THE IMPACT OF SPATIAL ENABLEMENT AND VISUALISATION ON BUSINESS ENTERPRISE DATABASES - WHAT YOUR DATA HAVE BEEN TRYING TO TELL YOU

Yiqun Chen¹, Abbas Rajabifard², Geoff Spring³, Judy Gouldbourn⁴, Gerald Francis Griffin⁵, David John Williams⁶

^{1,2,3,5.6}The University of Melbourne, Melbourne, Australia

⁴Volunteer Fire Brigades Victoria

¹yiqun.c@unimelb.edu.au, ²abbas.r@unimelb.edu.au, ³geoff.spring@unimelb.edu.au, ⁴j.gouldbourn@vfbv.com.au, ⁵gerarld.griffin@unimelb.edu.au, ⁶david.williams2@unimelb.edu.au

Abstract

Historically business enterprises have been gathering data as part of their "business as usual" operations. The evolution of the digital era has both enhanced this capability and increased the rate at which data is collected at unprecedented levels. The parallel evolution of spatially enabled data, data analytics and the visualisation of data presents opportunities to analyse spatio-temporal databases to a degree never before available. This ability provides the opportunity to incorporate the results of this analysis into corporate planning processes, policy and strategy development and risk identification and mitigation. However, this new capability may also identify deficiencies in historically utilised databases which have led to poor decision making and setting of policy and strategy that has unknowingly limited business performance, misdirected capital investment and impacted resource utilisation.

This paper will address these issues by understanding the concept of "concurrency" in database visualisation via a spatially enabled decision support tool developed by the Centre for Disaster Management and Public Safety (CDMPS), the University of Melbourne. A special case study is performed to analyse historic incidents and explore response capacities across Victoria. A snapshot of emergency management data has been subjected to data cleaning, aggregation and harmonisation processes to support our proposed analysis methodology. The output identifies key

components such as demands and supplies. Each of these components can be investigated at various temporal granularity levels such as daily, monthly and yearly. Besides statistics, the developed tool can also interactively manipulate the results on a 4D visualisation engine by using dynamic demand-supply heat maps and spider webs that precisely describe the concurrent characteristics. The developed system helps decision makers better understand when and where demands are triggered and how supplies are distributed in busy seasons and eventually identify research priority needs to enhance their workforce planning capability.

Keywords: spatio-temporal data analysis, disaster management and decision making, 4D visualisation.

1. INTRODUCTION

Disaster and emergency management is a global practice the need for which is rapidly growing to deal with events caused by environmental and or human intervention. The quality of the outcomes from the management of individual or multiple emergencies with their potential to scale to major disasters relies on the quality of the decision making processes underlying the preparation for, response to and the recovery from these events. Traditionally these processes have relied on personal experience and training is supplemented with information from disparate databases that is often not readily accessible.

The emergence of today's and the future's digital and IP based environment will continue to significantly change this situation so that the quality of outcomes and their associated decision making processes can be transparently examined, evaluated and improved using spatially enabled data, data analytics and visualisation tools. Importantly the application of these tools to historic databases should be expected to expose deficiencies in the data and data collection processes which will need rectification, particularly where these databases inform policy decisions and strategy development associated with the allocation and use of high-value discreet resources. A big challenge here is how to reveal the characteristics of spatio-temporal enabled data in an interactive and intuitive fashion (CDMPS, 2014; Rajabifard et al., 2015). It requires sound knowledge of data, articulated analysis methods and elaborate visualisation means.

This paper provides information about the methodology used to produce a set of tools for spatio-temporal database analysis and visualisation, especially to facilitate decision making process. The Centre for Disaster Management and Public Safety (CDMPS) in the University of Melbourne was engaged by Volunteer Fire Brigades

Victoria (VFBV) to develop a suite of such tools that could be applied to a sample from the Fire Incident Reporting System (FIRS) developed and used by the Country Fire Authority (CFA) in Australia to record data resulting from use of Fire Brigades in the management of major bushfires in the State of Victoria in Australia. The developed tools may be applied to the management of high-value discrete resources i.e. the lives of those involved in public safety activities that underpin community resilience to emergencies and disasters.

2. METHODOLOGY

We start a spatio-temporal analysis from tackling the temporal attributes. Usually, temporal values are treated as continuous variables; therefore, attributes directly attached to each time frame could be too sparse to form patterns. This indicates that a time window is required, within which non-temporal attributes can be aggregated (or accumulated) so that their characteristics are more identifiable. Here the concept of "concurrency" is introduced to formulate a universal analysis framework.

Given two timestamps t0 and t1 on the time axis **T**, where t0 < t1; then a time window is constructed and denoted as [t0, t1]. An event has a start time and end time denoted as st and et respectively, where st < et, then event lifespan LS can be denoted as [st, et] Figure 1 illustrates all six possible relationships between LS and **TW**, only four of them (green lines) can be considered as concurrent (or "ongoing") events within **TW**, they satisfy either of following criteria:

(1) st < t0 and et > t1
(2) st < t0 and et < t1
(3) st > t0 and et < t1
(4) st > t0 and et > t1

The criteria can be further simplified as:

(5) st < t0 and et > t0
(6) st > t0 and st < t1



Figure 1 Time window and concurrent events. Four green lines can be considered as concurrent events within time window [t0, t1].

Thus, given a time window [t0, t1], all concurrent events within that period can be selected out by adopting either criterion (5) or (6). For a long LS incident, it contributes the count to each TW it crosses. For example, the top green line in Figure 1 represents an incident that starts before t0 and ends after t1. This incident's LS covers two additional TWs before and after [t0, t1], therefore, it will be counted as a concurrent incident for each of three TWs. This process is a discretisation of continuous variables and can accurately measure the number of concurrent incidents in a given time window. The width of time window controls the granularity of discretisation, which also affects further spatio-temporal pattern analysis. There is no ubiquitous rule for picking granularity values. The decision highly relates to the temporal features of data as well as the goal of analysis. Setting an "impropriate" granularity might hinder the pattern identification process or impact the performance of analysis. In this work, to help VFBV identify temporal patterns for incident occurrence and resources allocation, three granularity levels (i.e., day, month and year) were selected to describe time windows. The raw FIRS data were also aggregated at these three levels respectively to speed up answering questions such as "which day is the busiest day in 2015" or "how the amount of distributed resources fluctuates in every January during the past 10 years".

As for spatial analysis, besides classic methods such as clustering (Diggle, 2013, Sturup et al., 2015), autocorrelation (Griffith, 2013), regression (Fortin et al., 2012) etc., this part can be enhanced by using advanced 4D spatial visualisation techs (CDMPS, 2014), which are regarded as a highly effective and intuitive way to recognise and understand spatial patterns. Again, there is no best visualisation means for all cases, the selection of various visualisation means is subjected to different purposes. Each visualisation means has its own merits as well as disadvantages; therefore, when interpreting the outputs, we should be aware of their limits and pitfalls.

In following sections, we will take a real world database as an example, and demonstrate how to identify and interpret the spatio-temporal patterns by adopting proposed methodology.

3. SAMPLE DATABASE – FIRE INCIDENT REPORT SYSTEM

Victoria is a state in the south-east of Australia and it is one of the most bushfire prone areas in the world (Jones, 2011). Since the 1850's community volunteers have come together to establish volunteer fire brigades and this led to the establishment of the Country Fire Brigades in January 1891 which was monitored by the Country Fire Brigades Board. Over the next fifty years, Victoria suffered a number of devastating bushfires including the 1926 Gippsland fire (60 fatalities), 1939 'Black Friday' bushfire in Narbethong (71 fatalities) and 1943/44 state-wide fires (51 fatalities and 700 injured) (Jones, 2011). The Stretton Royal Commission was established to conduct inquiries into the management of bushfires and this led to the establishment of the Country Fire Authority (CFA) on 2 April 1945 (Jones, 2011). Since this time the CFA has grown to become one of the largest volunteer and community-based emergency service organisations in the world. The CFA manages over \$240 million worth of assets and has an annual income of over \$500 million (CFA, 2014).

One of the most valuable assets of CFA Victoria is the historical incident reports recorded in FIRS (Fire Incident Report System). Back to 1990, CFA Victoria had used FIRS to manage incident related information. It is a giant database and has accumulated over 1 million records of incident and brigade logs over decades. For incidents (i.e., demands), FIRS contains key information such as when and where they occur and how long they lasts; for brigade (i.e., supplies), FIRS records where they are, how many resources (e.g., trucks and personnel) are taken to incidents as well as how long the support last. FIRS is a classic spatio-temporal database and well suits for our proposed concurrency analysis methodology.

3.1. Data Structure

Five critical data tables from FIRS are investigated in this paper for concurrency analysis. The entity relation diagram is shown in Figure 2 with only key attributes listed. FIRS has evolved over years and its database schema was also adjusted for several times to support system compatibilities. Though its structure becomes obscure to understand, it won't impede the data analysis process.

3.1.1. Table firs_primary_report_header

This table contains key information for an incident as well as its primary brigade, such as incident start time (*incident_datetime*), end time (*stop_datetime*), location (*geom*), its primary brigade id (*brigade_no*) and when its primary brigade is notified (*brigade_advised_datetime*).

3.1.2. Table firs_support_report_header

This table contains information for all support brigades of incidents, such as support brigade id (*brigade_no*), when a support brigade is notified (*brigade_advised_datetime*).It links to the incident table using foreign key (*primary_report_id*).

3.1.3. Table firs_brigade

This table contains information of brigades, such as name (*brigade_name*) and location (*geom*).

3.1.4. Table firs_report_resource

This table contains trucks dispatch logs for both primary and support brigades. It uses *report_type_flag* attribute ('P' for primary and 'S' for support) to differentiate which report header table its *report_id* should reference to. It also contains the timestamp for when a truck is sent out (*mobile_datetime*) and when it returns (*in_station_datetime*).

3.1.5. Table firs_report_personnel

This table contains people dispatch logs for both primary and support brigades. Similar to table *firs_report_resource*, it also uses *report_type_flag* attribute ('P' for primary and 'S' for support) to differentiate which report header table its *report_id* should reference to. However, it does not contain timestamps for when a person is sent out and when he returns.

Figure 2 Entity relation diagram for FIRS key tables. Only key attributes for each table are listed.



3.2. Data Preparation

For this research, CFA provided us with a database backup file containing the five selected data tables. It comprises all incident and response logs recorded between July 1999 and April 2016. Current FIRS utilises MSSQLServer as its database which lacks spatial data structure, storage and analysis support and is not useful for our work. PostgreSQL is an open-sourced, object-relational database system and with its PostGIS plugin, it supports for geographic objects (e.g., point, line, polygon primitives) and can be used as a spatial database for geographic information systems (GIS). We set up a PostgreSQL server on the NecTAR Research Cloud and used the following steps to port data from MSSQLServer to PostgreSQL:

- 1. Restore backup file to an instance of MSSQLServer
- 2. Create functions to dump data into PostgreSQL INSERT script files
- 3. Create data table schemas in PostgreSQL
- 4. Execute INSERT script files
- 5. Create point geometries based on values
- 6. Create (spatial) index for each table

From 1999 to 2016, FIRS has collected 688,910 incident reports attached with over 1.73 million trucks and 8.28 million personnel allocation logs respectively. With the assistance from the FIRS database admin from CFA, attributes incident_datetime and stop_datetime in data table firs_primary_report_header are selected to identify incident start time (st) and end time (et). However, a further data check shows that about 8% incident records have invalid stop datetime value, either it is an empty value or it contains a time before *incident datetime*. The data quality problem is most likely caused when manually logging information into FIRS. Though the proportion of issued records is relatively low, it does affect the understanding and judgment of surge analysis outputs, particularly when the visualised map. To overcome this, results are on we used the last truck return time (ltrt) of an incident to correct its end time if its original value is problematic. After the correction process, 99.3% incident records have valid st and et values and are ready for surge analysis.

Similarly, for the truck operation logs, we selected *mobile_datetime* (st) and *in_station_datetime* (et) from data table *firs_report_resource* to measure the period a truck committed to supporting an incident. However, FIRS does not specifically record the time involvement for personnel. It cannot identify when a CFA employee (or volunteer) is devoted to an incident and when he retreats from it. To measure the personnel allocation over time, we assume that all FIRS recorded personnel from a brigade are allocated to an incident from *brigade_advised_datetime* (st) of brigade till *stop_datetime* (et) of the incident. This might overestimate the actual number of involved personnel at a given time period

at brigade level, but certainly, gives a right indication of how personnel resources are allocated during the entire incident lifespan.

3.3. Data Aggregation

The main purpose of data aggregation in this work is to avoid complex database queries that will be repetitively executed in the application runtime. Aggregation data will also boost the performance of surge analysis.

By using the concurrency concept described in the previous section, nine aggregation data tables are created by crossing three aspects (i.e., incident, truck and personnel) in surge analysis with three granularity levels (i.e., year, month and day), as Table 1 shows.

	Year (18)	Month (204)	Day (6210)
Incident	aggr_incident_y	aggr_incident_m	aggr_incident_d
Truck	aggr_truck_y	aggr_truck_m	aggr_truck_d
Personnel	aggr_personnel_y	aggr_personnel_m	aggr_personnel_d

Table 1 Nine FIRS aggregation tables. 18, 204 and 6210 records will be created ineach table if it is aggregated by year, month and day respectively.

All aggregated tables share similar data structure with three key columns *st*, *et* and *total_num*. Surge peak analysis in section 4.1 will directly come from these tables.

3.4. Data Visualisation

Besides charts and statistics, it is imperative but more challenging to reveal the spatio-temporal patterns on the map to decision makers in an interactive and intuitive fashion. In this work, we introduce dynamic heatmaps, demand-supply lines and a set of queryable map markers to visualise surge interactively on a 4D map.

3.4.1. Incident (Demand) Heatmap

Incident heatmaps are ideal to illustrate spatial distribution patterns of incidents' concurrency intensity. It becomes a particularly useful when a set of time serial heatmap frames are stitched together so that the continuous change patterns can be observed. To build a single frame of the incident heatmap, the algorithm takes incident location and intensity aggregated at a given time period as inputs. The location of the incident is described by its longitude-latitude coordinate, which will

affect the heat cores' location. The intensity value of incident can be designated either as the total number of trucks or as the total number of personnel (default option). This parameter affects the coverage radius of heat cores.

3.4.2. Resource (Supply) Heatmap

Similar to incident heatmaps, resource heatmaps depict the spatial distribution patterns of concurrent allocation of resources. Unlike incident heatmaps which are created by using incident parameters, resource heatmaps are built upon brigade parameters including its location and the amount of its committed resources (the total number of trucks or personnel).

3.4.3. Demand-Supply Line

Heatmap offers an overview of the changing patterns of demand-supply spatial distribution over time; however it fails to reveal the relationships and details of demand-supply at a finer level. For example, by using heatmap only, we cannot tell where exactly the resource supplies of a single incident come from, nor can we tell which incidents a brigade is supporting simultaneously. Therefore, we introduce demand-supply lines to illustrate this missing nexus.

As its name indicates, a demand-supply line connects an incident to all brigades that send resources to it. From a brigade's standpoint of view, this line also shows all incidents that it is committed to at the same time. The demand-supply lines form a spider-web-like structure and reveal complex linkage patterns between incidents and brigades. In this work, we use the line width to illustrate the strength of the connection, which can be presented as the total number of trucks or personnel.

3.4.4. Queryable Map Markers

To make visualisation self-explained, we introduce a series queryable map markers, they are summarised in Table 2. If a brigade responds to multiple incidents simultaneously, multiple brigade markers will be created around the real location of the brigade so that demand-supply lines can be connected individually.

Markers	Description
•	Incident marker
	Primary brigade (pink) and support brigade (blue)
	Brigades that only send personnel (P) to an incident
	Brigades that only send truck (T) to an incident

Table 2 Map marker icons and descriptions.

N	Brigades that respond but not (N) send any resources to an
	incident

4. INVESTIGATION ON BLACK SATURDAY BUSHFIRES 2009, VICTORIA

The Black Saturday Bushfires were a series of bushfire started around the 7th of February 2009 in Victoria, Australia. It is the worst bushfire catastrophe in Australia history and killed 173 people, injured 414 people, destroyed 2,100 homes and displaced 7,562 people (BSB, 2014; Victoria Police, 2009). It is estimated the energy released by the Black Saturday Bushfires, was the equivalent of 1,500 Hiroshima atomic bombs and 1.1 million acres were burnt in total (Cameron et al, 2009; BSB, 2014).

Black Saturday provides a good example for investigating surge behaviour. In the following section, we articulate the analysis outputs during the period from 4th to 28th of February 2009.

4.1. Surge Peak Analysis

Figure 3 shows a single peak of concurrent incidents on the 7th of February 2009. On that day, FIRS recorded 407 incidents happened in parallel. The number plunged over the next three days till 10th of February and then fluctuated at 150. For CFA Victoria, the average daily concurrent incident number in busy season (from January to March) is 129 and for the rest of year, the number is 101. Given this, the critical fire situation in February 2009 was clearly a huge challenge to CFA Victoria.

Figure 3 Concurrent incident (blue line) analysis on Black Saturday Bushfires, from 4th to 28th February 2009.



In Figure 4, the concurrent trucks usage is turned on (green line), which tells the same story. The number increased sharply on 7th February to 2035, which is four times of 504 recorded on 6th February. The concurrent truck usage well correlates

with the number of concurrent incidents. For CFA Victoria, the average daily concurrent number of the truck on road in busy season (from January to March) is 367 and for the rest of year, the number is 250. It demonstrates that the CFA were under great pressure of resource allocations during the Black Saturday period.

Figure 4 Concurrent incident (blue line) and truck (green line) analysis on Black Saturday Bushfires, from 4th to 28th February 2009.



In Figure 5, the concurrent number of personnel (red line) allocated is also appended. Although the numbers are distorted due to our assumption, from the people resources perspective again, CFA was also stressed in February 2009. For CFA Victoria, the average daily concurrent number of people allocated in busy season (from January to March) is 4262 and for the rest of year, the number drops to 2051. The red line remains four times above the average number of busy season for two weeks and then follows by two more peaks appeared on 23rd and 27th of February. This indicates how exhausted CFA staffs were in that critical period.

Figure 5 Concurrent incident (blue line), truck (green line) and personnel (red line) analysis on Black Saturday Bushfires, from 4th to 28th February 2009.



4.2. Surge Visualisation

We ran a surge simulation for Black Saturday Bushfires from 4th to 28th February 2009. A series heatmaps were generated to illustrate the spatial distribution patterns of demand and supply over time. Our developed system can fluently and interactively visualise the continuous change patterns and offer an intuitive understanding of the surge behaviours. In this paper, the outputs of four days (4th, 7th, 15th and 23rd) were selected and put together to demonstrate the capability of surge analysis.

Figure 6 shows the incident heatmap over four days. On 4th February, main incidents (red areas on the map) concentrated in the mid of Victoria, but on 7th, the situation deteriorated in the mid and started sprouting to the east region. Isolated but critical incidents were reported in the north-west part as well. One week later on 15th February, the situation cleaned up particularly in the mid and east regions; but on 23rd February, the situations in west and east regions became worse again.

Figure 6 Incident heatmaps on 4th, 7th, 15th and 23rd February 2009. Red areas represent the location of major incidents and blue areas indicate the location of minor incidents.



Figure 7 shows the surge from the resource (i.e. personnel) supply perspective. CFA Victoria has 1724 brigades distributed across the entire state. On 4th February, the significant resources (white areas on the map) were mainly allocated to the mid and east regions; we can see there are widely scattered, insignificant supply dots

in the west and north part of Victoria as well. But the situation changed dramatically on 7th February. The mid, north-east, south-east and south-west were all under the pressure of significant supply. The situation remained intense and did not notably mitigate in next two weeks. This visualisation heatmap is well aligned with the surge peak analysis outputs shown in Figure 5, and it reveals how widely the brigades in Victoria were affected during the disaster.

Figure 7 Resource (personnel) heatmaps on 4th, 7th, 15th and 23rd February 2009. White areas represent where significant supplies locate and blue areas indicate where insignificant supplies locate.



Figure 8 overlaps the demand and supply heatmaps on one picture. It seems that both heatmaps well fit with each other geographically, especially in the mid region. It might give a wrong impression that supplies for incident mainly come from local brigades. Actually, that is far from the truth in Black Saturday.

Figure 8 Overlap incident heatmaps and personnel supply heatmaps on 4th, 7th, 15th and 23rd February 2009.



The demand-supply lines (yellow lines on the map) are drawn in Figure 9, which demonstrates the linkage between incidents (red marker) and brigades (blue/pink flags). On 4th February, resources located in the far west and north region were dispatched to support incidents in the mid and east regions. The long yellow lines indicate that the supplies came from distant brigades rather than local ones. In the following days, more brigades got involved in the north and west part to support both local and remote incidents. Figure 10 gives a detailed view of the demand-supply lines. On the map, a significant incident drew resources from 605 brigades across Victoria; the width of yellow lines stands for the number of people dispatched from each brigade. Quantitative incident information such as start-end time, number of involved brigades, personnel, trucks, average travel distance, etc. is accessible by clicking the incident marker on map. Comprehensive brigade response data including number of dispatched personnel, trucks, attendance time, support duration etc. is presented in a similar way.

Figure 9 Demand-supply lines (yellow) on 4th, 7th, 15th and 23rd February 2009. Table 2 shows the marker legend information.



Figure 10 A detailed view of demand-supply lines (yellow) on 7th February 2009. Table 2 shows the marker legend information

5. CONCLUSION

This paper proposed a new methodology for exploring the concurrency characteristics hidden in spatio-temporal databases. Taking the Fire Incident

Report System (provided by CFA Victoria) as an example, we started from scratch and went through each step of data preparation and data aggregation. Then we successfully developed a 4D online system and performed surge analysis for the Black Saturday Bushfires in 2009. The analysis results and visualisations intuitively reveal the details about how the incidents and resources were managed during the fatal calamity in Australia. It also strongly demonstrates the capabilities, effectiveness and potentials of our developed system as a spatially enabled decision support tool for stakeholders.

6. REFERENCES

- Bivand, R.S., E. Pebesma and V. Gómez-Rubio (2013). *Hello World: Introducing Spatial Data. In Applied Spatial Data Analysis with R.* Springer New York. pp. 1-16.
- BSB Black Saturday Bushfire (2014). A brief introduction website of Black Saturday Bushfire, at <u>http://www.blacksaturdaybushfires.com.au/</u> [accessed 5 May 2016]
- CDMPS Centre for Disaster Management and Publich Safety (2014). Intelligent Disaster Decision Support System (IDDSS) Project Technical Report, at <u>http://www.cdmps.org.au/download/iddss/IDDSS_Project_Technical_Rep</u> <u>ort.pdf</u> [accessed 11 June 2016]
- CFA Country Fire Authority (2014), Foundations of the CFA: Capstone, Country Fire Authority, Melbourne Australia.
- Day, J., Y. Chen, P. Ellis and M. Roberts (2016). A Free, Open-Source Tool for Identifying Urban Agglomerations using Point Data, Spatial Economic Analysis, 11(1), 67-91.
- Diggle, P.J. (2013). Statistical analysis of spatial and spatio-temporal point patterns. CRC Press.
- Fortin, M.J., P.M. James, A. MacKenzie, S.J. Melles & B. Rayfield (2012). Spatial statistics, spatial regression, and graph theory in ecology. *Spatial Statistics*, 1, 100-109.
- Griffith, D.A. (2013). Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization. Springer Science & Business Media.
- Jones, D.J. (2011), Report of inquiry into the effect of arrangements made by the Country fire Authority on its volunteers, Country Fire Authority, Melbourne Australia, at <u>http://www.cfa.vic.gov.au/fm_files/attachments/volunteer_and</u> <u>career/CFA_Inquiry_Report_and_Appendices-combined.pdf</u> [accessed 7 June 2016].

- Rajabifard, A,. R.G. Thompson and Y. Chen (2015). An intelligent disaster decision support system for increasing the sustainability of transport networks, *Natural Resources Forum*, 39, 83-96.
- Sturup, S., J. Day, Y. Chen, A. Wu, M. Budahazy (2015). Clustering Policy Effects on Industrial Agglomerations: A Spatial Study of Industrial Growth in Melbourne, *State of Australian Cities Conference (SOAC)*, Gold Coast, Australia, 9-11 December, 2015.
- Victoria Police (2009). Press conference: Bushfires death toll revised to 173, at <u>http://www.police.vic.gov.au/content.asp?Document ID=20350</u> [accessed 5 May 2016]