Using flow-comap technique to visualize spatial-temporal patterns of public bike sharing program and the effect of weather and contender events

Tiebei Li and Jonathan Corcoran

Abstract

Public Bicycle Sharing Programs (PBSPs) have become prominent features across city spaces worldwide. Despite the rapid rise of this new transport opportunity, there has been relatively limited research on the underlying dynamics of these schemes that is arguably underpinned by a lack of data made available to researchers. This paper redresses the deficit using trip level data from Brisbane's "CityCycle", which provide an opportunity to investigate the spatio-temporal dynamics of a large urban PBSP system and the effects of weather and calendar events. Employing a flow-comap technique we explore the impact of salient weather conditions and calendar events on the spatio-temporal dynamics of the case study PBSP. We conclude through highlighting how the results from the analyses may form part of an evidence base for policy makers with the potential to inform future PBSP expansions to further enhance uptake of this non-motorised urban transport mode.

T. Li (Corresponding author)
Urban Research Program, Griffith University, Brisbane, Australia
Email: t.li@griffith.edu.au

J. Corcoran
Email: jj.corcoran@uq.edu.au
School of Planning Architecture and Environmental Management, Univer-sity of Queensland, Brisbane, Australia.

1. Introduction

Public Bicycle Sharing Programs (PBSPs) have a long history dating back to an unsuccessful 1965 pilot scheme in Amsterdam, which was plagued by vandalism and theft. The recent explosion in size and scale of PBSPs has been facilitated by technological advancements allowing bicycle riders to be tracked using technology pioneered in a small scale program (25 bicycles during the scheme's pilot phase, eventually increasing to 75 bicycles) at Portsmouth University in 1996 (Black & Potter n.d.), and subsequently rolled out at a city-wide scale in Renne in 1998, followed by Munich in 2000. These early programs were replicated at still larger scales in Lyon in 2005 (4000 bicycles in the scheme (Henley 2005)) and then Paris in 2007 (7000 bicycles initially expanding to 23,600 (DeMaio 2009)). Rapid growth in PBSPs followed and, at the time of writing, PBSPs now cover five continents and in excess of 486 urban areas (Ahillen, et al. in review). Central to the success of the modern PBSP, is a system for tracking the identity of bicycle users, the result of which are rich datasets which have the potential to unveil the underlying dynamics of PBSPs, and provide insights into the contribution of these schemes to urban public transport.

Two distinct types of data capturing the dynamics of PBSPs have emerged in recent years: data capturing stocks, and data capturing flows. Stock data track variations in the number of bicycles docked at stations within a PBSP network. Such data can provide insights into spatio-temporal fluctuations in demand across a system (see Kaltenbrunner et al. 2010), but do not capture the underlying mobility behaviour (e.g. trips). Flow or trip data have emerged much more recently, and are consequently more limited in their availability. At present, the use of PBSP trip level data has been confined to a small number of studies that have been restricted in scope to measuring operational dimensions of the PBSP. One such example of an operationally-focussed study is the analysis of spatial flows by Jensen et al. (2010). The primary focus was the determination of average transport speed within the network. A second study of the Parisian PBSP (Vélib) by Nair et al. (2013), was concerned with the issue of establishing 'balance' across the bicycle network. To date, there has been little attempt to analyse these rich trip level datasets with the more generic goal of better understanding the spatio-temporal dynamics of human spatial mobility. It is the purpose of this paper to demonstrate how trip level data can be better interrogated to help shed new insights into the dynamics of PBSP journeys. We demonstrate the utility of these data by exploring how external effects, including weather conditions and calendar events (i.e. public holidays and weekends) impact on the spatial distribution and timing of PBSP trips in Brisbane, Australia.

The analysis conducted requires some innovative combination of existing spatial exploratory analysis tools. In particular, we embed flow maps (Thompson & Lavin 1996), which graphically present origin-destination flow matrices, into the comap (Brunsdon 2001), which displays spatial information conditionally on one or more external variables. This new analysis technique, we term the flow-comap, in combination with a suite of metrics, is demonstrated to be capable of illuminating the underlying spatio-temporal dynamics of our case study PBSP, Brisbane's CityCycle Scheme. The flow-comap is able to graphically depict subtle changes in usage patterns under different weather conditions (for example, a hot versus a cold day), and across various calendar events (for example, a public holiday versus a 'normal' working day).

The remainder of the paper is structured as follows: Section 2 provides a brief review of the emerging studies that have begun to explore the spatial databases of PBSPs. In Section 3, the data underpinning the study is detailed. In Section 4, the development of the flow-comap from the pre-existing flow map and comap techniques is discussed. In Section 5, results are presented followed by concluding remarks and a discussion of the insights that our exploratory data analysis brings in Section 6.

2. Spatial analytical approach on PBSP

The ultimate success of PBSPs is dependent on improved knowledge of the fundamental dynamics of the system (e.g. volume, pattern and timing of trips) along with the extent to which various system attributes (e.g. number of stations, climate regime etc.) affect usage. To date, considerably less research attention has been dedicated on the underlying dynamics of these schemes, which is arguably underpinned by a lack of data available to researchers. It is however crucial that planners not only understand the contexts where PBSPs are implemented, but also be able to evaluate, with the use of established metrics, the implications of evolving policies affecting PBSP use, in particular, so as to better respond to mobility needs of its populace.

Technological improvements introduced in new generation of PBSPs permit continuous monitoring of traffic flows of public bikes (Shaheen et al. 2010). As such, operators of public bicycle systems are able to gain access to real-time usage data of their networks. Collected via mobile devices or other ICT-based measures, this information can then improve our understanding of individual traveller's behaviour, offer real-time travel information, and also present personalised location-based services. Moreover, fine-grained data on the status of shared bicycles also enables an empirical measurement of the impacts of proposed system improvements or policy changes (e.g. fare restructuring) as well as the results of force majeure on the system (e.g. flooding).

A number of studies have utilised stock or usage data to explore spatial and temporal patterns. For example, examining Barcelona's shared bicycling system 'Bicing', Froehlich, Neumann and Oliver (2009) investigated user behaviour across stations in relation to location, neighbourhood, and time of day. Still in the European context, Borgnat et al. (2009) predicted the number of bikes hired per hour in Lyon's community bicycle program to describe the daily and weekly patterns. The prediction method involved several explanatory factors such as the number of subscribed users, the time of the week, the occurrence of holidays or strikes, and weather parameters. However, methods to identify and visualise spatio-temporal patterns (i.e. location and times when frequency of bike use is particularly high) based on flow or trip data have not been adequately examined in past studies, particularly, within the Australian context. There is, therefore, an imperative need to better understand the location, time and reasons for these individual uses to inform strategies to ensure a more successful PBSP implementation.

3. Data

Three databases are used in this study: (1) CityCycle trip data; (2) calendar events; and, (3) daily weather data. The three databases are integrated such that for any single bicycle trip it is possible to ascertain the weather conditions during which the trip occurred and whether it coincided with any salient calendar event. Each dataset is now described in further detail:

3.1 CityCycle trip data

CityCycle, Brisbane's self-service bicycle hire scheme, was launched in October 2010. At present, the system accommodates 150 docking station terminals distributed across Brisbane's city centre, stretching from the suburbs of Newstead in the north to West End in the south and St Lucia in the west (Figure 1). Payment for the system is based on a per-trip usage fee with the first 30 minutes being free of charge.

Disaggregate CityCycle trip data was drawn upon for a period covering 20 months (November 2010 to July 2012) and comprised a total of 285,714 trips. Each record contained the date and time of the bicycle release and return station. Spatial information describing the location of each of the 150 bicycle stations was also provided. Some aggregation of the spatial unit of analysis (from the individual bicycle to a larger unit) was required in order to enhance our capacity to visually depict the complex spatio-temporal flows using the flow-comap. This was deemed necessary given that the 150 CityCycle stations are located in relatively close proximity (between 300 to 500 metres) to one another. As such, the Australian Bureau of Statistics State Suburb was used as the unit of analysis resulting in the use of 16 areas for this study (Figure 1).

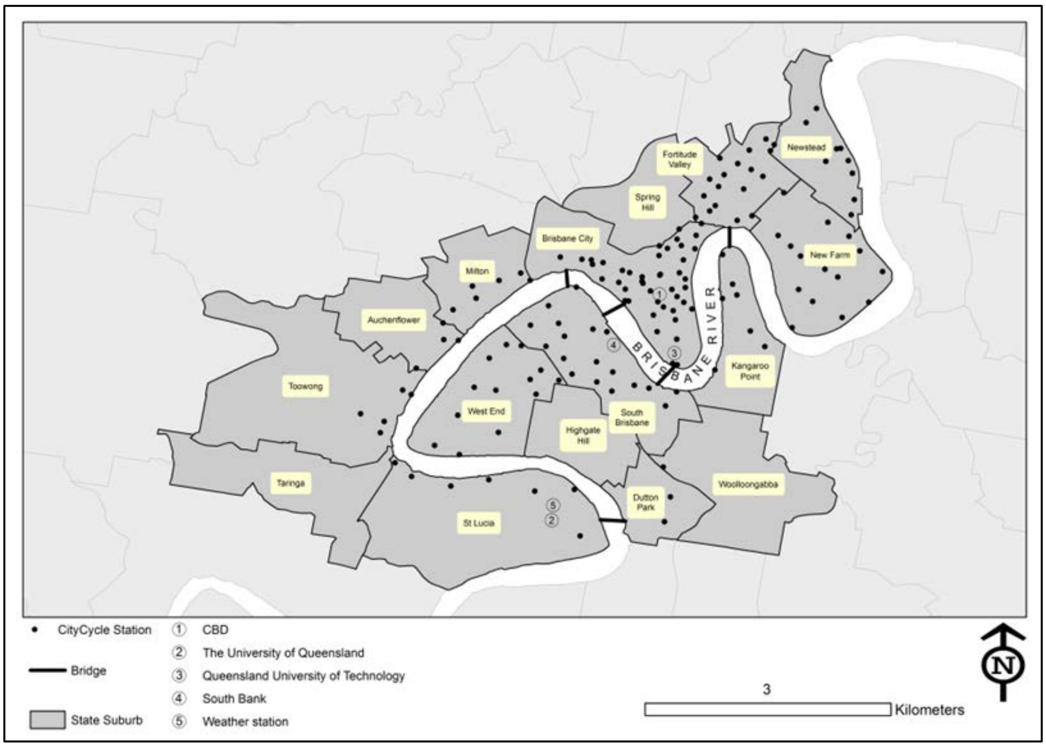


Fig. 1 Public bicycle stations in Brisbane

3.2 Calendar events

A database of all major calendar events during the study period was generated. The inclusion of specific dates or periods into a list of events was guided principally by previous research such as Brandenburg et al. (2007), and included all school and public holiday periods. School holidays (Department of Education, Training and the Arts (DETA) 2013) covered adjacent weekends and pupil free days but did not include non-adjacent weekends.

3.3 Weather

Data from a single weather station located central to the case study area was used; Data measurements included temperature, rainfall and wind speed. Given its centrality coupled with a relatively small study area (30 square kilometres), a single weather station was deemed adequate to represent the weather conditions across the CityCycle network. The data covered the same time period as the CityCycle database and reported weather measurements as daily averages. Weather data was integrated with the CityCycle trips on the basis of the closest set of weather measurements in time to indicate the weather conditions during which the trip took place.

4 Method

The CityCycle dataset contains trip level information in the form of an origin- destination matrix. The rows represent origin stations and the columns destination stations along with individual counts of transitions. Analysis of this type of dataset in raw form is not generally viable given that it consists of a large matrix of numbers with no geographic information included. This is particularly the case in multivariate situations where the difference in origin-destination matrices conditional on other variables (for example, hour of day) is of interest. The argument for the role of exploratory spatial data analysis has been made convincingly elsewhere (see for example, Haining et al. 1998; Anselin 1999) and would be strong in our case with the CityCycle dataset, where no simple descriptive statistic can be easily defined that captures the complex dynamics of an origin-destination matrix.

The flow map is a well-established cartographic technique. It can assist in our understanding of these matrices by mapping transitions between spatial units. Here, lines depicting the transitions between spatial units are typically appended with arrows to indicate flow direction and the width of the line is used to indicate the volume of flow. While the flow map can be used to develop an understanding of origin-destination matrices, it does not readily allow the incorporation of other variables such as weather and calendar events as a component of the visual output. In such circumstances, the analysis of bivariate spatial data can be analysed using a technique termed the comap in which plots with overlapping subsets of data are selected using non-spatial variables. These plots of raw data can be used to give a sense of how spatial relationships change conditional on one or more external variables (for example, how bicycle trips vary spatially according to how windy it was at the time the particular trips were made).

Flow mapping, a visual analytical tool to depict spatial interaction and movement, has a long history dating back to 1869 where Charles Minard first used the technique to depict Napoleon's army's advancement towards Moscow (Minard 1869). The development of computerised flow mapping tools, however, only commenced in the late 1980s with Waldo Tobler's flow mapper program. This program was designed to visualise discrete node-to-node movements (Glennon & Goodchild 2005). Since then, a number of significant efforts, particularly in computer science, have been made to develop more advanced tools to visualise flow information in various standalone applications (see for example, Phan et al. 2005; Guo 2009; Boyandin et al. 2010; Boyandin et al. 2011).

The flow map, on its own, only conveys pertinent information of an origin-destination matrix where the objective of the investigation is to study the matrix in isolation. In circumstances where there is a need to investigate the extent to which the origin-destination matrix changes as a function of other variables (for example, under certain weather conditions), another technique must then be employed. In this paper, we progress a promising new technique, the flow-comap. The flow-comap is the product of embedding the flow map into another spatial exploratory data analysis technique, the comap.

The comap (Brunsdon 2001) and its older, more generic cousin, the coplot (Cleveland 1994), provide an effective means of visualising multivariate data in order to assist with uncovering previously hidden patterns embedded within complex data. The basic idea is that a bivariate subset of raw data is selected based upon some condition (for example, a defined range of rainfall or a categorical variable such as weekend or weekday). This data is then plotted either in raw form in a scatter plot panel or a mapped kernel density surface. The comap and coplot embrace the small multiple principle (Tufte 1991) by using multiple panes of similar and typically overlapping regions, in order to illustrate how gradual changes can be observed as a function of external variables.

Embedding the flow map into the comap to form a technique we term the flow-comap, represents a straightforward extension of the two techniques which allows multivariate exploration of origin-destination data which otherwise would not be readily possible.

5 Results

We use flow-comap to visualize the spatial flows of CityCycle at the suburb scale. As illustrated by Figure 2, the number of trips between suburbs represented using a variable line thickness, where the width of the line is proportional to the total flow volumes. The choropleth classification of suburbs represents the within-suburb flows (i.e. where the bicycle release and return are in the same suburb), with the darker colour representing the higher number of internal trips. The GINI and CV indexes are calculated for each of map panel to quantify the relative spatial concentration of suburb-to-suburb flows for a given time window.

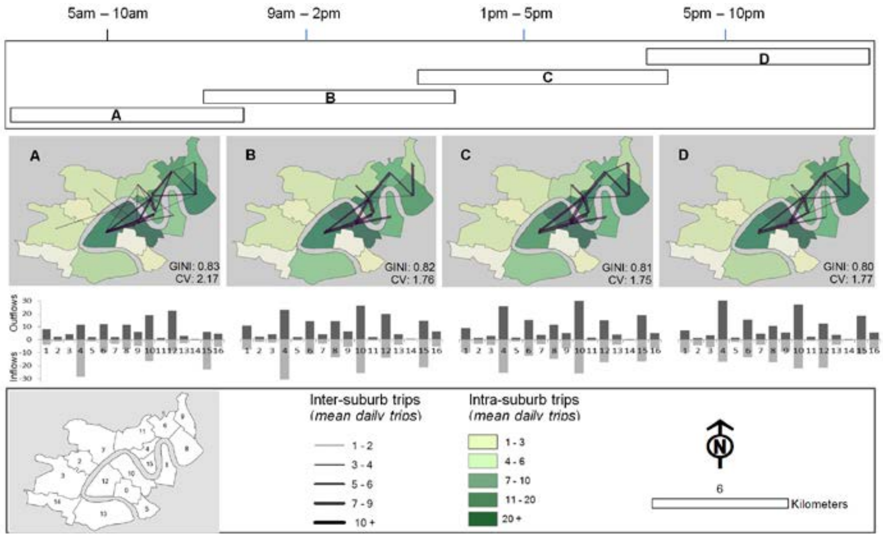


Fig. 2 Univariate flow-comap for public bicycle trips (by hour).

Firstly, we examine the spatial flows over a 24-hour period using a univariate flow-comap (Figure 2). Results highlight that early morning trips (between 5am and 10am) are more spatially dispersed. Later on in the morning and early afternoon (from between 10am and 2pm), trips tend to be more concentrated and spatially focussed around the CBD and the immediate surrounding suburbs. This spatial flow pattern continues until around 5pm where after (from 5pm to 10pm) trips begin to spread further into the suburbs but appear less dispersed than during the morning peak hours. There is also evidence that there are a relatively high proportion of self-contained trips that remain relatively constant across the 24-hour period at some suburbs.

Next, we examine the effect of specific calendar events (i.e. weekdays, weekends, public and school holidays) and hour on trip patterns simultaneously using a bivariate flow-comap (Figure 3). A number of specific observations can be made. First, CBD-based trips taking place during weekends, especially between the hours of 9am to 5pm are shown to be markedly less concentrated that those occurring during workdays. Trips taking place during the evening (i.e. after 5pm) show a significant reduction in weekdays that is not as marked during weekends. The effect of public holidays on the spatio-temporal patterns is very similar to weekend patterns. Trips occurring during school holidays differ very little from regular weekdays apart from a small reduction in the number of trips taking place between the suburbs and the CBD during peak hours.

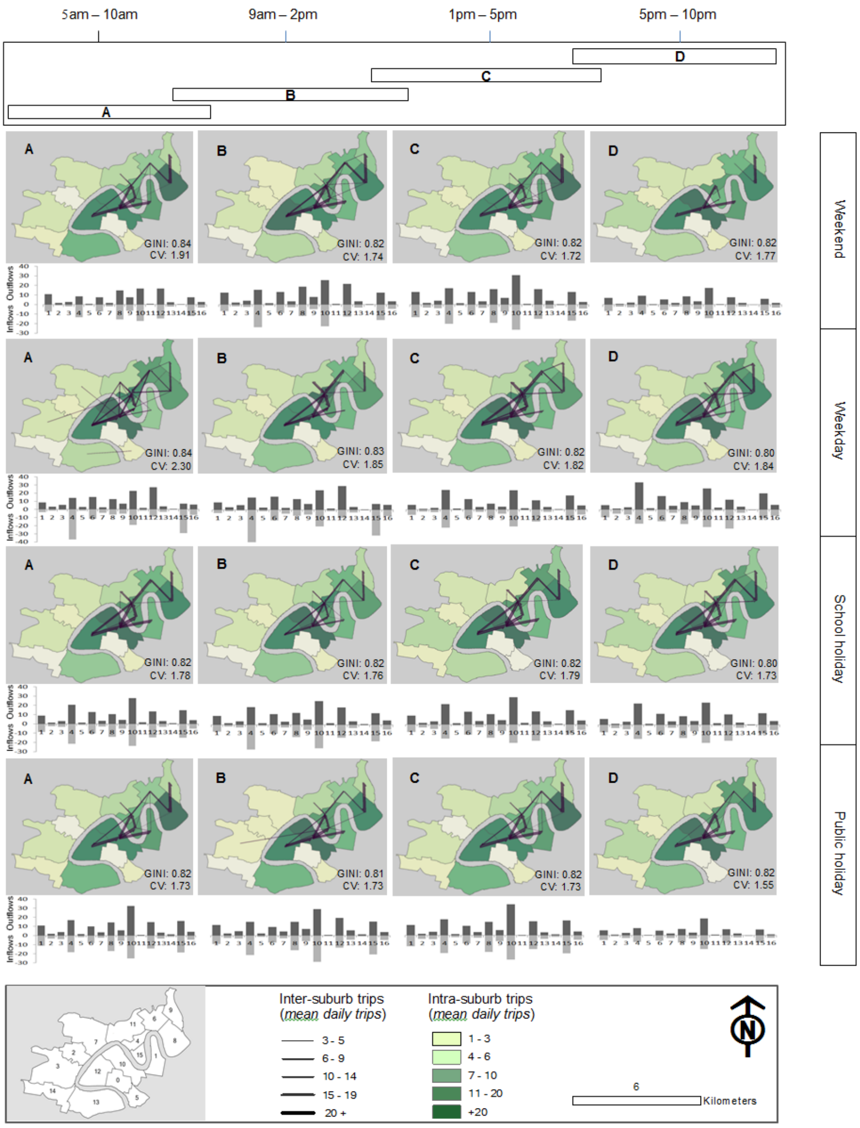
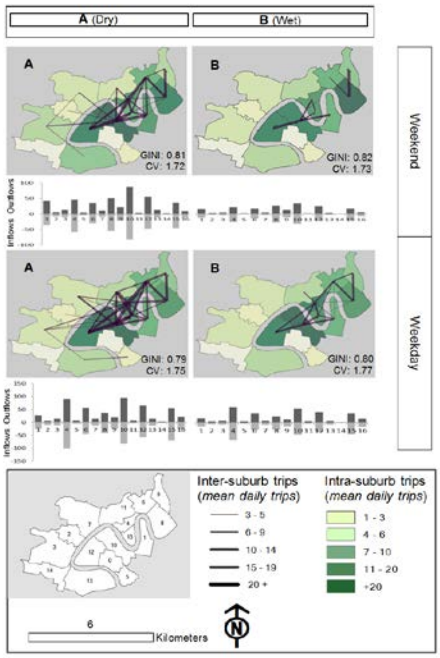
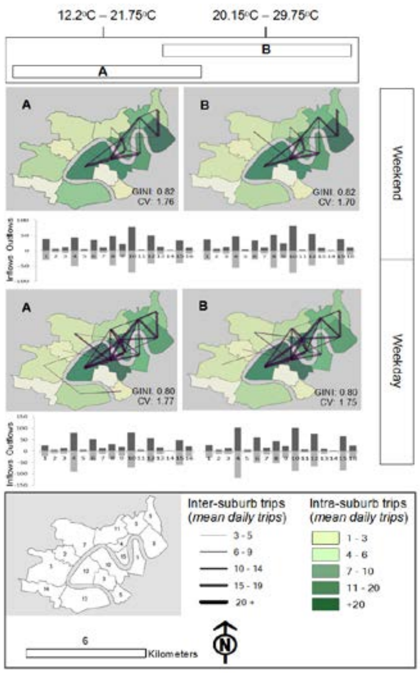


Fig. 3 Bivariate flow-comap for public bicycle trips (by hour of day and for specific calendar events).

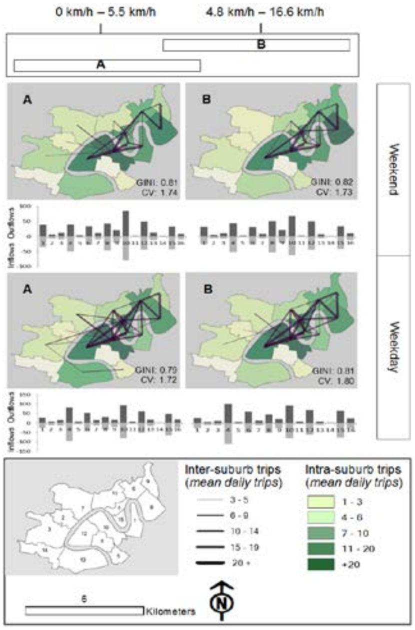
Finally, we examine the relationship between weather (i.e. rainfall, temperature and wind) and day of the week on trip patterns simultaneously using a bivariate flow-comap (Figure 4). Classifying each of the weather variables in two categories (for example, with rainfall into dry and wet) permits the investigation of the weather extremes on the location and timing of trips. Findings suggest that weather conditions exert certain influences that are highlighted by differences in the spatio-temporal patterning of trips. A number of specific observations can be made. First, there is clear evidence to suggest that rainfall affects trips, especially during the weekend, where there is a general system-wide decrease in the number of trips taking place. That said, there are a number of relatively short trips that persist during wet weather, especially between and within West End, South Bank, New Farm and Teneriffe. Wind was also shown to exert a notable impact on tips for both weekdays and weekends during which strong winds (in excess of 5.5 kilometres per hour) considerably reduced the number of longer distance trips. Compared with both rainfall and wind, temperature appears to exert much less of an influence on trips both during the weekday and weekend possibly highlighting the relatively consistent year-round sub-tropical climate in Brisbane.



(a) rain



(b) temperature



(c) wind speed

Fig. 4 (a)(b)(c): Bivariate flow-comap for public bicycle trips (by weather and specific calendar events)

6 Discussions

Developing an understanding of the complex spatio-temporal dynamics of PBSPs is critical to compiling an ongoing evidence base with the capacity to ensure the system is configured in a manner that meets the needs of the users. In addition, this evidence base has the potential to offer prospective insight into how the system might be modified (i.e. re-configured that might involve a variation on how the stations and bicycles are spatially distributed across the network) to better meet the changing needs of the user base. The data necessary to establish such an evidence base already exists (albeit disparately) in the form of disaggregate trip-level records (PBSP trips), the meteorological office (weather data) and the state Department of Education (calendar events). The challenge now for PBSP operators and researchers is to draw upon existing data, and where necessary, develop new tools and techniques to examine these data in a manner that has the potential to improve the use and uptake of these systems worldwide.

The first step to developing this evidence base is the testing and development of suitable tools with the capacity to quantify the salient factors influencing the PBSP dynamics. This paper has attempted to progress research in this area through the assembly of a new integrated database and development of the flow-comap to explore these new data. Results from our analysis highlight that at the system-wide level rain and wind are both significantly related to the number of trips taken. Stronger winds and rainfall significantly reduce the total number of trips taking place. These findings concur with previous studies that have examined the effect of weather conditions on cycling (see for example, Nankervis 1999; Brandenburg et al. 2007; Parkin et al. 2008; Heinen et al. 2011) suggesting that inclement weather conditions are significant detractors for both recreational and commuting bicycle trips. Temperature was found not to be significantly related to the number of trips taken, unlike the evidence presented in other studies including Brandenburg et al. (2007). However, our findings are not surprising given the relatively small temperature variations experienced in Brisbane’s sub-tropical climate, which are quite unlike the weather conditions reported in Brandenburg’s survey of Vienna-based cyclists. Here, temperatures were reported to vary between below minus 4 degrees Celsius to in excess of 40 degrees Celsius. In regards to the effect of calendar events, both public and school holidays were not found to exert a significant influence at the system-wide level. This differs from other studies on cycling (for example, Brandenburg et al. 2007) suggesting that these types of calendar events help to explain variations in both the number of recreational and commuter bicycle trips reflecting changes in the routine activities of individuals and families.

Examining the spatial flows in conduction with weather conditions highlighted that both rain and wind (specifically stronger winds) reduce the frequency of trips taking place, and in particular the longer trips. These findings align closely with existing research (for example, Heinen et al. 2011) suggesting that cyclists have limited capacity to protect themselves from inclement weather when compared to other transport modes such that poor weather has a significant impact on the number and length of trips undertaken. Despite showing no effect at the system-wide level, calendar events (in particular public holidays) were shown to exert some subtle variation in the spatial distribution of trips. The CBD experienced a reduction in the number of trips to and from the immediately surrounding suburbs, pointing to the adjustment in individuals’ routine activities during holiday periods.

There are at least two areas that could form the basis for follow-up research. First, based upon our case study presented in this paper, we advocate the integration of the flow-comap as a tool within a standard GIS environment. The various flow maps produced here have been generated individually using the "FlowMap" package (Guo 2009) (one flow map at a time), and then integrated into a final output using graphic design software. The development of a GIS-based equivalent would greatly facilitate their efficient generation and subsequent spatial exploration. Embedded within this tool should be the capacity to automate the generalization of the spatial unit such that the user can readily adjust the unit of analysis. In this case study, it was necessary to fix the spatial unit of analysis, i.e. the suburb. However a tool with the ability to adjust the spatial unit to either smaller or larger units on-the-fly would greatly assist in assessing the spatial scale and degree to which weather and calendar events impact the spatio-temporal dynamics of the PBSP. A second avenue for future research is concerned with broadening the scope of this current study so as to assess different situational and climatic contexts and to draw comparable data in order to establish if the reported findings are replicated in other PBSPs around the world.

7 Conclusions

PBSPs have rapidly risen as a new alternative transport mode in many cities worldwide. Despite their rapid growth, there has been relatively limited research that has explored their underlying spatio-temporal dynamics using trip level data. In most of the previous research into PBSPs data capturing, the stock data has been the predominant source of data to study their underlying dynamics. In this study, we have highlighted the utility of flow or trip-level data that offers new opportunities for research to examine the spatio-temporal dynamics of PBSPs.

In this paper, we have attempted to progress the previous work on PBSPs in two ways. First, we highlighted the utility of integrating three formerly disparate databases describing each bicycle trip along with databases of calendar events and weather conditions to provide the necessary foundation from which the underlying dynamics of the PBSP can be examined. Second, has been the development of a new visual analytic, the flow-comap, where we have demonstrated its potential to draw on an integrated database to uncover new insights into the association of bicycle trip patterns to various calendar events, and under certain weather conditions. Gaining a better understanding of these underlying dynamics is a first step to establishing a fully automated monitoring tool with the capacity to identify expected movements of bicycles around the system, given certain temporal and environmental circumstances. The analysis presented here has extended our knowledge of PBSP dynamics through adopting a quantitative approach to examine disaggregate-level relationships between bicycle trips and their complex relationships with weather and calendar events. Whilst findings reported in this paper are important for managers of PBSPs, we also recognise that more in-depth research is now needed to further explore the reasons for such trends. In conducting such research could in turn lead to operational benefits for PBSPs including the way in which bicycles are distributed across the system on wet weekends and windy weekdays to ensure that supply and demand are met in the most optimal manner possible.

References

Anselin, L. (1999). Interactive techniques and exploratory spatial data analysis. Geographical Information Systems: principles, techniques, management and applications, 1, 251-264.

Bell, M., Blake, M., Boyle, P., Duke-Williams, O., Rees, P., Stillwell, J. and Hugo, G. (2002). Cross-national comparison of internal migration: issues and measures. Journal of the Royal Statistical Society A, 165(3), 435-64.

Black, C., and Potter, S., n.d. Portsmouth bikeabout: A smart-card bike club Scheme. Accessed 14 August 2013, http://www.metrobike.net/index.php?s=file_download&id=11.

Boyandin, I., Bertini, E., and Lalanne, D. (2010). Using flow maps to explore migrations over time. Proceedings of Geospatial Visual Analytics Workshop in conjunction with The 13th AGILE International Conference on Geographic Information Science 2(3). https://diuf.unifr.ch/main/diva/sites/diuf.unifr.ch.main.diva/files/jflowmap-geova10.pdf

Boyandin, I., Bertini, E., Bak, P., and Lalanne, D. (2011). Flowstrates: An Approach for Visual Exploration of Temporal Origin - Destination Data. Computer Graphics Forum, 30(3), 971-980.

Brandenburg, C., Matzarakis, A., and Arnberger, A. (2007). Weather and cycling—a first approach to the effects of weather conditions on cycling. Meteorological applications, 14(1), 61-67.

Brunsdon, C. (2001). The comap: exploring spatial pattern via conditional distributions. Computers, environment and urban systems, 25 (1), 53-68.

Cleveland, W.S. (1994). Coplots, nonparametric regression, and conditionally parametric fits. Lecture Notes-Monograph Series, 21-36.

DeMaio, P. (2009). Bike-sharing: History, impacts, models of provision, and future. Journal of Public Transportation, 12(4), 41-56.

Department of Education, Training and the Arts (2013). School holidays calendar. Accessed 14 August 2013, http://education.qld.gov.au/public_media/calendar/holidays.html.

Glennon, A. and Goodchild, M. (2005). A GIS Flow Data Model. Flow White Paper. Accessed 14 August 2013, http://dynamicgeography.ou.edu/flow/.

Guo, D., 2009. Flow mapping and multivariate visualization of large spatial interaction data. Visualization and Computer Graphics, IEEE Transactions on, 15(6), 1041-1048.

Haining, R., Wise, S., & Ma, J., 1998. Exploratory spatial data analysis. Journal of the Royal Statistical Society: Series D (The Statistician), 47(3), 457-469.

Heinen, E., Maat, K., & Van Wee, B., 2011. Day-to-day choice to commute or not by bicycle. Transportation Research Record: Journal of the Transportation Research Board, 2230(1), 9-18.

Henley, J. (2005). Rentabike moves up a gear from curiosity to runaway success. The Guardian. August 12. http://www.guardian.co.uk/world/2005/aug/12/france.jonhenley/

Jensen, P., Rouquier, J.-B., Ovtracht, N., and Robardet, C. (2010). Characterizing the speed and paths of shared bicycles in Lyon. Transportation Research Part D: Transport and Environment. 15(8), 522-524.

Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., and Banchs, R. (2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. Pervasive and Mobile Computing, 6(4), 455-466.

Minard, C.J. (1869). Carte figurative des pertes successives en home de l'Armée Fançaise dans la campagne de Russie 1812-1813. Paris: Réginer et Dourdet.

Nair, R., Miller-Hooks, E., Hampshire, R. C., & Bušić, A., 2013. Large-Scale Vehicle Sharing Systems: Analysis of Vélib'. International Journal of Sustainable Transportation, 7(1), 85-106.

Nankervis, M. (1999). The effect of weather and climate on bicycle commuting. Transportation Research Part A: Policy and Practice, 33(6), 417-431.

Parkin, J., Wardman, M., and Page, M. (2008). Estimation of the determinants of bicycle mode share for the journey to work using census data. Transportation, 35(1), 93 109.

Phan, D., Xiao, L., Yeh, R., and Hanrahan, P. (2005). Flow map layout. Information Visualization, 2005. INFOVIS 2005. IEEE Symposium, 219-224. IEEE.

Thompson, W., and Lavin, S. (1996). Automatic generation of animated migration maps. Cartographica: The International Journal for Geographic Information and Geovisualization, 33(2), 17-28.

Tufte, E.R. (1991). Envisioning information. Optometry & Vision Science, 68(4), 322-324.