

Inferring commercial vehicle activities from GPS data.

Johan W. Joubert

STRC 2009

June 2009

STRC 2009

Inferring commercial vehicle activities from GPS data.

Johan W. Joubert
Industrial and Systems Engi-
neering
University of Pretoria
0002, Pretoria, South Africa
phone: +41 44 633 68 65
fax: +41 44 633 10 57
johan.joubert@up.ac.za

June 2009

Abstract

To address the underreporting of freight from a transport geography point of view, we present a descriptive paper of the time and spatial characteristics of disaggregated commercial vehicle activities. The activities were extracted from raw GPS data collected over a six-month period for more than 30,000 commercial vehicles. We provide a comparison of commercial activity metrics for the small province of Gauteng in South Africa. The province represents less than 1.5% of the country's land surface, yet produces in excess of 30% of the gross domestic product of the country. In response to the South African government's intent to better understand freight movement at a detailed, non-macroeconomic level, this paper is significant as it provides a means to support government's transport infrastructure investment decisions. Although the activity densities contradict some accepted expectations found in literature, it highlights some of the impacts of past political inequalities still evident in South Africa, again providing value inputs to policy making.

Although the study has been conducted within a South African context, the approach of extracting commercial activities will be relevant to an international community of researchers.

Keywords

GPS data processing; Commercial transport; Activity density

1 Introduction

As participants in the economy, we take part in various activities such as work, residential, leisure, shopping, and many others. These activities are separated in both space and time, implying that some form of transport will be required. But economic activities also imply the provision, distribution, and consumption of goods and services. These too are connected through different means of transport.

Hesse and Rodrigue (2004) argue that freight transport is not merely a derived demand. Logistics are no longer just means to overcome space, but has a critical time component. The extension of supply chain structures, outsourced and subcontracted across vast geographic areas have resulted in smaller consignments delivered more frequently. As supply chains increase in complexity, activities of production, distribution and consumption are increasingly difficult to separate.

The objective of this paper is to report on an exercise in which we aim to better understand the movements of commercial vehicles. In processing GPS points collected for more than 30,000 vehicles over a six month period, we extract and evaluate activity locations, densities of the activities, and characteristics of the chains of activities that vehicles follow. GPS studies about people movement and their activities abound. Consider, for example, the works of Yalamanchili *et al.* (1999); Draijer *et al.* (2000); Wolf *et al.* (2001, 2004); Marchal *et al.* (2005); Tsui and Shalaby (2006); Schüssler and Axhausen (2009).

Ambrosini and Routhier (2004) review urban freight modelling efforts across nine industrialized countries. The methods and tools included theoretical models, interviews, surveys, and experiments. In response to the review, Figliozzi (2007) notes that studies concerning vehicle activity chains are absent. Although still based on zonal aggregates, Friedrich *et al.* (2007) report on recent models where activity chain generation has been accounted for. As with other top-down multiclass assignment models based on the four step modelling approach, the aggregate models lack the ability to capture individual stakeholders' behaviour.

In this paper we address that lack of understanding of commercial vehicle chains. In Section 2 we explain our process of extracting and inferring activities from GPS data. Section 3 reports on some of the initial results. The results look at the both vehicle activity and chain characteristics, and activity densities from a geographic point of view. The paper concludes with a brief research agenda.

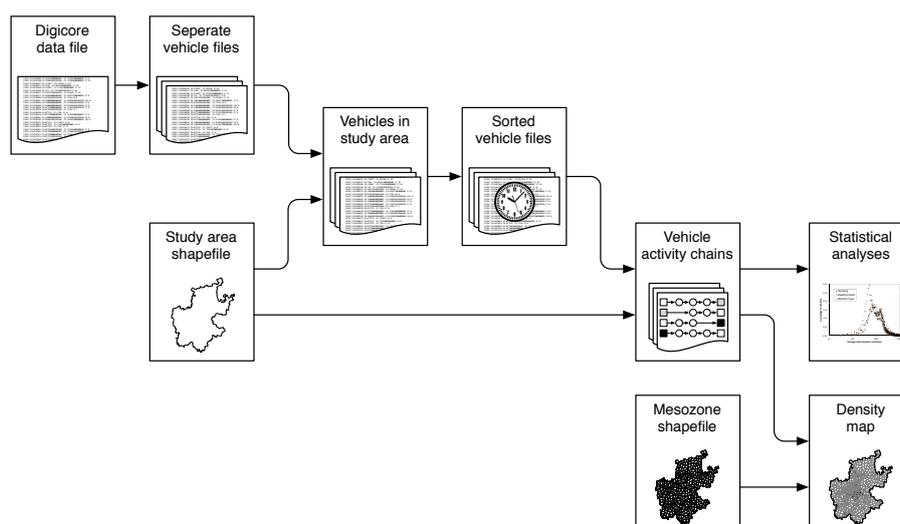


Figure 1: Process flow for the vehicle activity chains

2 Activity inferring

The data used to infer commercial vehicle activities was provided by *DigiCore Holdings Limited*, a commercial fleet management company based in South Africa. The company provides vehicle tracking service to its customers, with tracking devices not only recording location information, but also various engine and electronic triggers, including temperature, water levels, braking, ignition activity, tampering with the GPS unit, triggering of the panic alarm, and opening and closing of vehicle doors, to name but a few.

A single file containing all the GPS log records for commercial vehicles for the six months 1 January 2008 to 30 June 2008 was provided to us. We then followed the process depicted in Figure 1 to extract commercial activity chains from raw GPS data.

Each log record has six fields: 1) a unique vehicle identifier; 2) a seconds-based Unix time stamp; 3) a longitude and 4) latitude value, both in decimal degrees according to the WGS 84 reference coordinate system used by global positioning systems; 5) a vehicle status identifier used by *DigiCore* for fleet management purposes; and 6) a speed value. With a total file size in excess of 30 Gigabyte, the first step was to split the data into seperate files, one for each vehicle. A total of 31,041 vehicle files were identified, and each had to be sorted chronologically according to the time field. Since sorting was considered a time-intensive operation, we limited the files to only those vehicles that had GPS log records within the province of Gauteng. Once identified, the vehicle files were sorted. Gauteng was chosen since it is the economically most active province in South Africa. Accounting for only 1.5% of the land surface, it accounts for more than 30% of the country's Gross Domestic Product (GDP).

In this study we focussed on ignition-related triggers to identify activity start and stop times. An activity start was identified as the point when an *ignition off* trigger is received, and an

activity stops when the ignition is turned on again. To eliminate any false starts or false stops, such as a failed engine start attempt requiring the driver to switch the ignition off and back on before starting again, we provided an arbitrary threshold for activity duration of 8 minutes. The threshold is assumed realistic since commercial vehicles are involved in time-consuming loading, unloading and refuelling activities. Using vehicle status triggers to identify activities, as opposed to recurring geo-locations, allowed us to avoid identifying a heavy vehicle waiting to cross a busy intersection, or short stops at toll gates, as potential activities.

The complete list of sequential activities was then categorised as being either a *major*, lasting in excess of 5 hours, or a *minor* activity, lasting less than 5 hours. Figure 2 shows the activity durations of all activities before chains were identified. The threshold of 5 hours was

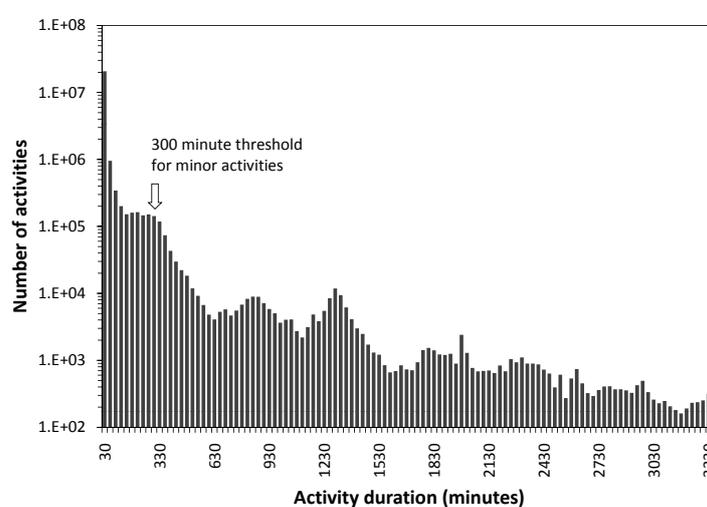


Figure 2: Histogram of all activity durations. The threshold of 300 minutes (5 hours) between minor and major activities is indicated.

arbitrarily chosen since no empirical evidence was found suggesting an intuitive, definite and distinguishable difference in activity durations.

Consider, as one detailed example, the GPS logs, and identified activities presented in Figure 3. This particular vehicle only performed activities within a very small extent during the entire study period, and illustrates the locations of both major and minor activities. We did not map the vehicle movement to the road network, mainly because the GPS log records in our data set were too infrequent (300 seconds under normal operating conditions when no other trigger is received).

Whenever a major activity was identified, a new activity chain was created, starting with the major activity. Subsequent minor activities were added sequentially until the next major activity was identified. Once all activities were chained, the chains were cleaned: any chain not starting and ending with a major activity was removed to ensure only complete chains were evaluated. Chains containing only two major activities were considered mere relocations, and since they

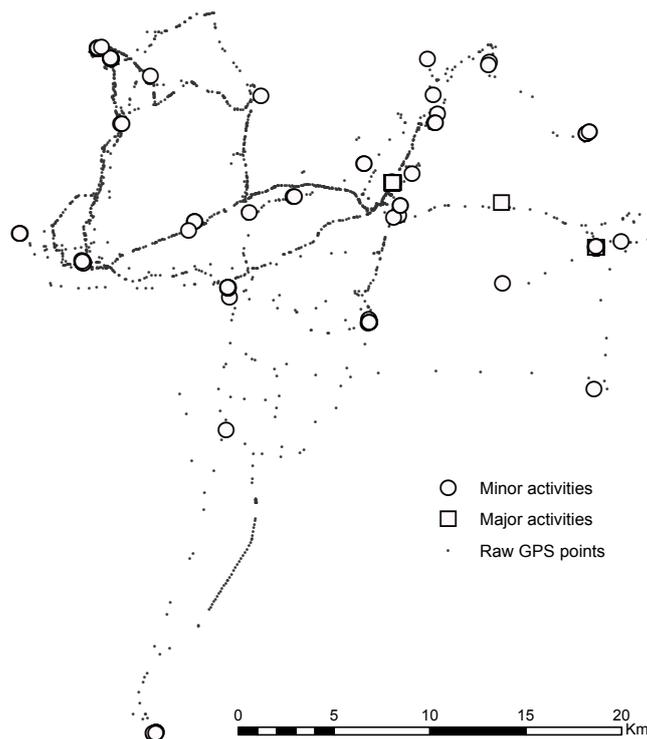


Figure 3: Example of extracting vehicle activities from raw GPS data.

made up a very small proportion of the activity chains, they were also removed.

3 Results

The more than 10.5-million identified activities were spread across southern Africa, all the way into the southern regions of the Democratic Republic of Congo. Of interest to us were the activities across South Africa. We estimated the density of commercial vehicle activities through the kernel density estimation implementation of *ESRI ArcGIS*—popularized by Silverman (1986)’s quadratic kernel function. Conceptually, kernel density estimation fits a smooth surface over each point. The volume under the surface equals the estimate number of activities in that grid cell. The density map of minor activities illustrated in Figure 4 used a grid size of 500×500 meters, with a search radius of 5000 meters. Activity densities for major activities follows a very similar pattern, but due to space the figure has been omitted here. Three areas of interest are highlighted in the figure: the economically active province of Gauteng; KwaZulu-Natal with activity concentration around the port of Durban; and the Western Cape with the activity densest around the port area.

Once the activities and activity chains were extracted, we were able to further distinguish between *within* and *through* traffic. In Figure 5 we observe the percentage of vehicles that spend

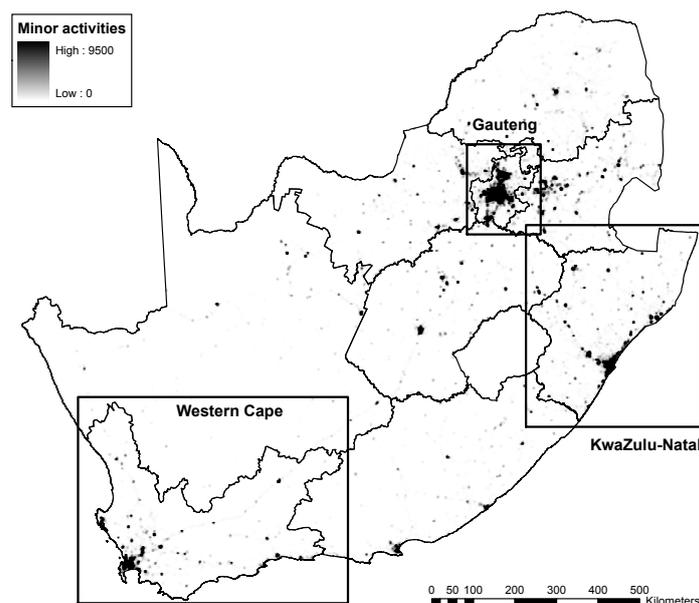


Figure 4: The extent of minor vehicle activities throughout South Africa. The three prominent activity centra are indicated

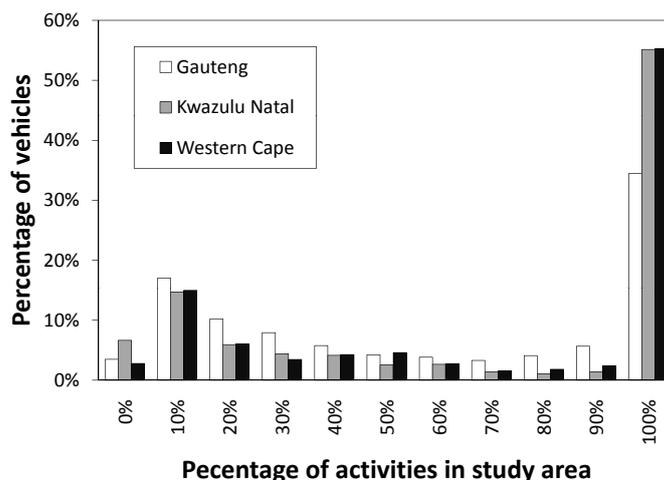


Figure 5: A histogram comparing the proportion of vehicle activities that vehicles associated with the study area spend within the study areas.

a certain percentage of their time (number of activities) in one of the three areas. To explain how to interpret the figure, consider for example the bars at the 100% mark of the percentage of activities axis. In the case of Gauteng, 35% of vehicles that pass through Gauteng, have more than 90% of all their activities in Gauteng, the other activities are outside. About 55% of vehicles that travel in Kwazulu Natal or the Western Cape, have more than 90% of its activities in that area. All vehicles with more than 90% of their activities in a single area is considered *within* vehicles, while the remainder are considered *through* vehicles.

For the remainder of the paper, we will evaluate the chain characteristics of vehicles passing

through Gauteng only. The average number of activities per chain, as depicted in Figure 6 differ between *within* vehicles and *through* vehicles, with the latter having much longer chains. The chain duration, depicted on the right in Figure 6, also shows *through* vehicles having

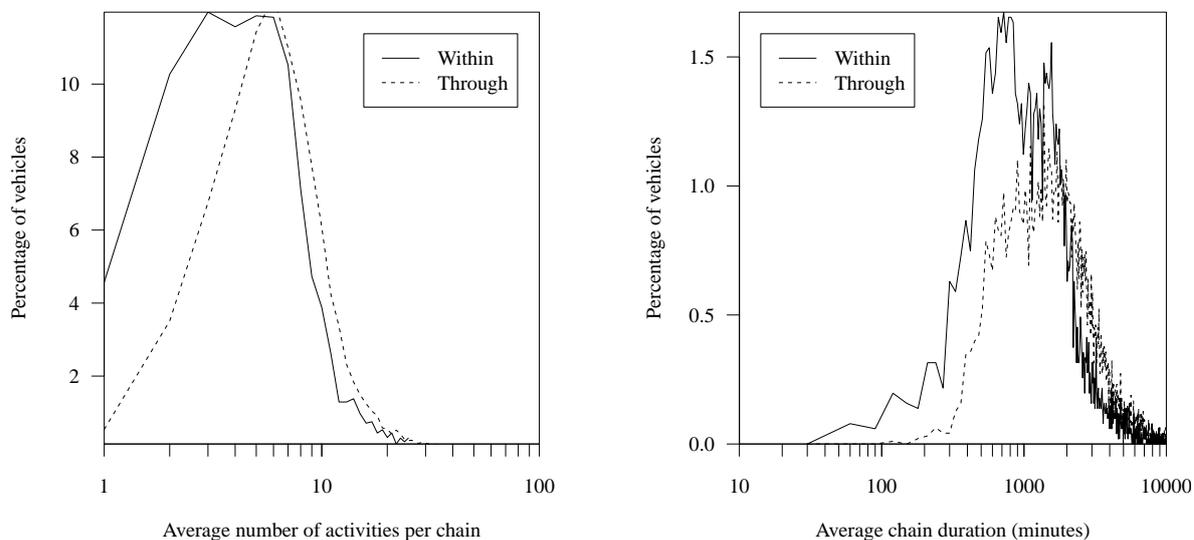


Figure 6: A comparison of the average number of activities per chain (left), and the average chain duration (right) between *within* and *through* vehicles.

longer chains. Since through traffic may be associated with long-haul transportation, the longer duration of through vehicle chains is intuitively expected.

The activity start times are depicted in Figure 7. The start times for minor activities of *through*

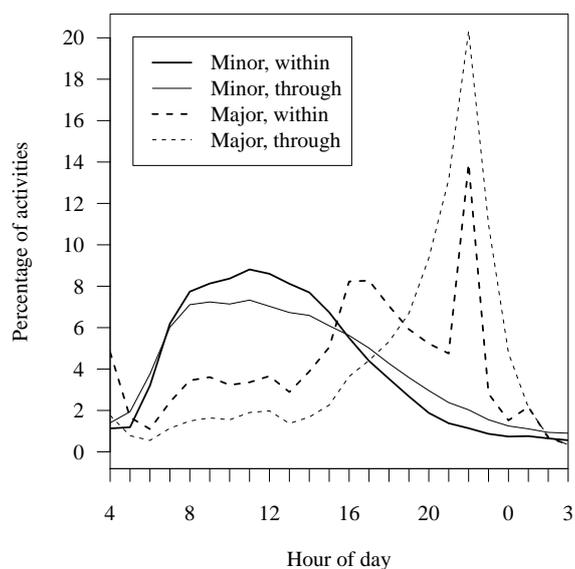


Figure 7: Activity start times.

vehicles is slightly more spread out than that of *within* vehicles, yet following a very similar distribution. The two distributions of major activities are less similar, although both seems to

peak at 22:00 in the evening. *Through* vehicles seem to have a much more distinct distribution, while the *within* vehicles' distribution is multimodal. We also expected that as the number of minor activity starts decreases, the number of major activity starts increases.

Knowing not only when the activities occur, but also where, allows us to a spatial element to the activities. Figure 8 illustrates the number of minor activities starting between 08:00 and 08:59 (as an example). The demarcation used is that of the Geospatial Analysis Platform

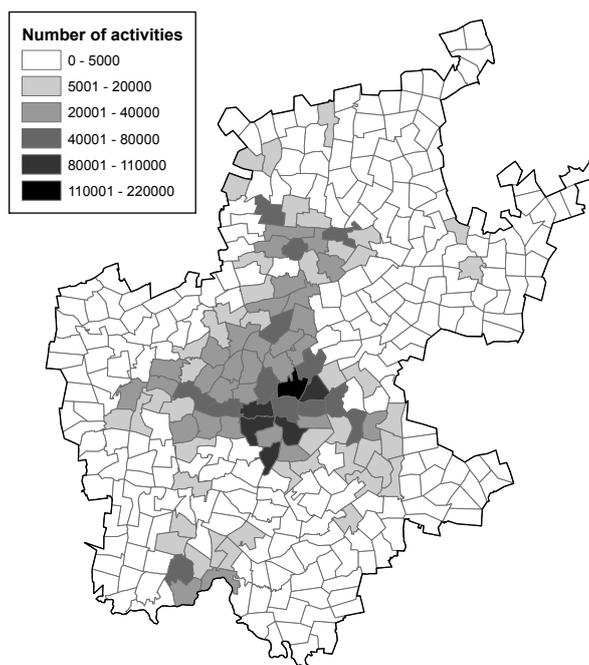


Figure 8: Spatial distribution of minor activities starting between 08:00 and 08:59.

(GAP) mesozones (CSIR Built Environment, 2007, 2009; Naudé *et al.*, 2007). Having local knowledge of the area, the high density of activities in more industrial areas are expected, yet without having further information about the underlying land use, we can do very little statistical confirmation, relying merely on observational inference.

4 Conclusion

In this paper we reported on the initial inferring of commercial vehicle activities from GPS data without having any additional information about the vehicles, or its activities. The results allow us to accurately sample from the start time, duration, and chain length distributions. Such sampling makes it possible to generate synthetic commercial vehicle populations for future simulations.

Further refinements to this paper are due. We have made a number of arbitrary choices in

inferring activities, such as minor activities lasting being between 8 and 300 minutes. These parameters must be established and tested more rigorously. The results reported on in this paper provide a decent base to understand the movement of commercial vehicles, but still requires a lot more thorough statistical inference before being considered rigorous.

5 Acknowledgement

The author wish to thank the South African National Research Foundation (NRF) for partially sponsoring the research (Grant FA2007051100019). Also, thanks to Prof Kay W. Axhausen of the Institute for Transport Planning and Systems (IVT) at ETH Zürich where the research was conducted.

References

- Ambrosini, C. and J.-L. Routhier (2004) Objectives, methods and results of surveys carried out in the field of urban freight transport: An international comparison, *Transport Reviews*, **24** (1) 57–77.
- CSIR Built Environment (2007) SA Geospatial Analysis Platform, Version 2: Spatially disaggregated and interpolated socio-economic and accessibility indicators based on 2004 magisterial district data, obtained from Global Insight's Regional Economic Focus (REX version 2.0c (190)), Collaborative project with the Policy Advice and Coordinating Unit (PCAS) of the Presidency, co-funded by GTZ.
- CSIR Built Environment (2009) Geospatial Analysis Platform and NSDP Spatial Profiles. Accessed 13 March from <http://www.gapweb.co.za>.
- Draijer, G., N. Kalfs and J. Perdok (2000) Global positioning system as data collection method for travel research, *Transportation Research Record*, **1719**, 147–153.
- Figliozzi, M. A. (2007) Analysis of the efficiency of urban commercial vehicle tours: Data collection, methodology, and policy implications, *Transportation Research Part B: Methodological*, **41** (9) 1014–1032.
- Friedrich, M., T. Haupt and K. Nökel (2007) Freight modelling: Data issues, survey methods, demand and network models, in K. W. Axhausen (ed.) *Moving through nets: The physical and social dimensions of travel*, chap. 9, 249–277, Elsevier, Amsterdam, The Netherlands.
- Hesse, M. and J. Rodrigue (2004) The transport geography of logistics and freight distribution, *Journal of Transport Geography*, **12** (3) 171–184.

- Marchal, F., J. Hackney and K. Axhausen (2005) Efficient map matching of large global positioning system data sets: Tests on Speed-Monitoring experiment in Zürich, *Transportation Research Record: Journal of the Transportation Research Board*, **1935**, 93–100.
- Naudé, A., W. Badenhorst, L. Zietsman, E. Van Huyssteen and J. Maritz (2007) Geospatial analysis platform – versions 2: Technical overview of the mesoframe methodology and south african geospatial platform., *Technical Report*, **CSIR/BE/PSS/IR/2007/0104/B**, Council for Scientific and Industrial Research, Pretoria, South Africa.
- Schüssler, N. and K. Axhausen (2009) Processing GPS raw data without additional information, Forthcoming in *Transportation Research Record*.
- Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.
- Tsui, S. and A. Shalaby (2006) Enhanced system for link and mode identification for personal travel surveys based on global positioning systems, *Transportation Research Record*, **1972**, 38–45.
- Wolf, J., R. Guensler and W. Bachman (2001) Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data, *Transportation Research Record*, **1768**, 125–134.
- Wolf, J., S. Schönfelder, U. Samaga, M. Oliveira and K. Axhausen (2004) Eighty weeks of global positioning system traces: Approaches to enriching trip information, *Transportation Research Record*, **1870**, 46–54.
- Yalamanchili, L., R. Pendyala, N. Prabakaran and P. Chakravarthy (1999) Analysis of global positioning System-Based data collection methods for capturing multistop Trip-Chaining behavior, *Transportation Research Record*, **1660**, 58–65.