Artificial Intelligence for Transportation: Advice, Interactivity and Actor Modeling: Papers from the 2015 AAAI Workshop

Concept of a Data Thread Based Parking Space Occupancy Prediction in a Berlin Pilot Region

Tim Tiedemann and **Thomas Vögele**
DFKI GmbH, Robotics Innovation Center
Robert-Hooke-Straße 1, 28359 Bremen, Germany

Mario M. Krell and **Jan H. Metzen**
University of Bremen, Robotics Research Group
Robert-Hooke-Straße 1, 28359 Bremen, Germany

Frank Kirchner
DFKI GmbH, Robotics Innovation Center and
University of Bremen, Robotics Research Group
Robert-Hooke-Straße 1, 28359 Bremen, Germany

Abstract

In the presented research project, a software and hardware infrastructure for parking space focussed inter-modal route planning in a public pilot region in Berlin is developed. One central topic is the development of a prediction system which gives an estimated occupancy for the parking spaces in the pilot region for a given date and time in the future.

Occupancy data will be collected online by roadside parking sensors developed within the project. The occupancy prediction will be implemented using "Neural Gas" machine learning in combination with a proposed method which uses data threads to improve the prediction quality.

In this paper, a short overview of the whole research project is given. Furthermore, the concept of the software framework and the learning methods are presented and first collected data is shown. The prediction method using data threads is explained in more detail.

1 Motivation

There is a significant lack of parking spaces in most larger cities today. In particular in densely populated inner-cities, parking spaces are rare, both in residential areas and in commercial zones. This has a direct impact on inner-city traffic. Previous research has shown that the search for a parking space creates a significant amount of additional traffic (Shoup 2006; VDA 2009; Delatte et al. 2014). Subsequently, this might even lead to a loss of quality of life when people avoid moving their car again once they have found a parking space, as found in previous studies (Delatte et al. 2014, Wimobil).

On the other hand, as parking spaces are a rare commodity they can be sold expensively and are thus a valuable asset for often underfunded municipalities. Efficient methods for the active management of parking spaces rank high on the agenda of urban management plans.

To avoid unnecessary traffic and to better use the available parking spaces the driver should be informed where all spaces are occupied and where in the vicinity of his target he can find an available parking space. In contrast to the already existing systems all public roadside parking spaces need to be covered: Here, (1) often unused capacities can be found and (2) usually the most preferred parking spaces (free and close to the destination) are located.

2 Related Work

Current solutions for parking space management, which cover mostly only parking garages, are not sufficient because on the one hand, public roadside parking is favoured by most drivers over parking garages, and on the other hand there are often unused, albeit hidden, capacities available. Examples are fixed parking guidance traffic signs showing the way to parking lots (with or without a display that gives the current available capacity). App- or website-based versions of such a solution are parking map services (e.g., Parkopedia) that also cover only parking lots and that usually do not give any information about the current occupancy. The parking app system developed by the Australian city of Perth goes a step further (Perth Parking App). There, a map of *available* parking spaces can be found. Furthermore, an app-based payment option is also supported. However, there are no roadside parking spaces covered, either.

More related to the solution presented in this paper are projects that include parking sensors. In the Belgian city of Kortrijk a system "SENSIT" by Nedap is installed (Vandewinckele 2014). This project already finished the pilot phase with 35 parking spaces. However, the focus of this solution is in short-term parking where, e.g., loading/unloading zones have to be kept available. Thus, the information of the occupancy sensors is only given to parking attendants and not to users seeking for available parking spaces. The manufacturer offers parking guidance displays but an access via a public website or a public app is not available.

Another disadvantage of the solution chosen in the city of Kortrijk is the sensor setup with separated and fixed parking bays with one monitoring detector per bay. The detector is mounted in or on the parking space surface. Similar setups are planned or already installed in other cities like London and Moscow.

The solution presented in the following sections combines two of the objectives mentioned above: first, the detection of parking violations with data access and data visualization for parking attendants. Second, a public access and route planning tool available for users searching for an available parking space. An occupancy prediction is included to guide users when or before they start their trip. An automatic accounting option is prepared and a website and an app access will be implemented, too. With the decision to use top-view sensors, the detection of arbitrarily laid out (slotted) roadside parking spaces or even roadside parking without any parking bay markers (unslotted) including “second-row” parking is possible.

A very interesting proposal with comparable properties of the occupancy detection method was published by Mathur et al. (2010). They used ultrasonic sensors attached to the passenger side of a vehicle to collect parking occupancy data while driving. The sensors measure distances to passenger-side obstacles and the measured values are classified as “occupied” or ”available” space. With three vehicles in a 2 month time frame they collected a total of more than ≈ 500 miles of data. As drawbacks of their setup they reported (potential) problems when driving at higher speeds and detecting parking occupancy on multi-lane roads (among others issues). In their data collection there was only data from roads with single lanes. Coric and Gruteser (2013) use the same data collection method to automatically generate a map of legal and illegal parking spaces.

Further publications from other application domains that propose methods comparable to those discussed in this paper are presented in Section 5.2.

3 Proposed Solution

The research project “City2.e 2.0”, which provides the framework for the research work presented in this paper, brings together municipalities (Senate Department for Urban Development and Environment, Berlin), mobility information service providers and IT-service developers (VMZ Berlin[1]), large IT and infrastructure providers (Siemens AG), and research institutes specialized both on the technical (DFKI RIC[2]) and legal and economic (IKEM[3]) dimension of the problem (City2.e 2.0). The objective is to develop a solution for reliable monitoring and prediction of parking space availability.

The underlying idea for the City2.e 2.0 solution is that if drivers are informed as early as possible about the availability of parking spaces in a given region, the amount of unnecessary traffic can be significantly reduced and the parking spaces in that region can be optimally used.

In City2.e 2.0, a solution for the predictive management of public and semi-private roadside parking spaces is developed and evaluated in a real-world test environment in Berlin. The solution integrates sensors for the monitoring of parking spaces, a backend system for the collection of real-time data, and tools for the prediction of parking space occupancy. As user interface a website and an app are planned.

A prediction system is needed to guide the user to a parking space available at the time of his arrival. This system learns the occupancy for the specific location and time based on long-term gathered data. These learned predictions can also give the user good estimates for future travels. Details of the machine learning approach used in this project are described in Section 5.

4 Project’s Pilot Region

The project’s main test area is to be set up in 2015. It will be located in a public area at the Bundesallee in Friedenau (borough Tempelhof-Schöneberg) in Berlin (see Figure 1). The set up will include occupancy sensors covering more than 50 parking spaces. Parts of the covered parking spaces are close to commercial zones while other parts are in more residential areas. Some parking spaces are in parallel to the street lanes, some are perpendicular. Most of the parking spaces do not have any separation marks painted on the ground. Furthermore, the sensors will be able to detect “second row” parking cars (occupying additional regular parking spaces), too. Also covered are parking spaces located at charging stations for electric cars.

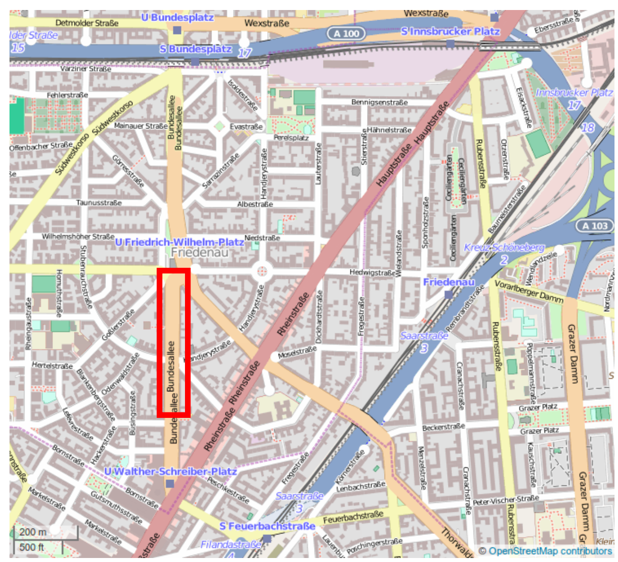

Figure 1: Location of the project pilot region in Friedenau in the Berlin borough Tempelhof-Schöneberg. (© OpenStreetMap contributors, www.openstreetmap.org).

During the project runtime the parking spaces of the pilot region will be available to the public like they were before (with the same potential limitations, e.g., for disabled people, for residents’ parking passes, or for electric cars). This ensures real-world conditions for the project’s final studies. An IT infrastructure will be set up including a web front end with a route planner accessible to the public, too. Via this

[1] VMZ Berlin Betreibergesellschaft mbH

[2] German Research Center for Artificial Intelligence, Robotics Innovation Center

[3] Institute for Climate Protection, Energy and Mobility

front end the parking sensor data and the predictions (and their influence on the routing results) are accessible by the parking space users.

In this pilot region the cars and their drivers are not identified or tracked. Especially, the number plates are not scanned. Just the occupancy of a parking space is detected. To identify residents' parking passes, maybe a side channel communication solution will be tested. For the pre-studies in this and former projects, hash values of the number plates and of vehicle properties were used. These hashes cannot be used to identify a vehicle or its owner. Privacy preserving issues are dealt with by the other project partners.

5 Methods

To study a directed, parking-space-related user navigation, it is required to set up a real-life fully-equipped test area. The equipment includes

- sensors to detect occupied and available parking spaces,
- an IT infrastructure to collect and distribute occupancy data,
- a prediction system to supply future occupancy data, and
- a navigation system to guide the user to available spaces.

The sensor is developed and built by the project leader Siemens AG. It is designed as a roadside, top-view sensor to be mounted to walls or street lights. The IT infrastructure is set up by VMZ Berlin. Also the multi-modal routing software was developed and implemented by VMZ Berlin. The software can cover the path between parking spaces and the actual start and target position.

The prediction system is developed and implemented by the DFKI Robotics Innovation Center. It consists of (1) a data collection and machine learning based prediction basis (including, e.g., class prototypes, see below) and (2) a query response system generating parking space occupancy predictions for a specific demanded time in the future using the learned prediction basis. To automatically compare different methods and parameterizations of the algorithms, the processing will be implemented using the signal processing and classification environment pySPACE (Krell et al. 2013). This framework supports regression, and clustering support will be added, too.

5.1 Parking Occupancy Prediction Basis

The data basis and the desired prediction results have two dominant properties:

1. The amount of collected data is rather small. A temporal resolution of 15 minute slots is sufficient for both, original sensor data collection and as temporal resolution of the predictions. The data basis per sample consists of just one value, namely the number of available parking spaces per parking "lot". Each parking "lot" in this particular sense covers a group of a few up to a few dozen of parking spaces and is combined in one prediction. Thus, for each lot and each 15 minute slot in the future, there is one value (the number of available parking spaces) to be predicted. Hence, the data set to be acquired per year and per lot is as small as 35 KB (or even with 64 bit alignment and additional overhead for indices and pointers about 1 MB).

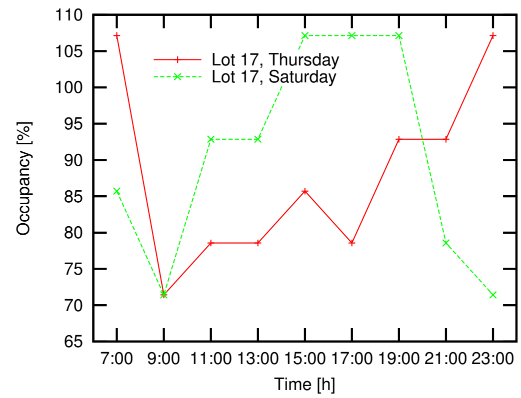

Figure 2: Example occupancies of one area of 14 parking spaces at a side street (lot 17). Plotted are the data of one Thursday and one Saturday. As can be seen, the occupancy change over the day can differ a lot for the same parking lot. In most of the lots there are no separation marks painted on the ground. The number of cars fitting in one lot is estimated by standard parking space sizes. Hence, more than 100% occupancy can be found if smaller cars and/or smaller inter-car distances and/or cars are present that cross the lot borders. Data is taken from the Wimobil project (Wimobil).

2. As first studies have shown, there is a dominant 24 hour periodicity plus weekly repetitions (day of the week dependency) in the parking occupancy data (see Figure 2). This leads to a typical occupancy behaviour which mainly depends on the location (e.g., residential vs. commercial area) and the day of the week. Under the same conditions the occupancy is not expected to vary a lot (Wimobil). However, it can be expected that certain external (and maybe unknown) conditions lead to different classes of occupancy behaviour (e.g., holidays, winter weather, street works, sports/cultural events).

These properties led to the following conclusions regarding the prediction methods:

1. The whole sensor data can be logged and can be taken into account when generating predictions or when adapting the prediction basis to new sensor data (details see below).
2. The predictions have to be learned for each lot independently and need the time of day and the day of the week as input.
3. To learn the standard behaviour within one class (e.g., usual Thursday, no holidays, dry weather) and to separate data of different classes (e.g., holidays vs. no holidays) a clustering method has to be applied. Since the reason for a change in the parking behaviour is potentially unknown (depends on hidden variables) an unsupervised method needs to be used. The first candidate method to be tested will be the neural gas algorithm (Martinetz, Berkovich, and Schulten 1993). The two tasks *clustering* and *proto-*

type generation (within one class) may have to be separated, see Section 5.2.

4. In a previous study, occupancy was studied at three days for 71 lots (Wimobil). There, it could be found that some parking lots seem to have similar occupancy behaviours while others differ a lot. While in general it could be an option to combine the learning for predictions of similarly behaving lots this will most likely not be done here. Given the small data set sizes and the needed temporal resolutions, the advantage of reducing the computational effort and the needed data space is not relevant in this project. Furthermore, there are just a few and rather different lots planned. Thus, the prediction quality is expected to be impaired. However, a combination of several lots' data in learning to improve the results is still an option that might be studied.

To avoid overfitting within one occupancy prediction class, the data needs to be averaged or combined in some other way to a prototype time series. In Figure 3 the variations over five weeks are given for the whole pilot area (currently 59 parking spaces are observed manually) together. The data is currently collected manually by the project partner Berlin Senate Department to get a long-term view over the future test and demonstration setup. Each plotted curve was sampled at the same day of the week and at the same time of day. One method to combine the samples (besides averaging) could be the neural gas algorithm (potentially in combination with the separation of different behaviour classes).

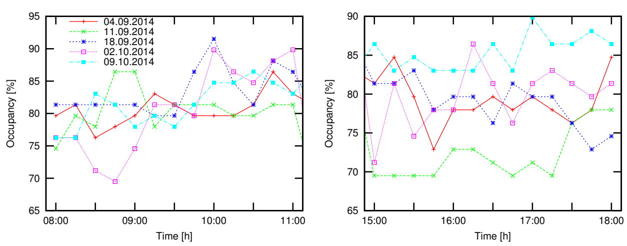

Figure 3: Example of variations in the parking occupancy data over five weeks (for all lots together with 59 parking spaces in total). The left plot shows morning data, right plot afternoon. Data was collected by the Senate Department for Urban Development and Environment, Berlin.

5.2 Data Threads Based Predictions

From the data processing perspective, the task of the project is time series prediction. Prominent solution algorithms are Neural Gas (Martinetz, Berkovich, and Schulten 1993) and Kalman filters (Okutani and Stephanedes 1984; Harvey 1990). Also least squares have been used, to predict financial time series (Van Gestel et al. 2001). In the context of urban time series prediction, the most prominent example is traffic density estimation (Wang and Papageorgiou 2005; Papageorgiou 1983; Munoz et al. 2003; Okutani and Stephanedes 1984; Zuefle et al. 2008). Other examples are the analysis on water consumption (Zhou et al. 2000; Cardell-Oliver 2013), air pollution forecasting (Niska et al. 2004; Schweizer et al. 2011), and waste management (Dyson and Chang 2005).

In contrast to many other time series prediction problems, the amount of data is rather small here (De Gooijer and Hyndman 2006). For example, if you take an application using electroencephalogram (EEG) data, the dimensionality and the temporal resolution is much higher (Kirchner et al. 2013). Moreover, the different occupancy behaviour classes can differ a lot and new classes might come up often and quickly (e.g., street works, company move, concerts or sports events). In such cases, accessing old data (yesterday's or three years ago) and identifying the best matching class even if it is a small one (e.g., *only* yesterday's data) could improve the prediction quality a lot. However, there will be no means to use further external data to identify the correct class for a queried prediction.

Proposed is a solution using threads of data i.e. chunks of the timeline of previously recorded occupancy behaviours. The expression "threads" is chosen deliberately to distinguish it from general time series which can be unlimited in start or end. The threads used here have particular start and end points. Furthermore, there can exist several threads in parallel for the same time period (as explained below).

The proposed method is a two-step approach. The first step of learning is an (unsupervised) clustering of the logged data into classes of similar parking behaviour (similar over 24 hours). In the second step all 24-hours-threads of one class are taken to generate a prototypical parking behaviour of this class. For example, this prototype could be the average of all threads or Gaussian distributions fitted to the threads.

The recall is a two-step process, too: First, the identification of the best matching class and, second, the computation of the actual prediction based on the selected threads. To identify the best matching behaviour class, not only the time of day and the day of the week is used but also the current occupancy situation at the time of the query. By this means, the maximum of the data available (the most recent measurements) can be used to adapt to very recent changes in the parking behaviour. Moreover, this method exploits the temporal relationship between a) the recorded data matching the current measurements and b) the recorded data following a) up to the data set equivalent to the queried prediction (see Figure 4 for an example plot). This means that if, for example, new road works (started shortly before the time of the query t_q) influence the latest measurements then the temporal relationship between previous road works' beginnings and previous road works' subsequent parking behaviours can be used for a better prediction of the new road works' behaviour.

Additionally, as all recorded data are available, a very good estimation of the prediction error can be computed and given to the user. This is also possible because of the two-step recall process (see Figures 4 and 5): At first only the learned time series *prototypes* of all classes are compared with the recent measurements (i.e. one per class). Then all original data threads associated with the winning class (i.e. all data that generated this class) are followed to their partic-

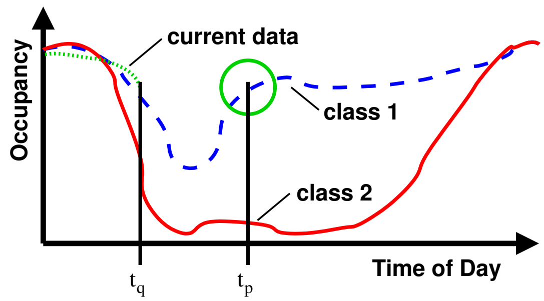

Figure 4: Visualization of the occupancy prediction method using data threads. At point t_q the user queries for a parking space occupancy prediction for the point t_p. "class 1" (dashed blue curve) and "class 2" (solid red curve) are two potential time series (prototype data threads) learned before (e.g., Thursday in winter weather, holidays vs. Thursday but dry working day). The dotted green curve depicts the current measurements from the sensor. For these current measurements at t_q (or a time span some hours before up to t_q) the best matching class is taken. This winning data thread is followed up to t_p to find the occupancy value to be predicted.

ular prediction value related to t_p. This leads to a range of occupancies that in the history followed an occupancy situation as it is currently measured. Now, all selected data threads together show the distribution of occupancies (of the winning class) in the past. The occupancies of other parking behaviour classes are filtered out in the first step. Thus, we expect to end up in a distribution and an error estimation based on that distribution which fits well to the situation at the time t_p. In case there are multiple winning classes, this range includes all those associated data threads' values related to t_p. Perhaps, a soft-assignment will be needed here by weighting the N closest ("winning") classes differently. Depending on the user's preferences a standard deviation, a minimum, or a maximum of all this range of historical values can be given. Even more complex – but in this application more appropriate – estimations can be computed, for example, the probability (given all historical values related to t_p) that there will be at least one free parking space.

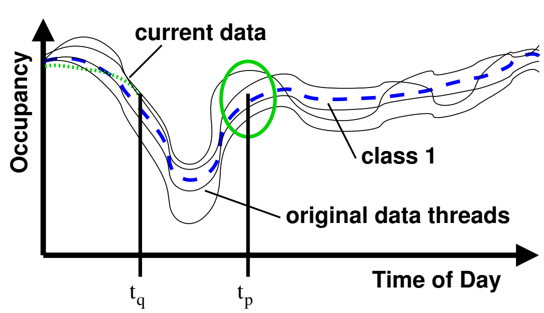

Figure 5: Visualization of the occupancy prediction method using data threads (step 2). The thin lines depict the original data threads (sensor measurements).

However, of course there will be the case of predictions for the farther future ($t_p \gg t_q$) where there is no relationship to the current measurements anymore. In this case, the class selection needs to be changed. Possible could be the selection of an "overall standard" class or, of course, the selection of all classes. This needs to be studied when the parking space sensors are set up and data can be collected. While the user should be able to give any future date, the selection of the best class(es) is designed for the focus of the prediction time span (which is the next minutes and hours, up to one day).

6 Summary and Outlook

Presented was the concept of a parking space occupancy prediction method using a combination of Neural Gas vector quantization and the raw data timeline in data threads. The combination of a machine learning clustering method and the original temporal relations of the raw data is supposed to lead to good prediction results. However, the applicability of the presented methods and the quality of the prediction results need yet to be studied with the real data.

Furthermore, the whole research project in which the prediction system is embedded was described. The main focus of the project is a practical real-world test of public roadside parking detection in a Berlin pilot region. Included is an IT infrastructure with an inter-modal routing planner using the current parking occupancy sensor data as well as the parking space occupancy predictions.

The main pilot area in Berlin will be set up in 2015. When online, all parking occupancy data will be logged and used to train the occupancy prediction system (planned to run for at least half a year). At the end of the project an evaluation of the whole system and of the prediction quality will be carried out.

A Acknowledgements

Supported by the Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety on the basis of a decision by the German Bundestag, grant no. 16EM2051-5.

References

Cardell-Oliver, R. 2013. Water use signature patterns for analyzing household consumption using medium resolution meter data. *Water Resources Research* 49(12):8589–8599.

City2.e 2.0 website. http://www.erneuerbar-mobil.de/de/projekte/vorhaben-im-bereich-der-elektromobilitaet-von-2013/kopplung-der-elektromobilitaet-an-erneuerbare-energien-und-deren-netzintegration/city2e.

Coric, V., and Gruteser, M. 2013. Crowdsensing maps of on-street parking spaces. In *Distributed Computing in Sensor Systems (DCOSS), 2013 IEEE International Conference on*, 115–122. IEEE.

De Gooijer, J. G., and Hyndman, R. J. 2006. 25 years of time series forecasting. *International journal of forecasting* 22(3):443–473.

Delatte, A.; Kettner, S.; Schenk, E.; and Schuppan, J. 2014. *Multimodale Mobilität ohne eigenes Auto im urbanen Raum*. http://www.ivp.tu-berlin.de/fileadmin/fg93/Forschung/Projekte/City_2.e/IVP_Projektbericht_City2e_Lange_Fassung.pdf (in German).

Dyson, B., and Chang, N.-B. 2005. Forecasting municipal solid waste generation in a fast-growing urban region with system dynamics modeling. *Waste Management* 25(7):669 – 679.

Harvey, A. C. 1990. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.

Kirchner, E. A.; Kim, S. K.; Straube, S.; Seeland, A.; Wöhrle, H.; Krell, M. M.; Tabie, M.; and Fahle, M. 2013. On the applicability of brain reading for predictive human-machine interfaces in robotics. *PLoS ONE* 8(12):e81732.

Krell, M. M.; Straube, S.; Seeland, A.; Wöhrle, H.; Teiwes, J.; Metzen, J. H.; Kirchner, E. A.; and Kirchner, F. 2013. pySPACE - a signal processing and classification environment in Python. *Frontiers in Neuroinformatics* 7(40). https://github.com/pyspace.

Martinetz, T.; Berkovich, S.; and Schulten, K. 1993. Neural-gas' network for vector quantization and its application to time-series prediction. *Neural Networks, IEEE Transactions on* 4(4):558 –569.

Mathur, S.; Jin, T.; Kasturirangan, N.; Chandrasekaran, J.; Xue, W.; Gruteser, M.; and Trappe, W. 2010. Parknet: drive-by sensing of road-side parking statistics. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, 123–136. ACM.

Munoz, L.; Sun, X.; Horowitz, R.; and Alvarez, L. 2003. Traffic density estimation with the cell transmission model. In *American Control Conference, 2003. Proceedings of the 2003*, volume 5, 3750–3755.

Niska, H.; Hiltunen, T.; Karppinen, A.; Ruuskanen, J.; and Kolehmainen, M. 2004. Evolving the neural network model for forecasting air pollution time series. *Engineering Applications of Artificial Intelligence* 17(2):159 – 167. Intelligent Control and Signal Processing.

Okutani, I., and Stephanedes, Y. J. 1984. Dynamic prediction of traffic volume through Kalman filtering theory. *Transportation Research Part B: Methodological* 18(1):1–11.

Papageorgiou, M. 1983. *Applications of Automatic Control Concepts to Traffic Flow Modeling and Control*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Parkopedia. http://www.parkopedia.com.

City of Perth Parkings mobile app. http://www.cityofperthparking.com.au/mobile-app.

Schweizer, I.; Bärtl, R.; Schulz, A.; Probst, F.; and Mühläuser, M. 2011. Noisemap-real-time participatory noise maps. In *Proc. 2nd International Workshop on Sensing Applications on Mobile Phones (PhoneSense11)*, 1–5.

Shoup, D. C. 2006. Cruising for parking. *Transport Policy* 13(6):479–486.

Van Gestel, T.; Suykens, J.; Baestaens, D.-E.; Lambrechts, A.; Lanckriet, G.; Vandaele, B.; De Moor, B.; and Vandewalle, J. 2001. Financial time series prediction using least squares support vector machines within the evidence framework. *Neural Networks, IEEE Transactions on* 12(4):809–821.

Vandewinckele, J.-P. 2014. Retailers benefit from multiple parking space utilization. http://www.nedapidentification.com/uploads/2014%2520update/Article%2520Kortrijk%2520JPvdW%2520in%2520Vexpansie_v5.pdf.

VDA. 2009. *Annual report 2009*. http://www.vda.de/en/downloads/636/.

Wang, Y., and Papageorgiou, M. 2005. Real-time freeway traffic state estimation based on extended Kalman filter: a general approach. *Transportation Research Part B: Methodological* 39(2):141–167.

Wimobil. *(not published yet)*. http://www.erneuerbar-mobil.de/de/projekte/foerderung-von-vorhaben-im-bereich-der-elektromobilitaet-ab-2012/ermittlung-der-umwelt-und-klimafaktoren-der-elektromobilitaet/wimobil.

Zhou, S.; McMahon, T.; Walton, A.; and Lewis, J. 2000. Forecasting daily urban water demand: a case study of melbourne. *Journal of Hydrology* 236(34):153 – 164.

Zuefle, A.; Schubert, M.; Kriegel, H.-P.; and Renz, M. 2008. *Statistical Density Prediction in Traffic Networks*. chapter 62, 692–703.