_multi-attraction_tourist_does_distance_travelled_from_country_of_origin_make_a_difference
- Fournier-Viger, P. (2010). *A hybrid model for supporting learning in procedural and ill-defined areas* [Doctoral dissertation, Université du Québec à Montréal]. https://www.researchgate.net/publication/49121771_Un_modele_hybride_pour_le_support_a_l'apprentissage_dans_les_domaines_proceduraux_et_mal_definis
- Grinberger, A., Shoval, N., & Mckercher, B. (2014). Typologies of tourists' time-space consumption: A new approach using GPS data and GIS tools. *Tourism Geographies, 16*(1), 105 - 123. <https://doi.org/10.1080/14616688.2013.869249>
- Kumar, K.M.V.M., & Srinivas, P.V.S. (2011). Algorithms for mining sequential patterns. *International Journal of Information Sciences and Application, 3*(1), 59-69. http://www.irphouse.com/ijisa/ijisav3n1_9.pdf
- Lew, A., & McKercher, B. (2006). Modeling tourist movement: A local destination analysis. *Annals of Tourism Research, 33*(2), 403–423. [doi:10.1016/j.annals.2005.12.002](https://doi.org/10.1016/j.annals.2005.12.002)
- Li, Z. (2014). Spatiotemporal pattern mining: Algorithms and applications. In C.C. Aggarwal, & J. Han (Eds.), *Frequent pattern mining* (pp. 283-306). Springer International Publishing. [10.1007/978-3-319-07821-2_12](https://doi.org/10.1007/978-3-319-07821-2_12)
- Ni, B., Shen, Q., Xu, J., & Qu, H. (2017). Spatio-temporal flow maps for visualizing movement and contact patterns. *Visual Informatics, 1*(1), 57-64. <https://doi.org/10.1016/j.visinf.2017.01.007>
- Raun, J., Ahas, R., & Tiru, M. (2016). Measuring tourism destinations using mobile tracking data. *Tourism Management, 57*, 202-212. <https://doi.org/10.1016/j.tourman.2016.06.006>
- Talpur, A., & Zhang, Y. (2018, October 23-26). *A study of tourist sequential activity pattern through Location Based Social Network (LBSN)* [Conference session]. 2018 International

Conference on Orange Technologies (ICOT), Nusa Dua, BALI, Indonesia. 10.1109/ICOT.2018.8705895

The Tourism Act. (2004). 14th Congress. 2nd Regular Session. Metro Manila. [http://tourism.gov.ph/Downloadable%20Files/Updated_RA_9593_and_IRR_\(as_of_01_Nov_2020\).pdf](http://tourism.gov.ph/Downloadable%20Files/Updated_RA_9593_and_IRR_(as_of_01_Nov_2020).pdf)

Thimm, T., & Seepold, R. (2016). Past, present and future of tourist tracking. *Journal of Tourism Futures*, 2(1), 43-55. 10.1108/JTF-10-2015-0045

Xia, J., Ciesielski, V., & Arrowsmith, C. (2005). *Data mining of tourists' spatio-temporal movement patterns - a case study on Phillip Island* [Paper presentation]. 8th International Conference on GeoComputation, University of Michigan, USA. <http://www.geocomputation.org/2005/Xia.pdf>