EUROPEAN TRANSPORT CONFERENCE 2020

Integrative analysis of traffic and situational context data to support urban mobility planning

Sofia Cerqueira
INESC-ID, Instituto Superior Técnico, Universidade de Lisboa & LNEC
Elisabete Arsénio
Laboratório Nacional de Engenharia Civil (LNEC), Department of Transport
Rui Henriques
INESC-ID, Instituto Superior Técnico, Universidade de Lisboa

ABSTRACT
European cities are placing a larger emphasis on urban data consolidation and analysis for optimizing public transportation in response to urban mobility dynamics. In spite of the existing efforts, traffic data analysis often disregards vital situational context, such as social distancing norms, public events, weather, traffic generation poles, or traffic interdictions. Some of these sources of situational context data are still private, dispersed or unavailable for the purpose of planning or managing urban mobility. The Lisbon City Council has already started efforts for gathering of historic and prospective sources of situational context in semi-structured repositories, triggering new opportunities for context-aware traffic data analysis.
In this context, this paper adds value to the current theory and practice with three major contributions. First, we propose a methodology to integrate situational context around urban mobility in descriptive and predictive analysis of traffic data, with a focus on the following major spatiotemporal traffic data structures: i) geo-referenced time series data; ii) origin-destination tensor data; iii) raw event data. Second, we introduce additional principles for the online consolidation and labeling of heterogeneous sources of situational context. Third, we offer compelling empirical evidence of the impact produced by situational context aspects on urban mobility, with particular incidence on public passenger transport data gathered from card validations along the bus (CARRIS), subway (METRO) and bike sharing (GIRA) modes in Lisbon.
The research reported in this paper is anchored in the ongoing contributions made available in the pioneer research and innovation ILU project, a project that joins the Lisbon city Council and two research institutes with the aim of applying current advances in the field of artificial intelligence to move towards context-aware and sustainable passengers' mobility.

Keywords: *Situational Context; Sustainable Mobility; Data Science; Smart Cities; Public Transport*

1. INTRODUCTION
Worldwide, urban centers entail complex changes in urban mobility patterns to address climate and environmental goals. Mobility dynamics are being challenged by the emerging COVID-19 pandemic crisis. Urban decarbonization initiatives encompassing enhanced public transport, infrastructure interventions towards active modes such as walking and cycling, along with sharing services, are aimed to provide a better fit with individual mobility needs and the distribution of traffic generation poles. In this light, most European cities such as Lisbon in Portugal are pursuing technological solutions to face the challenge of self-actualizing public passenger transportation systems in

accordance with the real traffic dynamics. To this end, efforts are being established towards collection, consolidation and exploration of traffic data from public and private, passive and active modes of transport.

Traffic dynamics are situated, i.e. dependent on a high multiplicity of situational factors including spatiotemporal context (generated and attracted traffic by urban poles along the day), meteorological context, occurring events, calendrical context, or traffic interdictions (e.g. due to road accidents and construction works). Despite their well-recognized impact on urban mobility, sources of urban situational context are commonly disregarded for two major reasons. First, some of these sources are unstructured, private or unavailable, preventing their automated consolidation. Second, existing principles for context-aware traffic data analysis remain largely dispersed. In fact, state-of-the-art contributions for context-aware descriptive and predictive tasks generally fail to model the joint impact that these multiple sources of context exert on urban mobility.

In addition, existing works generally fail to separate the important role of both historical and prospective sources of context. For instance, historical weather and event data are important to assert the true impact of weather and event variables in demand, while prospective events and weather forecasts are essential to support predictions.

The research work presented in this paper proposes a comprehensive set of principles to incorporate historical and prospective sources of situational context into descriptive and predictive models of urban traffic data. Initial evidence on the impact produced by different sources of context on urban traffic is further provided, motivating the relevance of pursuing context-aware analysis to support urban mobility decisions.

We tackle the problem of incorporating situational context from three distinct angles: i) how to perform context-aware analysis for different traffic data structures, including geo-referenced time series data, origin-destination tensor data, and raw event data (e.g. smart card validations on the public transport network); ii) how to consolidate the different sources of situational context; and iii) how to adapt the learning in accordance with the targeted task, including modeling, pattern discovery, anomaly and outliers detection and forecasting tasks. To this end, the Lisbon city is used as a case study for implementing the proposed research contributions.

In recent years, the Lisbon City Council (CML) established protocols to monitor multiple sources of traffic data and situational context. Heterogeneous sources of situational context data are being consolidated in semi-structured repositories, offering unique opportunities for context-aware traffic data analysis. Among the existing initiates propelled by the city Council, the ILU project, a pioneering research and innovation project that unites two research institutes in an effort to reveal the hidden knowledge behind urban data and make advances in data science towards developing the field of context-aware and integrated traffic data for urban mobility analysis and optimization. The contributions reported in this work are anchored in the empirical observations gathered along the first stage of this project, providing a study case of interest to be followed by other European cities.

This paper is structured as follows: *section 2* introduces essential background; *section 3* surveys related work on context-enriched data analysis; *section 4* combines existing and novel principles within a coherent methodology for context-aware traffic data analysis; *section 5* preliminary evidence collected from Lisbon mobility data highlighting the relevance of incorporating sources of situational context; *section 6* presents final remarks.

2. BACKGROUND

This section provides a structured view on traffic data analysis (*section 2.1*), identifies major sources of situational context data (*section 2.2*), and introduces the Lisbon's urban ecosystem as our case study (*section 2.3*).

2.1 Traffic data analysis

Four major sources of traffic data are generally monitored in urban centres:

- road traffic data produced by: i) stationary devices positioned along the city, typically inductive loop counters found in the major road junctions of a city; and ii) geolocalized speed meters from active mobile devices used along private and public road trips (GoogleMaps[1], WAZE[2] and TomTom[3] data are paradigmatic examples);
- automated fare collection (AFC) data from public transport operators, generally consisting of smart card validations from users at stations or vehicles. For each validation, a record is issued with the passenger identifier, timestamp, boarding or alighting location and, for validations inside vehicles, additional details pertaining to the vehicle and route. In cities, such as Lisbon, the ticketing systems of public carriers are consolidated, offering the possibility to trace multimodal user movements along the public transportation network;
- public transport planning data, generally termed General Transit Feed Specification (GTFS) data, encompassing all information pertaining to schedules, routes/lines, stations/stops of the public carriers in a standardized format;
- individual trajectory data collected from mobile devices during active modes of transport.

The above traffic data structures are generally mapped into new spatiotemporal data structures more conducive to the subsequent data analysis. Generally, we find four major representations of traffic data:

- geo-referenced time series data (*section 2.1.1*) mapped from all listed sources of traffic data, generally providing heterogeneous views on traffic dynamics along fixed intervals of time, such as passenger volume (card validations) at a given station or road traffic speed and frequency at a particular road junction;
- origin-destination tensor data (*section 2.1.2*) mapped from paired entry-and-exit card validations of users along the public transport network, as well as from trajectories produced from road and active modes of transport;
- raw event data (*section 2.1.3*), comprising three major types of mobility events: i) smart card validations, ii) events associated with road traffic congestions (as generally produced by GoogleMaps, TomTom and WAZE when speed limits along certain roads decay significantly), and iii) individual trajectories.

[1] https://www.google.com/maps

[2] https://www.waze.com/en-GB

[3] https://www.tomtom.com/en gb

2.1.1 Traffic time series analysis

Traffic data can generally be represented as georeferenced time series data, pairs (time series, coordinates), where a time series is a set of ordered observations $x_{1..T} = (x_1, \ldots, x_T)$, each observation x_t being recorded at a specific time point or interval t. Time series can be univariate, $x_t \in \mathbb{R}$, or multivariate, $x_t \in \mathbb{R}^m$, where $m > 1$ is the multivariate order (number of variables).

Given a time series x, descriptive tasks aim at extracting patterns or modelling x, while predictive tasks generally aim to forecast the *h* upcoming observations, $x_{T+1\,..\,T+h}$, from available observations $x_{1..T'}$. These tasks typically focus on a single variable at a time, being complementary variables in multivariate time series used to aid the description or prediction of the target variables.

Classical approaches for time series analysis generally rely on statistical principles, including decomposition, auto-regression, differencing and exponential smoothing operations. Time series can be decomposed into trend, seasonal, cyclical, and irregular components using additive or multiplicative models (Jain, 2001). Although these models are inherently descriptive in nature, the components can be projected along a time horizon for predictive ends. Alternatively, auto-regressive and moving-average models can be combined to either describe or forecast stationary time series (Wei, 2006). These models can be extended with simplistic differencing operations for dealing with non-stationarity, and their learning guided in the presence of seasonal terms. Exponential smoothing operations can be alternatively considered to describe and predict time series (Wei, 2006). Holt-Winters offers a form of triple exponential smoothing sensitive to level, trend and seasonal behavior.

Advances from machine learning aim at mitigating the challenges faced by classical approaches. In this context, time series are generally segmented to compose a dataset that guides the learning of the target descriptive and predictive models under a specific loss criterion. Among the wide-diversity of contributions on (multivariate) time series analysis, two major groups of approaches are here highlighted. First, distance-based approaches for time series description and prediction that rely on similarities between (multivariate) time series (lazy learning) and expectations (barycentre computation) (Cai et al., 2016). Second, neural network approaches rely on the composition of simple linear functions (neurons) to learn complex mappings. In the context of time series analysis, the mapping can either be descriptive (e.g. auto-encoders) or predictive (e.g. regressors) (Bao et al., 2017).

2.1.2 Origin-destination traffic data analysis

Smart card validations or individual trajectory data can be mapped into origin-destination (OD) tensors, generally a numeric three-dimensional $location_{origin} \times location_{destination} \times variable$ matrix inferred along a specific region and time period (Munizaga and Palma, 2012). Origin and trip destinations are generally given by some spatial criterion of interest (e.g. station, zoning scheme, or geographical mesh). Variables generally reveal information pertaining to origin-destination trips, typically the total volume of individuals or vehicles moving along each origin-destination pair.

In the context of automated fare collection data, entry and exit validations per transport operator should be paired to identify a trip segment. As a single trip may require commutes within a single carrier or between different carriers, trip segments should be then properly concatenated in order to identify the true origin and destination of a single whole trip. Segment concatenation typically relies on spatial and temporal assumptions specifying the maximum acceptable walking distances and waiting times per user (e.g. commuter) (Yang and Diez-Roux, 2012).

Similar principles are commonly applied to individual trajectory data, where the time spent by individuals at a single location or disconnected/untracked periods should provide the desirable splitting-concatenating criteria. Additional challenges may be observed in OD tensor inference from public passenger transport data for each operator that only require card validations at the entry or exit of a trip. In this context, entry or exit estimators need to be implemented (Munizaga and Palma, 2012).

A final consideration should be stated regarding the time period for inferring OD tensors. OD tensors can be either inferred from different calendar periods (such as weekdays or holidays), possibly spanning an arbitrarily-high number of days. This possibility gives rise to two different types of OD tensor data structures (Van Der Zijpp, 1997). First, single OD tensors whose variables (e.g. passenger volume) generally correspond to the daily average (e.g. daily passenger volume) along the targeted time period. Second, OD tensor series, i.e. an ordered set of OD tensors where each tensor generally corresponds to a single day or period within a day.

Given a single OD tensor, descriptive tasks generally aim at finding frequent patterns or detecting vulnerabilities. In the context of public transport, in addition to passenger volume, alternative statistics can be computed for each origin-destination pair, including mean and variance estimators of the i) number of commuting trips per passenger, ii) overall travel time per passenger, iii) travel time for commuting trips per passenger, or iv) walking distance spent in commuting trips per passenger. Understandably, these new variables provide important views on critical trips (origin-destination pairs) with regards to commuting needs and waiting times. In the context of predictive tasks, single OD tensors are generally assumed as ground truths for OD movements to be observed in the near future.

To surpass the inability to capture trends and seasonal factors within single OD tensors, series of OD tensors can be alternatively considered for both descriptive and predictive ends (Van Der Zijpp, 1997). In this context, an OD tensor series can be generally seen as a multivariate time series with a multivariate order as high as the number of origins, destinations or origin-destination pairs. In this context, the listed classical and machine learning principles in *section 2.1.1* for time series analysis can be considered for extracting patterns, modeling OD distributions and forecasting upcoming OD movements (Wei, 2006; Jain et al., 2016; Gamboa, 2017).

2.1.3 Raw traffic data analysis

An event is a tuple $e = (x, s, \tau)$, where:

- $x = (x_1, .., x_m)$ is the observation, either univariate ($m = 1$) or multivariate ($m > 1$) depending on the number of monitored variables;
- s is the spatial extent of the observation x. The spatial extent s can be any spatial representation associated with the event, such as a geographic coordinate or a trajectory;
- τ is the temporal extent of the observation x, either given by a time instant or a time interval.

A spatiotemporal event dataset is a collection of events, $E = \{e_1, e_2, .., e_n\}$, each event producing a (multivariate) observation recorded along specific spatial and temporal context.

Notable examples of event traffic data are raw smart card validations where the observation identifies the user, the spatial extent is a coordinate (typically bound to a station) and temporal extent is the timestamp of the validation.
Trajectory data can also be seen as event data, in this context the spatial extend corresponds to the trajectory, i.e. an ordered set of coordinates, and the temporal extent to the time interval where the trajectory was travelled.

Descriptive and predictive tasks from event data generally resort to one of the previously covered spatiotemporal data structures introduced in *sections 2.1.1 and 2.1.2*. First, a spatial and temporal granularity can be fixed and estimators, such as counts and other measures which can be applied to events aggregated by the placed spatiotemporal criteria in order to produce time series. Once event data is mapped into time series, the traditional modelling, motif analysis, novelty detection and forecasting approaches can be readily applied Lin et al. (2007). Second, raw card validations and trajectory event data can be alternatively mapped into OD data structures as we previously covered.

Additional approaches to learn descriptors and predictors from raw event data are nevertheless available, including distance-based and generative approaches from event-sets (Henriques et al., 2015), dedicated neural processing architectures (Dabiri and Heaslip, 2019) and episode mining (Wu et al., 2013).

2.2 Sources of situational context

The available sources of situational context consolidated by the Lisbon City Council and to be considered in the targeted traffic data analysis tasks include:

- public events, including: a) conventions, b) festivals, c) concerts, and d) sport events. The historic and prospective events are currently sourced from two major sites: 1) the planned usage of large halls, stadiums and wide open areas in the city of Lisbon, and 2) the cultural agenda of the city;
- historic, current and 10-day forecasted weather record data sourced from three meteorological stations maintained by Instituto Português do Mar e da Atmosfera (IPMA). Weather variables with potential impact on traffic include temperature, humidity, wind intensity, precipitation, nebulosity, and visibility;
- relevant occurrences in the city, including: a) road accidents, b) medical emergencies, c) fires and floods, d) logistical help and falling structures, e) transport requests, f) conservation and complaints, and g) rescue and civil protection;
- ongoing and planned construction road works (traffic conditioning events) characterized by a set of trajectories with (possibly non-convex) interval of obstruction and accompanying details (including the number of affected ways and whether interruption is spasmodic);
- urban planning of the city with the localization of traffic generator-attractor poles, including:
  - [*commercial*] malls, commercial permits, markets, terminals;
  - [*utils*] citizen spaces, conservatories, social stores, councils, parks, public authorities;
  - [*education*] public and private schools, universities, institutes;
  - [*health*] hospitals, health centres, clinics;
  - [*sport*] sports facilities ;
  - [*cultural*] concert halls, monuments, museums, movie; theatres, amphitheatres;
  - [*transportation*] public transport networks;
  - [*leisure*] recreational spaces;

- other sources of interest including context around active modes of mobility (such as cycling roads and maps with the walking potential of urban roads and street segments), as well as calendrical context encompassing annotations associated with academic breaks, festivities, holidays and weekly-monthly-yearly seasonal factors.

2.3 Case study

The Lisbon city Council has established numerous initiatives to pursue a sustainable urban mobility. In the context of urban data monitoring and consolidation, the following sources of traffic data are currently being consolidated:

- automated fare collection data from the public transportation network, including card validations and the GPS positioning of public vehicles;
- bike sharing data from the Lisbon's public bike sharing system (GIRA[4]), including trip records per user, user feedback on bicycle's condition, bike charging information, bike malfunction and repair status, among others;
- road traffic data from three major types of sensors: i) inductive loop detectors in major road junctions in the city, offering discrete views on traffic flow; ii) individual mobile devices with GPS producing aggregated views on traffic congestions (geolocalized speed data), explored in the context of partnerships with WAZE and TomTom; and 3) privacy-compliant cameras in major roads;
- other sources: data pertaining to emerging modes of transportation, including private scooter traffic data. An entry requirement for new private operators is precisely the full disclosure of trip records.

The public transport network in the Lisbon Metropolitan Area (LMA) which includes the Lisbon city council and more 17 municipalities, covers more than 12 transport operators, being CARRIS (the major bus operator in the City of Lisbon) and METRO (the subway operator) distinctively large in passenger volumes. Apart from active modes of transportation, the providers of bus, subway, railway and inland waterway modes of transport are currently operating under an integrated fare collection system, enabled through the VIVA card initiative. The VIVA card initiative, firstly established between METRO and CARRIS, currently encompasses other major carriers operating within (or interfacing with) the city of Lisbon which operate under a single consolidated ticketing system.

3. RELATED WORK

Recent attention has been paid to the incorporation of contextual data to enhance the understanding of mobility dynamics and support traffic data analysis (Leite et al., 2020). Usually these factors are divided on whether they can be planned (Latoski et al., 2003) – including football matches, concerts, festivals, construction works, urban planning – or not (Soua et al., 2016) – weather, air quality, traffic accidents, emergencies. The former factors are often mentioned as Planned Special Events (PSE), as earlier introduced by Latoski et al. (2003). Some of the challenges of integrating spatiotemporal contextual information and its role in the development of smart cities are discussed by Sagl et al. (2015).

[4] https://www.gira-bicicletasdelisboa.pt

Different types of situational context have been considered in previous research works aiming at analyzing traffic dynamics, namely: weather and occurrences of potential relevance from Twitter data (Soua et al., 2016); accident and weather records, sport matches, festivals, and other crowded events inferred from social media data recurring to natural language processing techniques (Tang et al., 2019; Wibisono et al., 2012; Rodrigues et al., 2017; Kwoczek et al., 2014). Tempelmeier et al. (2019) introduced an approach for enriching traffic data with available sources of situational context exclusively retrieved from the web, including events from social media, weather portals, and traffic warnings.

Two major classes of context-sensitive approaches for traffic data analysis can be identified from the existing literature. First, approaches that aim to describe and predict traffic dynamics by segmenting data into chunks according to the available situational context and using only context-specific chunks for understanding and forecasting demand. Second, approaches able to embed the context directly into the models by capturing correlations with the context and using these correlations as corrections to automatically adjust descriptive and predictive models.

3.1 Context-selective analysis

Context-selective analysis of traffic data segments the available data – whether time series, OD series or raw traffic events – into chunks, each chunk sharing similar context. Data chunks are then used with the purpose of learning context-sensitive descriptive and predictive models. Hence, we can see descriptors and predictors of traffic as a set of arbitrarily different context-specific models. More advanced structures can, however, be considered (Li et al., 2001). For instance, Li et al. (2015) proposes the use of decision trees where the branches are placed in accordance with the calendrical contextual data and the roots contain context-specific models.

Kwoczek et al. (2014) proposed a method to predict and visualize traffic congestion caused by planned special events. Public events are generally characterized by two waves of congestion: people arriving and leaving the event. The authors recognized the difficulty of estimating the impact of these waves (the popularity of event) and, to address this observation, developed a distance-based approach to predict the waves using nearest neighbors from past PSEs, showing that event-sensitive predictions yield improvements. However the presented solution relies heavily on historical event data which can be scarce.

El-Assi et al. (2017) provide a multi-level model considering the impact of land use, built environment, and weather measures on bike share ridership. Similarly, Tran et al. (2015) consider the problem of predicting bike sharing system flow. To this end, they propose the use of a regression model and further consider the effects from five categories of context variables: public transport, socio-economic, topographic, bike-sharing network, and leisure variables.

3.2 Context-based corrections

The second option is to allow the learning to analyze context and its integration with traffic data, thus allowing the available context to shape the learned models. Gallop et al. (2012) explore complex serial correlation patterns between weather and bike traffic and use this effect to adjust the error terms of the classic autoregressive integrated moving average (ARIMA) models. They further suggestion that this correction can be used to affect historical data in an effort to create context-independent models to facilitate specific traffic data analysis tasks.

Rodrigues et al. (2017) introduced a Bayesian additive model (BAM) to predict the number of public transport trip arrivals in a given place. By employing a Bayesian additive framework, the proposed model is also capable of decomposing a time series into components that reflect the contributions of routine behavior and individual special events. The authors formulate the problem as a model-based machine learning approach, which seeks to create a model tailored to this particular problem, instead of using standard algorithms and transform the problem to fit them. The model is built on the assumption that a base routine component and a variable number of event components exist. The components are summed to obtain the total number of observed arrivals in a given area. Variables used to parameterize the algorithm include the start/end time of trips, its duration, if it is a multi-day event or not and the event's category. To extract the event's category, they propose an approach that uses results from web search engines. Rodrigues et al. (2017)show that this approach outperforms other models that integrate context in their predictions. The proposed method has the additional advantage of giving information of each individual event's influence, making the model highly interpretable.

To analyze the impact of social events, Tomaras et al. (2018) propose the use of a metric, influence factor, to measure the influence of an event in the nearby bike station. The influence factor is a ratio between sum the drop-off and pick-up when an event happens and drop-off and pick-up in a typical day. This factor can be used as a correction factor for descriptive and predictive models.

Ashqar et al. (2019) consider a data-fusion approach towards the analysis of context-enriched bike demand. They propose the use of random forests to rank context predictors and consider them to develop a forecasting model using a guided forward step-wise regression approach. They found that time-of-day, temperature and humidity are significant predictors.

Thomas et al. (2009) studied cycle flows from utilitarian and recreational paths in the Netherlands. A bi-level model for predicting the demand for cycling was used. The lower level describes how cyclists value the weather. The upper level is the relation between demand and this weather value. Most fluctuations are described by the model.

Principles to incorporate context within neural networks have been additionally proposed. Thu et al. (2017) propose multi-layer perceptron regressors from multi-source context data to predict bike pick-up demands in New York city considering clusters of stations based on their geographical locations and transition patterns. The proposed networks combine weather factors (temperature, wind speed, and visibility) and taxi trip records. Despite its relevance, temporal dependencies between observations are disregarded. Pan et al. (2019a) incorporate weather record data at the input layer of LSTMs to improve the prediction of bike sharing demand for balancing of distribution of bikes across stations. Results evidenced improvements against context-unaware LSTMs. Recent contributions on deep learning research also show the possibility of incorporating specific forms of calendrical and spatial awareness (Pan et al., 2019b; Casas et al., 2019).

4. PRINCIPLES FOR CONTEXT-AWARE ANALYSIS OF TRAFFIC DATA

4.1 Automated acquisition and consolidation of context data

Two major principles are suggested for the automated acquisition of situational context. First, social media, weather portals, online calendars of festivities, cultural agendas, theatre sites, and online news can be periodically explored with the aim of retrieving specific context sources of interest. Wibisono et al. (2012); Tang et al. (2019); Tempelmeier et al. (2019) gather principles towards this end. Despite the importance of web data mining, the acquisition of situational context data from the web is generally subjected to uncertainties related with data quality and availability. In addition, information pertaining to the urban poles and relevant city occurrences (such as road interventions) are generally dispersed and unstructured, being hard to infer the spatial and temporal extent of such context.

Second, in cities with well-established efforts towards the gathering and provision of situational context, the acquisition step can be simplified. In this context, periodic routines can be executed to extract context from structured or/and semi-structured sources maintained by the city Councils and other entities. Illustrating, the Lisbon city Council (CML) stores the sources of situational context listed in section 2.2 using semi-structured representations (JSON) at the Lisboa Aberta portal. As a result, the dynamic retrieval and labeling of context data can be done in a fully automated fashion (Leite et al. 2020). In addition to the listed sources of situational context, citizen notifications for reporting unexpected occurrences are further allowed recurring to "*Na minha rua*" portal and consolidated in real-time.

Some of these sources may be further subjected to preprocessing procedures to guarantee the absence of errors and the adequate imputation of missing data. In addition, the sources of events can be automatically annotated in accordance with their typology and duration. The historical and propspective duration of some of these events, such as construction works, is maintained in these repositories, For events without such information, the user can specify default rules containing expectations on the average event duration in accordance with its typology. An illustrative rule is that sport events approximately impacts entry validations at public transport stations 120 minutes before the game and up to 60 minutes after the game.

The available sources of traffic data and context data can the be consolidated by identifying the shared dimensions between sources (Melnykova et al. 2018), generally including the time dimension and space dimensions (whether associated with point, origin-destination or trajectory annotations). Considering a multi-dimensional schema, this consolidation step enables a coherent navigation throughout the records along different geographies, time periods and modes of transport. Data extraction facilities should be able to adequately index spatial, temporal and modal information for the efficient retrieval of both traffic and context data.

4.2 Context-aware data analysis

The integrative analysis of traffic data against their situational context offers unique opportunities to understand and anticipate traffic dynamics. To this end, statistically significant correlations between mobility dynamics and their situational context can be first identified and then used to enhance the targeted descriptive and predictive models.

Under the previously suggested consolidation schema, comprehensive correlations can be found between traffic records and their accompanying situational context using principles from multi-dimensional subspace clustering and relational pattern mining (Horvatic et al. 2011; Henriques and Madeira 2018; Dzeroski 2003). Once correlations are identified, they can be used as corrections to automatically adjust descriptive and predictive models (Gallop et al. 2012; Tomaras et al. 2018).

In alternative, traffic data can be segmented in accordance with the available situational context – traffic records under comparable events and calendrical, meteorological and spatial context. Context-specific slices can be then selected for learning context-specific predictors and descriptors (Li et al. 2015; El-Assi et al. 2017).

Finally, the available context can be seen as additional variables, whether static or temporal, used to augment traffic data. The context-enriched traffic data can then be subjected to machine learning approaches able to deal with heterogeneous, multivariate and spatiotemporal data structures. Neural network approaches are best prepared towards this end, as they can receive an arbitrary-high number of inputs and be architecturally-wise able to capture their inherent temporal, spatial and modal dependencies (Thu et al. 2017; Pan et al. 2019b; Casas et al. 2019)

The application of these three groups of principles – *context-driven corrections, context-driven data segmentation* and *context-driven data augmentation* – can be pursued irrespectively of the underlying data structure. Sections 4.2.1– 4.2.3 instantiate some these principles and complement them with dedicated principles for the context-aware analysis of time series, OD and traffic event data.

4.2.1 Context-aware analysis of traffic time series

Modeling and predicting demand has been the most research task. In the surveyed related work, different principles were recovered to support these tasks in the presence of situational context (Kwoczek et al. 2014; Tran et al. 2015; Rodrigues et al. 2017; Ashqar et al. 2019; Thomas et al. 2009).

Context-driven corrections have been proposed for both classical and machine learning approaches for time series analysis. In contrast, context-driven data segmentation is not suggested for classical descriptive and predictive approaches, such as SARIMA and Holt-Winters, due to the observed temporal discontinuities.

Context-enriched neural network approaches. In the context of the ILU project, recurrent neural network layering have been extended to incorporate both historical and prospective sources of context to guide traffic time series modeling and forecasting. One of the best performing architectures for the urban traffic data in Lisbon is a sequential composition of long short term memory (LSTM) components and/or gated recurrent units (GRU). To incorporate historical sources of context data, we take advantage of the fact that LSTMs are inherently prepared to learn from multivariate time series with an arbitrarily-high order. In this way, context variables can be combined at the input layer to guide the learning task by relying on masking principles. For instance, calendric masks can mark weekdays or academic periods and breaks, situational masks mark periods where events of interest may impact the demand observed at a given geography, and weather masks contain as many variables as weather attributes of interests. Prospective sources of context, including weather forecasts or planned events, may be available along the horizon of prediction.
Considering the introduced sequential composition of LSTMs as the target architecture, prospective context data can be inputted into the last LSTM component to adjust predictions. In this way, prospective context can be used as a denoiser or regularizer of the forecasted time series within predictive tasks.

4.2.2 Context-aware analysis of OD tensor data

Two major groups of principles are devised in accordance with the two major OD data structures introduced section 2.1.2: i) single OD tensors where trip statistics are inferred for the whole time periods from a given calendar, and ii) OD tensor series where the statistics pertaining to the monitored OD trips are inferred from multiple periods under a sliding time windows. Context-aware tasks along the first scenario generally aim at understanding or forecasting the impact of a given context (such as an historical or prospective event or weather context) in traffic. In this light, the time periods along a given calendar with similar situational context (events of same magnitude or comparable weather) can be marked, and two OD tensors inferred from periods marked as having comparable context and the remaining periods. Differences between OD entries can be extracted using principles from contrast set mining (Novak et al., 2009) or visually highlighted (Liao et al., 2020). Difference analysis can be done on traditional OD variables, such as passenger volume, as well as other variables capturing changes on the average time taken to complete a given OD trip or the adopted transportation modes. The gathered differences between context-sensitive and insensitive OD trips provide a simple yet effective way of supporting context-oriented mobility plans, such as decisions related with public transport strengthening needs in the presence of impactful events.

Context-aware tasks along the second scenario – OD tensor series – generally resort to the introduced principles for context-aware time series analysis. To capture the inherent spatial nature of OD data, particular attention should be placed to guarantee that the selected learning approaches are able to capture dependencies across series (OD entries). In this context, cross-series correlation factors should be explored in distance-based approaches using multivariate distances with cross-variable dependencies (Banko and Abonyi, 2012) and in neural processing approaches using ´ spatiotemporal convolutions or graph representations (Qi et al., 2019).

4.2.3 Context-aware analysis of raw traffic data

Temporal association rules can be inferred from traffic events for the purpose of describing (forecasting) available (upcoming) events (Liu et al., 2011). The learning of these rules can be conditional to historical or prospective context in order to consider meaningful temporal dependencies between events for a given context.

In addition, context-aware generative approaches placing Markov assumptions, such as hidden Markov models, can be also considered to model order dependencies between events for either describing available data or predicting the most probable event occurrences (Henriques et al. 2015). For instance, considering events to be given by smart card validations, a generative model can be learned from a single passenger in order to abstract its daily transportation choices. The same probabilistic model can then be used to understand the most probable choices for a given day under a specific context.

Context-aware analysis of raw card validations is also important to guide the estimation of entry or exit validations in carriers with non-mandatory entry or exit validations, as well as to generally impute missing data (Trepanier et al. 2007). Illustrating, depending on the weather, a given user may change its transportation preferences, affecting the patterns of transportation usage. As such, the earlier introduced segmentation-based and correction-based principles can be further applied towards entry-exit estimators.

The principles listed in section 4.2.1 can be further extended to learn descriptors and predictors from traffic events. The inherent flexibility and suitability of deep learning approaches for event data make them also a good candidate for context-aware analysis by virtue of adding both static and temporal context variables (Dabiri and Heaslip, 2019; Zhu et al., 2019). In this light, dedicated architectures, data mappings of event data to sparse time series, and graph-based views to simultaneously capture spatiotemporal dependencies are available in literature. In addition to deep learning approaches, context-aware distance-based approaches can be parameterized with distance functions able to assess the similarity between arrangements of events (Henriques 2016).

5. RESULTS

Considering the Lisbon city as a case study, this section explores the diversity of urban data sources collected by the Lisbon City Council in order to measure the impact produced by heterogeneous sources of situational context – planned events (*section 5.1*), traffic interdictions (*section 5.2*), weather records (*section 5.3*), traffic generation/attraction poles (*section 5.4*) and calendrical context (*section 5.5*) – in different modes of transport.

5.1 Planned events

Figures 1 to 3 show the impact of different events – sport matches and concerts – in different transport modes – bus (CARRIS) and subway (METRO) – using different traffic data structures – time series and OD data. Figure 1 depicts the impact generated by a soccer match in the volume of passenger boardings in two bus stops of *CARRIS* (the major bus carrier in Lisbon) near to the "*Luz*" stadium in the period between October, 2 and October, 15, 2018. The soccer match occurred in Sunday, October 7 at 17h30. The subsequent Sunday is highlight to facilitate comparisons, providing strong evidence of the disruptive nature of the event in passenger demand. The sudden passenger peak could be easily mistaken as a statistical outlier in the absence of context data.

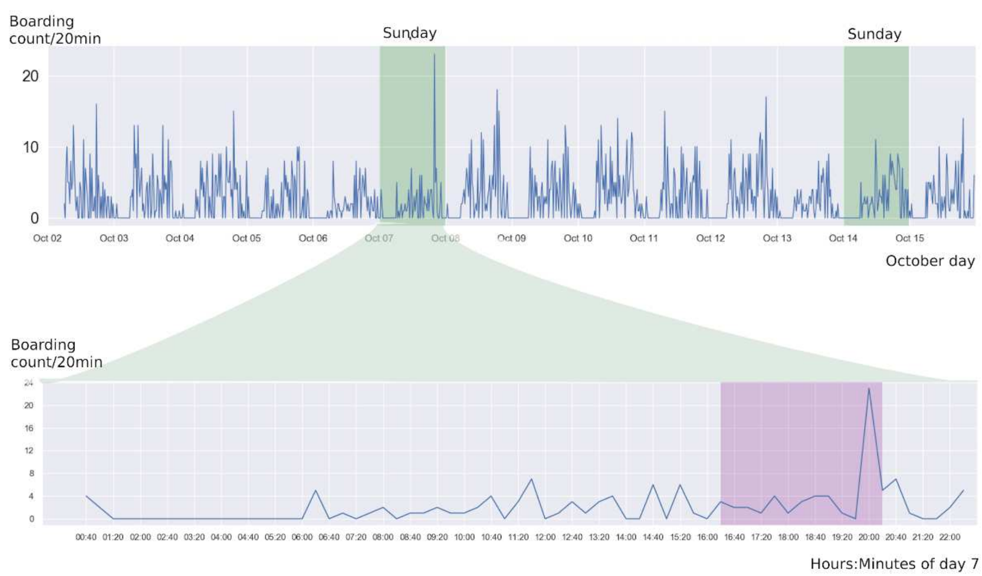

Figure 1A. Boarding passenger volumes at the *Colégio Militar* bus stop: impact of a soccer match at *Luz* stadium (Oct 7, 17h30).

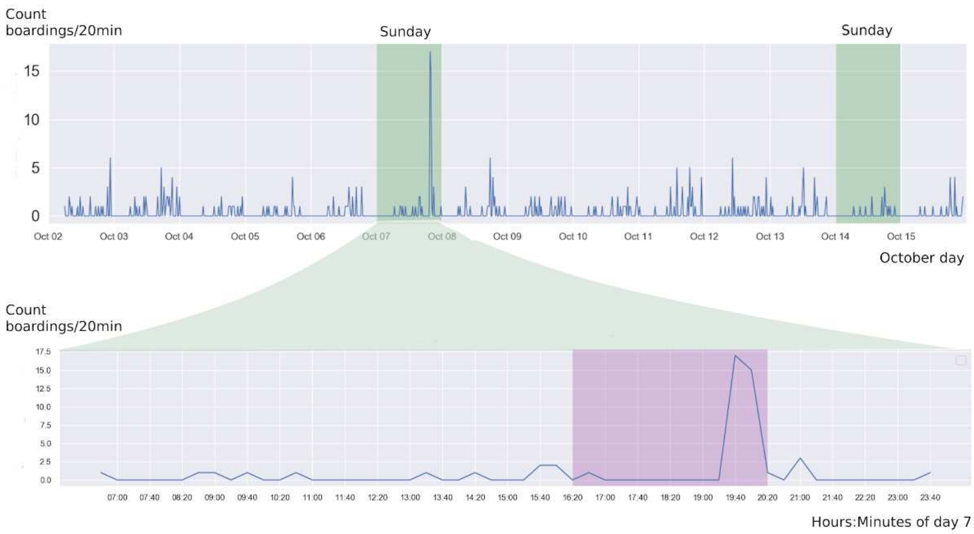

Figure 1B. Boarding passenger volume at *Estádio da Luz* stop: impact of a soccer match at *Luz* stadium (Oct 7, 17h30).

Figure 2 shows the impact of the alternative soccer matches when considering OD matrices dynamically inferred from paired entry-exit card validations at METRO (subway operator). The targeted soccer matches occurred along two consecutive Wednesdays – October, 23rd and 30th. Considering OD matrices inferred from identical weekdays in the presence and absence of the matches, the differences are clearly highlighted by the aggregate OD statistics or heatmap visualizations. In particular, Figure 2A measures the entry volume of passengers between 4:00 pm and 8:00 pm in Wednesdays at *Colégio Militar* station, while Figure 2B measures the exit volume of passengers at the same station between 9:00 pm and 11:00 pm. The difference on the volume generated between the two consecutive games is hypothesized to be associated with the popularity of matches – being the first match between *Benfica* and Lyon, *and* the second between *Benfica* and *Portimonense* – supporting the importance of considering the content of the event.

Figure 2A. Total passenger entry counts at the station Colégio Militar, October 2019 .

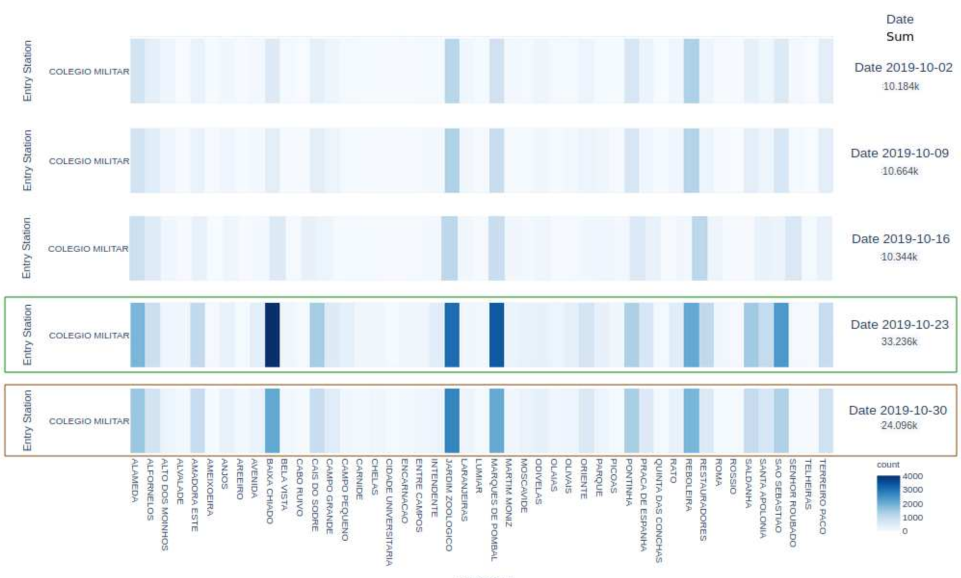

Figure 2B. Total passenger exit counts at the metro station Colégio Militar, October 2019.

A similar analysis was carried out for the *Oriente* station, near the *Altice Arena* concert hall, showing a considerable increase in passenger volume on Wednesday, October 1 2019, which cultural records revealed to be associated with a 2-hour concert by Michael Bublé. It is verified that there is again a similarity between the patterns of the following Wednesdays after the event, while in the concert data an increase of 30 percent (111 percent) of exit (entry) passenger volume against average was observed.

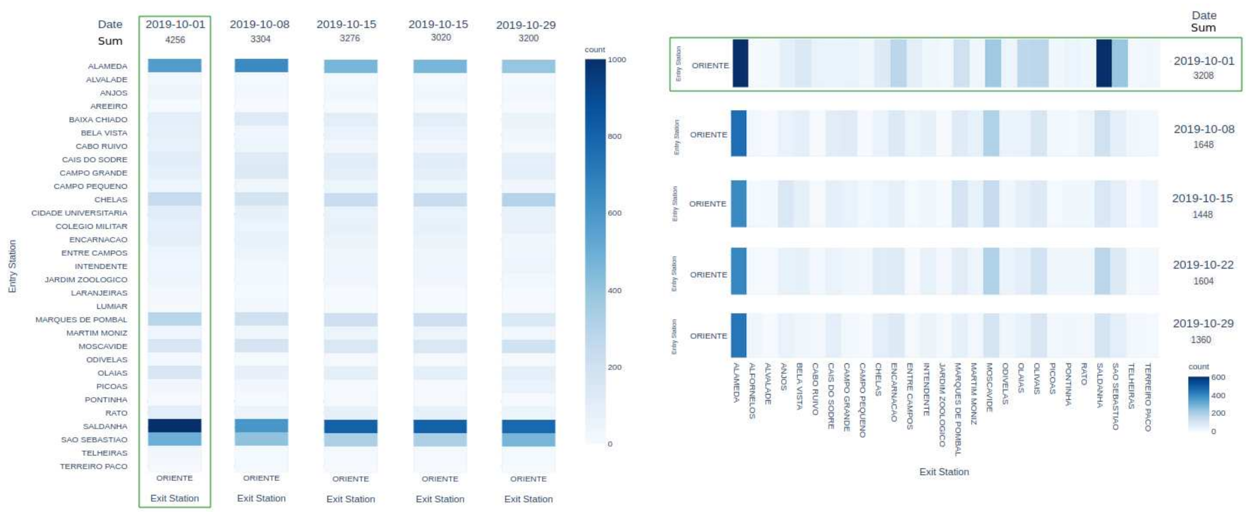

Figure 3. Total passenger exit and entry counts, respectively, at the metro station Oriente, in October 2019.

5.2 Traffic interdictions

To explore the effects the road closures and other comparable interdictions have in traffic, Figure 4 provides an analysis of the impact of three traffic interdictions in bus passengers' validations. To identify the stops that are in the vicinity of the critical road segments, we created a tool that automatically detects the potentially affected bus stops and subway stations from planned closure segments, as well as their periods under the impact of the interdictions. Figure 4 captures interdictions associated with three consecutive Sundays due to the occurrence of special events in the city, each assigned with a color in graphic visualizations. On October 14, between 8am and 14pm, occurred a marathon producing interdictions nearby the bus stop "Restauradores" , the bus stop "Av da Liberdade" and with finish line near to the bus stop "Praça do Comércio". This interdiction is represented with purple color. On October 21 at 10 am, a bike event was acclaimed at "Praça do Comércio". This interdiction is represented with yellow color. On October 28 at 12 am, an event inserted in the repository without content disclosure occurred at "Restauradores" and "Av. da Liberdade". The associated interdiction is represented in green. For this analysis, we consider passenger volume time series at three stops: "Restauradores" (Figure 4A), "Praça do Comércio" (Figure 4B), and "Av. Liberdade" (Figure 4C). Once more, these figures show strong evidence in favour of traffic disruption. On day 14, we observe zero boardings in the three locations during the marathon duration due to vehicle blockage, enclosed by peaks associated with the trips to the event and to return home. On day 21, boardings are further impacted at "Praça de Comércio" around 10 am, and on day 28 at "Restauradores" and "Av. da Liberdade" after 12 am.

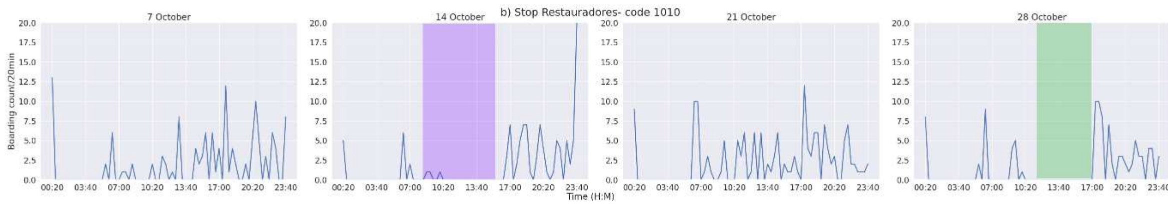

Figure 4A. Bus passenger validations at stop *Restauradores* (road interdictions signalled in purple and green).

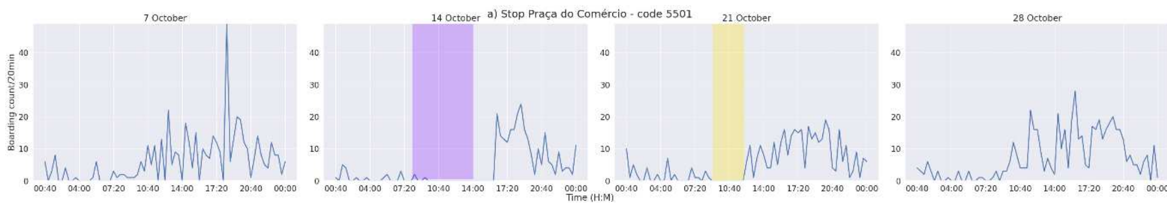

Figure 4B. Bus passenger validations at stop *Praça do Comérico* (road interdictions signalled in purple and yellow).

Figure 4C. Bus passenger validations at stop *Av. da Lisberdade* (road interdictions signalled in purple and green).

5.3 Urban planning: traffic generation poles

To assess the impact that traffic generation and attraction poles exert on the demand along the public transportation network, we provide the possibility to visualize maps of the passenger volume for the different carriers within the city of Lisbon together with a wide-diversity of poles with potential impact on traffic dynamics.

Figure 5 shows the geographical distribution of the passenger demand during 2018 for three distinct modes of transport: i) bus (CARRIS) in Figure 5A, ii) bicycle (GIRA) in Figure 5B, and iii) subway (METRO) in Figure 5C. As spatial criteria, we consider the TAZ (traffic analysis zoning) schema, providing geographical units used in transportation planning models.

Figure 5A: Bus demand distribution. Figure 5B: Bike demand distribution. Figure 5C: subway demand distribution.

Gathered views on transport demand and traffic congestion can be augmented with information pertaining to the available situational context, including city poles with capacity to attract generate traffic. Figure 6 provides a comprehensive listing of those poles including: tourist and leisure poles (Figure 6A), health-related can cultural poles (Figure 6B), commercial poles (Figure 6C) and education and institutional poles (Figure 6D).

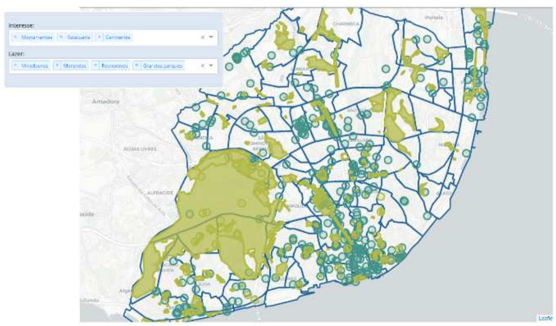

Figure 6A. Parks and leisure spaces (light green), tourist attraction poles (green).

Figure 6B. Hospitals, health centres and clinics (red) and concert halls, large theatres (green).

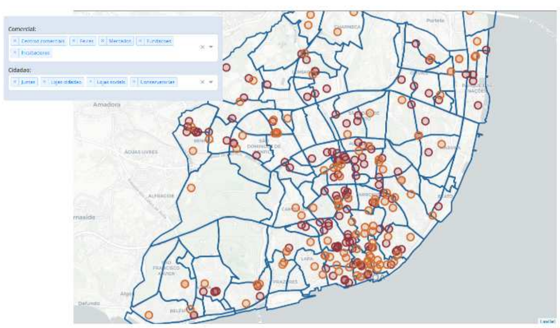

Figure 6C. Commercial poles (malls, permits, markets), citizen spaces and social stores.

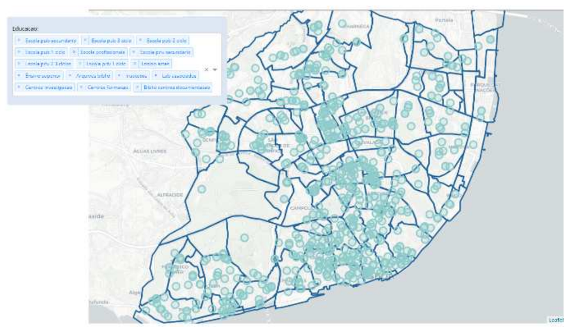

Figure 6D. Education poles (public/private schools and universities, institutes).

5.4 Meteorological context

Figure 7 assesses the impact of weather on bicycle demand. To this end, data from the public bike sharing system in Lisbon (GIRA) and three weather stations were explored. The correlation using the simplistic Pearson coefficients after removing seasonal factors in the bike check-in and check-out demand is represented in Figure 7. The GIRA stations chosen for the analysis were the stations with identifiers 406, 407, 408, 416 and 417, located in the Saldanha roundabouts and the area behind Instituto Superior Técnico. For the Pearson correlation analysis, we consider two off-peak periods - from 11h to 13h and 14h to 16h - along working days. Periods with missing data were also removed.

We can observe in Figure 7 from the depicted Pearson correlations that bike demand is positively (yet softly) correlated with the temperature and negatively correlated with the wind intensity. No delineated correlations were found for the precipitation and humidity.

			station 406	station 407	station 408	station 416	station 417	station sum
temperature	check-in	11-13	0.147	0.178	0.491	0.043	0.05	0.239
		14-16	0.127	0.255	0.05	0.05	0.088	0.138
	check-out	11-13	0.112	-0.171	0.273	0.19	-0.057	0.09
		14-16	0.303	0.082	-0.065	-0.065	0.115	0.167
precipitation	check-in	11-13	0.124	0.161	0.151	0.251	-0.07	0.161
		14-16	-0.204	0.017	0.005	-0.163	-0.011	-0.119
	check-out	11-13	-0.423	-0.146	-0.42	-0.124	-0.237	-0.414
		14-16	0.146	-0.344	-0.205	-0.267	0.287	-0.068
wind	check-in	11-13	-0.029	-0.044	-0.033	-0.248	-0.41	-0.288
		14-16	-0.122	-0.276	-0.116	-0.201	-0.251	-0.268
	check-out	11-13	-0.417	-0.412	-0.398	-0.147	-0.258	-0.501
		14-16	-0.14	-0.471	-0.404	-0.332	0.097	-0.337
humidity	check-in	11-13	0.067	0.278	0.235	-0.112	-0.008	0.111
		14-16	0.08	0.111	0.027	0.058	0.081	0.1
	check-out	11-13	-0.107	0.021	0.113	-0.24	-0.09	-0.088
		14-16	0.244	0.199	-0.159	-0.168	-0.042	0.001

Figure 7. Pearson correlation between the weather data and the check-ins and check-outs at GIRA bike stations for intervals of 2 hours in a day (11h to 13h and 14h to 16h) using data from 7/1/2019 to 28/2/2019.

Complementarily, Figure 8 shows the practical impact of considering weather context data to support forecasts of bike check-in demand at the GIRA network. To this, we consider LSTM forecasters in the absence of context and weather-enriched LSTM forecasters using the principles introduced in section 4.2.1 in order to account for both historical and prospective weather records. This analysis shows that the incorporation of meteorological context clearly guides the forecasting for the selected time series, as the magnitude of errors against true observations along the horizon of prediction significantly decreases in the presence of weather records.

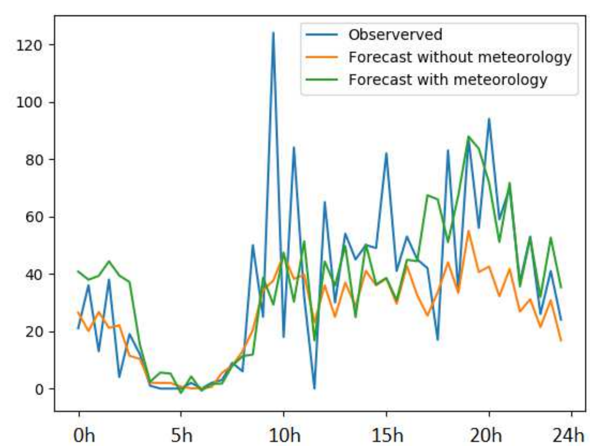

Figure 8. LSTM forecasts for an aleatory testing data instance in the absence and presence of meteorological context: impact of historical and prospective weather in guiding demand forecasting at GIRA network stations.

5.5 Calendrical context

Although, traffic data analysis generally considers calendrical constraints as sources of context. For instance, classical approaches are able to model seasonal factors capturing daily, weekly and yearly seasonalities, while segmentation in machine learning approaches considers time windows and shifting factors sensitive to similar seasonal aspects. Still, the arbitrary presence of festivities, local holidays or even the academic calendar are often disregarded, thus impact traffic data analysis.

Figure 9 provides a simple illustration of how an holiday can affect the average daily volume of passengers along different bus routes. Each route is subdivided into four time series, representing the same week day (Friday). A national holiday (Friday, October 5) is highlighted in green. The drop in passenger demand is extremely marked and should be carefully account to not hamper the learning of the target descriptive and predictive models.

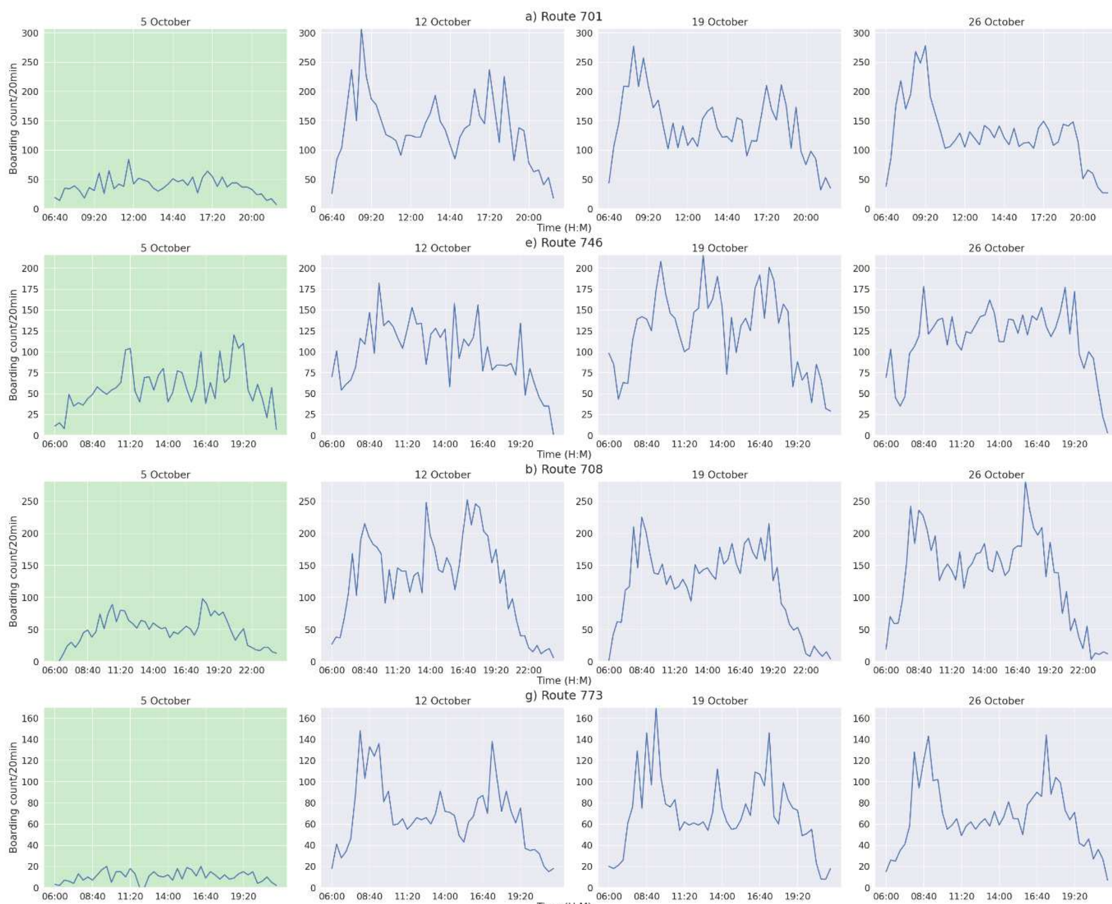

Figure 9. Effect of a national holiday (October 5) on passenger volume along four different bus routes.

Complementarily, Figure 10 provides the average daily volume of check-ins and check-outs at the stations of the Lisbon's public bike sharing network (GIRA). In particular, we show the average number of check-ins (Figures 10A and 10C) and check-outs (Figures 10B and 10D) for each weekday along three different months. The analysis of demand is shown for all the stations in Lisbon (Figures

10A and 10B) and for a cluster of stations nearby *Instituto Superior Técnico* (Figures 10C and 10D). This analysis clearly shows that traffic dynamics clearly vary among weekdays (with peaks generally observed on Mondays and Fridays) as well as throughout different periods within a year.

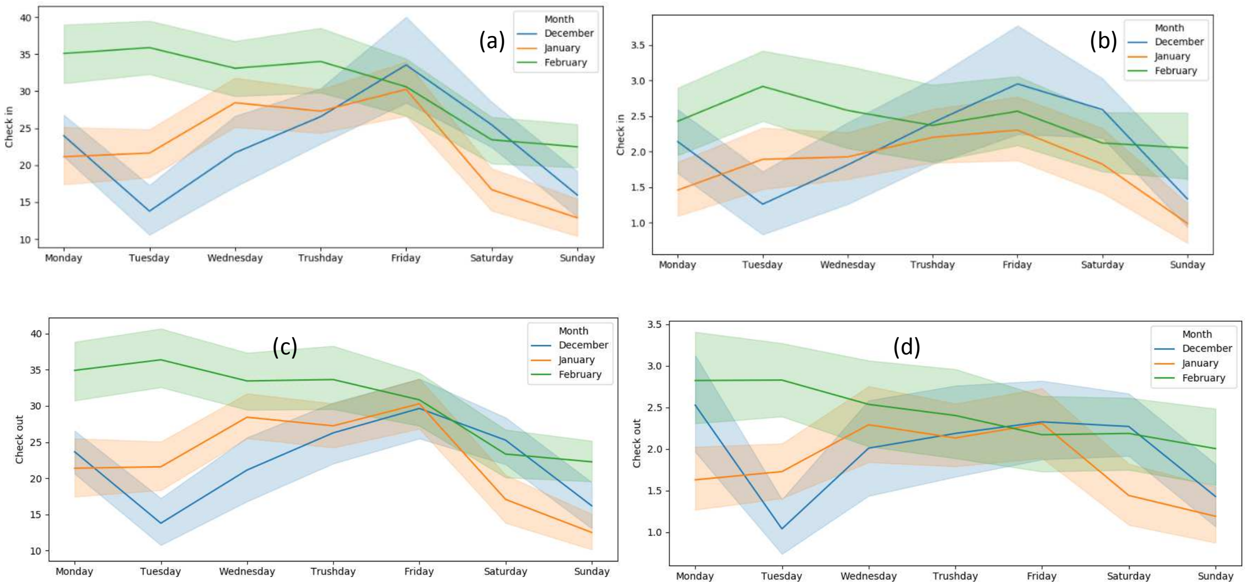

Figure 10. Volume of check-ins and check-outs for all GIRA bike stations (A and B) and the IST cluster of three stations (C and D): average number per weekday (and associated bounds) along three consecutive months.

6. CONCLUDING REMARKS

Moved by the undoubted observation that accessible context factors strongly shape public transport demand or traffic congestions, this work provides a structured view on how to incorporate heterogeneous sources of situational context into traffic data analysis for mobility planning. To this end, we survey and discuss important principles towards the integrative analysis of traffic data and its situational context in accordance with:

i) the selected sources of traffic data and their representation, whether georeferenced time series, origin-destination data or raw event traffic data;
ii) the type of situational data, whether historical or prospective, and whether situational, meteorological or/and calendrical; and
iii) the targeted task, whether descriptive or predictive.

The role placed by context-driven corrections, context-driven data segmentation and context-driven data augmentation in traffic data analysis is further discussed.
Using the city of Lisbon as case study, we gather comprehensive results that stress the importance of incorporating sources of historical and prospective context data for describing and predicting urban mobility dynamics, irrespective of the underlying data representation.

Acknowledgments
The authors thank the support of *CARRIS*, *METRO* and *Câmara Municipal de Lisboa*, in particular the attention provided by the *Gabinete de Mobilidade* and *Centro de Operações Integrado*. This work is further supported by national funds through *Fundação para a Ciência e Tecnologia* under project ILU (DSAIPA/DS/0111/2018) and INESC-ID pluriannual (UIDB/50021/2020).

Bibliography

H. I. Ashqar, M. Elhenawy, and H. A. Rakha. Modeling bike counts in a bike-sharing system considering the effect of weather conditions. Case Studies on Transport Policy, 7(2):261–268, 2019.

Z. Banko and J. Abonyi. Correlation based dynamic time warping of multivariate time series. ´ Expert Systems with Applications, 39(17):12814–12823, 2012.

W. Bao, J. Yue, and Y. Rao. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PloS one, 12(7):e0180944, 2017.

P. Cai, Y. Wang, G. Lu, P. Chen, C. Ding, and J. Sun. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. Transportation Research Part C: Emerging Technologies, 62:21–34, 2016.

S. Casas, C. Gulino, R. Liao, and R. Urtasun. Spatially-aware graph neural networks for relational behavior forecasting from sensor data. arXiv preprint arXiv:1910.08233, 2019.

S. Dabiri and K. Heaslip. Developing a twitter-based traffic event detection model using deep learning architectures. Expert systems with applications, 118:425–439, 2019.

S. Dzeroski. Multi-relational data mining: an introduction. ˇ ACM SIGKDD Explorations Newsletter, 5(1):1–16, 2003.

W. El-Assi, M. S. Mahmoud, and K. N. Habib. Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in toronto. Transportation, 44(3):589–613, 2017.

C. Gallop, C. Tse, and J. Zhao. A seasonal autoregressive model of vancouver bicycle traffic using weather variables. In Transportation Research

Board 91st Annual Meeting, number 12-2119, 2012.

J. C. B. Gamboa. Deep learning for time-series analysis. arXiv preprint arXiv:1701.01887, 2017.

R. Henriques. Learning from high-dimensional data using local descriptive models. PhD thesis, PhD thesis, Instituto Superior Tecnico, Universidade de Lisboa, Lisboa, 2016.

R. Henriques and S. C. Madeira. Triclustering algorithms for three-dimensional data analysis: a comprehensive survey. ACM Computing Surveys (CSUR), 51(5):1–43, 2018.

R. Henriques, C. Antunes, and S. C. Madeira. Generative modeling of repositories of health records for predictive tasks. Data Mining and Knowledge Discovery, 29(4):999–1032, 2015.

D. Horvatic, H. E. Stanley, and B. Podobnik. Detrended cross-correlation analysis for non-stationary time series with periodic trends. EPL (Europhysics Letters), 94(1):18007, 2011.

A. Jain, A. R. Zamir, S. Savarese, and A. Saxena. Structural-rnn: Deep learning on spatio-temporal graphs. In Proceedings of the ieee conference on computer vision and pattern recognition, pages 5308–5317, 2016.

R. K. Jain. A state space model-based method of seasonal adjustment. Monthly Lab. Rev., 124:37, 2001.

S. Kwoczek, S. Di Martino, and W. Nejdl. Predicting and visualizing traffic congestion in the presence of planned special events. Journal of Visual Languages & Computing, 25(6):973–980, 2014.

S. P. Latoski, W. M. Dunn, B. Wagenblast, J. Randall, M. D. Walker, et al. Managing travel for planned special events. Technical report, United States. Joint Program Office for Intelligent Transportation Systems, 2003.

I. Leite, A. C. Finamore, and R. Henriques. Context-sensitive modeling of public transport data. 2020.

W. Li, J. Han, and J. Pei. Cmar: Accurate and efficient classification based on multiple class-association rules. In Proceedings 2001 IEEE international conference on data mining, pages 369–376. IEEE, 2001.

Y. Li, Y. Zheng, H. Zhang, and L. Chen. Traffic prediction in a bike-sharing system. In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL '15, pages 33:1–33:10, New York, NY, USA, 2015. ACM.

Y. Liao, J. Gil, R. H. Pereira, S. Yeh, and V. Verendel. Disparities in travel times between car and transit: Spatiotemporal patterns in cities. Scientific reports, 10(1):1–12, 2020.

J. Lin, E. Keogh, L. Wei, and S. Lonardi. Experiencing sax: a novel symbolic representation of time series. Data Mining and knowledge discovery, 15(2):107–144, 2007.

W. Liu, Y. Zheng, S. Chawla, J. Yuan, and X. Xing. Discovering spatio-temporal causal interactions in traffic data streams. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 1010–1018, 2011.

N. Melnykova, U. Marikutsa, and U. Kryvenchuk. The new approaches of heterogeneous data consolidation. In 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), volume 1, pages 408–411. IEEE, 2018.

M. A. Munizaga and C. Palma. Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from santiago, chile. Transportation Research Part C: Emerging Technologies, 24:9–18, 2012.

P. K. Novak, N. Lavrac, and G. I. Webb. Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. Journal of Machine Learning Research, 10(2), 2009.

Y. Pan, R. C. Zheng, J. Zhang, and X. Yao. Predicting bike sharing demand using recurrent neural networks. Procedia computer science, 147: 562–566, 2019a.

Z. Pan, Y. Liang, W. Wang, Y. Yu, Y. Zheng, and J. Zhang. Urban traffic prediction from spatio-temporal data using deep meta learning. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 1720–1730, 2019b.

Y. Qi, Q. Li, H. Karimian, and D. Liu. A hybrid model for spatiotemporal forecasting of pm2. 5 based on graph convolutional neural network and long short-term memory. Science of the Total Environment, 664:1–10, 2019.

F. Rodrigues, S. S. Borysov, B. Ribeiro, and F. C. Pereira. A bayesian additive model for understanding public transport usage in special events. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(11):2113–2126, Nov 2017.

G. Sagl, B. Resch, and T. Blaschke. Contextual sensing: Integrating contextual information with human and technical geo-sensor information for smart cities. Sensors, 15(7):17013–17035, 2015.

R. Soua, A. B. Koesdwiady, and F. Karray. Big-data-generated traffic flow prediction using deep learning and dempster-shafer theory. 2016 International Joint Conference on Neural Networks (IJCNN), pages 3195–3202, 2016.

L. Tang, Z. Duan, and Y. Zhao. Toward using social media to support ridesharing services: challenges and opportunities. Transportation Planning and Technology, 42(4):355–379, 2019.

N. Tempelmeier, Y. Rietz, I. Lishchuk, T. Kruegel, O. Mumm, V. Miriam Carlow, S. Dietze, and E. Demidova. Data4urbanmobility: Towards holistic data analytics for mobility applications in urban regions. In Companion Proceedings of The 2019 World Wide Web Conference, WWW '19, pages 137–145, New York, NY, USA, 2019. ACM.

T. Thomas, C. Jaarsma, and S. Tutert. Temporal variations of bicycle demand in the netherlands: The influence of weather on cycling. 2009.

N. T. H. Thu, C. T. P. Dung, N. Linh-Trung, H. V. Le, et al. Multi-source data analysis for bike sharing systems. In 2017 International Conference on Advanced Technologies for Communications (ATC), pages 235–240. IEEE, 2017.

D. Tomaras, I. Boutsis, and V. Kalogeraki. Modeling and predicting bike demand in large city situations. In 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom), pages 1–10. IEEE, 2018.

T. D. Tran, N. Ovtracht, and B. F. D'arcier. Modeling bike sharing system using built environment factors. Procedia Cirp, 30:293–298, 2015.

M. Trepanier, N. Tranchant, and R. Chapleau. Individual trip destination estimation in a transit smart card automated fare collection system. Journal of Intelligent Transportation Systems, 11(1):1–14, 2007.

N. J. Van Der Zijpp. Dynamic origin-destination matrix estimation from traffic counts and automated vehicle identification data. Transportation Research Record, 1607(1):87–94, 1997.

W. W. Wei. Time series analysis. In The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2. 2006.

A. Wibisono, I. Sina, M. A. Ihsannuddin, A. Hafizh, B. Hardjono, A. Nurhadiyatna, W. Jatmiko, et al. Traffic intelligent system architecture based on social media information. In 2012 International Conference on Advanced Computer Science and Information Systems (ICACSIS), pages 25–30. IEEE, 2012.

C.-W. Wu, Y.-F. Lin, P. S. Yu, and V. S. Tseng. Mining high utility episodes in complex event sequences. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 536–544, 2013.

Y. Yang and A. V. Diez-Roux. Walking distance by trip purpose and population subgroups. American journal of preventive medicine, 43(1):11–19, 2012.

J. Zhu, C. Huang, M. Yang, and G. P. C. Fung. Context-based prediction for road traffic state using trajectory pattern mining and recurrent convolutional neural networks. Information Sciences, 473:190–201, 2019.