# FORECASTING DAILY DEMAND FOR BIKE SHARING USING WEATHER DATA

**Cantelmo Guido***
**Department of Civil, Geo and Environmental Engineering**
**Technical University of Munich**
**Munich, Germany**
g.cantelmo@tum.de

**Kucharski Rafał**
**Department of Transportation Systems**
**Cracow University of Technology**
**Cracow, Poland**
rkucharski@pk.edu.pl

**Antoniou Constantinos**
**Department of Civil, Geo and Environmental Engineering**
**Technical University of Munich**
**Munich, Germany**
c.antoniou@tum.de

## 1 Introduction

Bike Sharing Systems proved to be an effective scheme to complement public transportation and car sharing services. If usually this mobility solution is associated to sustainable urban development, reduction in greenhouse gases, health benefits and reduction of on-road vehicles, recent studies show that this scheme also brings significant economic benefits for the urban economy [1]. Properly designed Bike Bharing platforms can, in fact, improve spatial-connectivity of transport systems and deliver time-savings that far exceed commonly claimed benefits [1].

Like for other transport modes, the success of BSS depends on an optimal balance between the supply and the demand. As spatial and temporal fluctuations of bike rentals lead to a sub-optimal distribution of bikes between different urban areas, the main challenge to handle BSS in an efficient way is to understand the underlying structure of its demand avoiding supply imbalance [2, 3].

This paper contributes to the existing research on this topic by introducing a new approach to forecast the expected demand for bike sharing. First, we aggregate centre of gravity of trips and synthesize these aggregated data into vectors to identify similar daily mobility patterns and understand the systematic mobility pattern of a day-type. Then, we refine this general classification using contextual data (e.g. weather). Under this assumption, we develop a framework to infer recursive behaviour from a historical database of bike-sharing trips and to use this information to predict the daily demand for Bike Sharing.

While different approaches have already been proposed in the literature, the proposed methodology brings two main contributions. First, it is a low dimensional approach. This means that the number of parameters to be calibrated in order to achieve a good estimation is limited. In this paper, satisfactory results are achieved with a one-parameter model. Second, we show that – if weather data are included in the model – prediction capabilities can largely improve. We illustrate the method with publicly available trip data from the New York public bike system. We collect and synthesize trip data from summer 2016 (over 4 million bike trips) and use these data to make an accurate prediction of the daily demand for BSS service at a city level.

## 2 Model

In this work, we assume that the demand for bike-sharing services can be modelled as the combination of two components: a *systematic* component, composed of highly predictable travel patterns identified at the clustering level, and a *non-systematic* component, which is highly irregular and ill-predictable. Under this assumption, this section introduces a new framework to infer recursive behaviour from a historical database of bike-sharing trips. Contextual data (weather data) are then used to study their heterogeneity and daily variability and to predict the demand for a group of stations.

*This research has been partially sponsored by the European Commission through the MSCA-COFUND EuroTech program - INTERMODE project

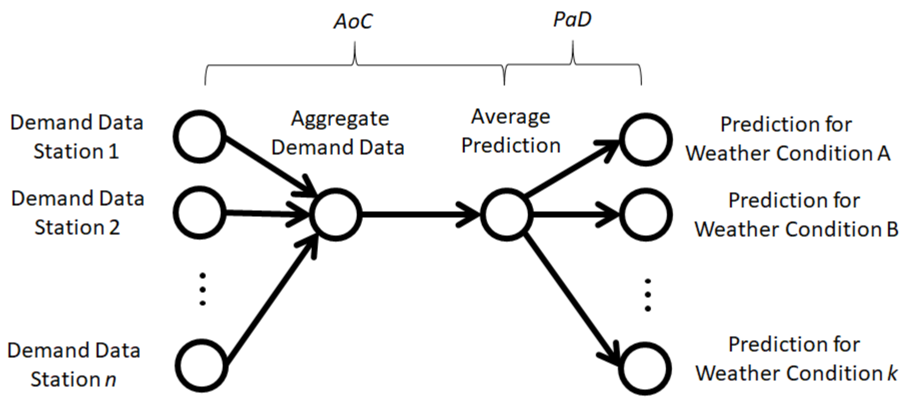


Figure 1: LD-BSS conceptual framework

The conceptual framework of the proposed Low Dimensional model for Bike Shasing Systems (LD-BSS) demand forecasting is showed in Figure 1. In essence, for a given group of stations and trip data, we calculate the so-called *Vectors of aggregated movements*. Then, days with similar mobility patterns (indetified through similarity of above vectors) are grouped to provide generic prediction of the systematic demand based on the day-type (working, non-working, ...). Finally, additional information about historical weather condition is used to study the variability of the demand and provide weather-specific demand predictions.

We can thus break down the proposed approach into two main steps, named *Aggregation and Clustering* and *Prediction and Disaggregation*.

**Aggregation and Clustering (AaC):** In this phase, we synthesize mobility data into vectors of aggregated movements. They connect the centre of gravity of trip origins with the centre of gravity of trip destinations in AM and PM peaks respectively. This synthetic representation of mobility allows reducing the set of recorded trips into a compact structure convenient for further processing. Such vector formulation allows for comparison with classical similarity measures (e.g. cosine similarity). Thanks to this, we can formulate the clustering problem and build the similarity measure between mobility patterns.

**Prediction and Disaggregation (PaD):** The *AaC* phase groups historical observations on bike-sharing trips in clusters of similar day type. However, for the same day type, different demand values can be observed. This is an expected output, as the demand for bike sharing systems not only fluctuates within the day but also changes with respect to other phenomena, among the other, season and weather conditions [4]. As a consequence, the prediction model identifies the most likely mobility pattern for a given season and meteorological condition, which are well-known to be the main elements influencing the bike sharing demand, and disaggregate this information in demand values that can be used to provide weather specific demand forecasting [4, 2].

### 2.1 Vectors of aggregated movements

Numerous trips, comprising a mobility pattern form a complex system. Typical cardinalities are thousands or even millions of trips, with various origins, destinations and start times and durations. It is far from obvious when two mobility patterns are similar.

We presume the fundamental set features like cardinality, total and mean trip duration are not enough to identify mobility pattern similarities and dissimilarities. Therefore we propose the following mapping.

We start from the concept of *gravity centre* (mass centre), an arithmetic mean of trip origins and/or destinations. Then we introduce *vectors* spanning between them. Due to the particular meaning of peak hours in the mobility patterns, we introduce a vector for *AM* and *PM* peak hours. For generic mobility pattern $M$ we introduce centre of gravity for origins (eq.1) and destinations (eq.2) and the vector of movement (eq.3) spanned between them.

$$O_M = E(O_i : i \in M) \tag{1}$$

$$D_M = E(D_i : i \in M) \tag{2}$$

$$\vec{V} = \overrightarrow{O_M D_M} \tag{3}$$

From the daily mobility we analyse trips of the $AM$ and $PM$ peaks (which are most important). Peaks are identified from the average recorded temporal profile as the two busiest morning and afternoon hours. From mobility pattern two subsets are selected: $M_{AM} = \{T_i : T_i \in M, t_i \in AM\}$, and $M_{PM} = \{T_i : T_i \in M, t_i \in AM\}$ respectively. Two vectors of movement computed for the two subsets of mobility pattern for the mapping utilized in the method:

$$M \rightarrow \{\vec{V}_{AM}, \vec{V}_{PM}\} \tag{4}$$

In fact, the above mapping transforms any number of trips into four points: *AM* origin and destination, *PM* origin and destination. Such interpretation synthesizes all main characteristics of mobility patterns.

### 2.2 Similarity measure

With such representation, pairwise comparison of days, which was troublesome for a set of recorded trips, becomes possible. We propose to compare two generic vectors $\vec{V}$ and $\vec{V'}$ with a cosine similarity (eq.5) which returns similarity from range 0 to 1, 1 for vectors of equal length and direction and 0 for either orthogonal vectors or vectors of different lengths. Importantly, cosine similarity does not use the actual location. Nonetheless, it happens to be sensitive to the day type.

$$S(\vec{V}, \vec{V'}) = \frac{\vec{V} \cdot \vec{V'}}{|\vec{V}||\vec{V'}|} \tag{5}$$

### 2.3 Clustering

Clustering procedure yields cluster membership map, i.e. labels each day with a cluster id $c(M_i)$, consequently the cluster is the subset of days belonging to a given cluster $C = \{M_i : c(M_i) = C\}$.

In this research, we apply the *agglomerative hierarchical clustering algorithm*. Specifically, we exploit the implementation proposed in *Scikit-learn*, which is an open-source library developed in Python [5]. Yet any alternative method can be applied.

### 2.4 Cluster decomposition

If a sufficient number of observations is available, results from the clustering procedure can be adopted to forecast the mobility demand $M_c$ for current cluster and associated day-types (working day, holiday, etc.) with (eq.6).

$$M_p = E(M_i : i \in C) \tag{6}$$

However, the predicted mobility patter $M_c$ for a given day should change significantly according to weather conditions and seasonality. We thus include additional information concerning the average temperature to further classify mobility patterns based on the average temperature of each day.

In this work, we adopt the concept of temperature class $\Theta_c$ to combine weather data within the proposed LD-BSS framework. Let us define $\theta_i$ as the average temperature related to a certain set of observations $M_i$. Let us also define $\theta_{low}$ and $\theta_{up}$ as the minimum and maximum average temperature for a certain class $\Theta_c$, respectively. Then, each observation $M_i$ can be associated to a certain class, as showed in (eq.7):

$$C^{\Theta_c} = \{\theta_{low} < \theta_i \leq \theta_{up} : i \in \Theta_c) \tag{7}$$

Where $C^{\Theta_c}$ represents the subset of observations belonging to the temperature class $\Theta_c$. The improved prediction model can then be written as:

$$C^* = C \cap C^{\Theta_c} \tag{8}$$

$$M_p = E(M_i : i \in C^*) \tag{9}$$

Which provides the most likely mobility pattern for a given cluster and temperature.

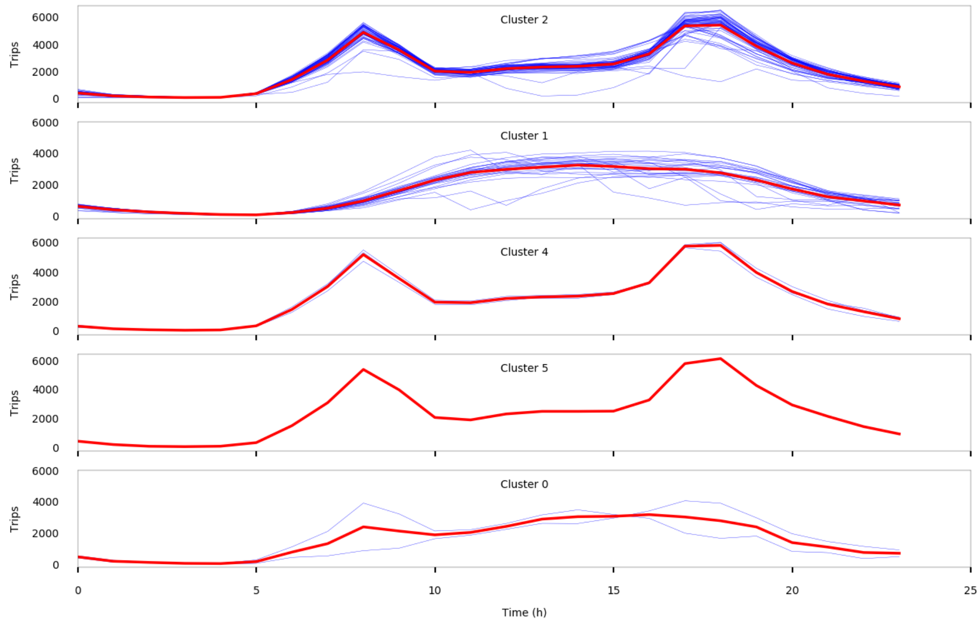


Figure 2: Demand Profiles within each Cluster - Observations (Black) and Prediction (Red)

# 3 Preliminary Results

We tested the proposed LD-BSS method with publicly available trip data from the New York City bike system.We collect and synthesize trip data between 01/06/2016 to 30/09/2016 (over 4 millions bike trips). City Bike NYC allows using one of over 12 000 bicycles to travel between more than 750 stations located in New York City and Jersey City,New Jersey.

For the same time period, historical weather data are also available(Source: https://w2.weather.gov/climate). In this analysis, data about the average temperature in New York Central Park area have been assumed to be representative of the entire study area. The database has been divided into two parts. About 70% of the data - 85 days in total - have been used to train the model.Then, this model has been adopted to predict the demand for the reaming 37 days.

Two different experiments have been performed to investigate the relationship between vectors of movements and weather data. First, to use only vector of aggregated movements during the training phase and to leverage the decomposition scheme proposed in Section 2 to further improve predictions (Experiment I). Then, we investigate the effect of combining weather data and vectors of aggregated movements within the clustering (Experiment II). On the one hand, this solution simplifies the model by removing the decomposition phase. On the other hand, additional parameters need to be properly calibrated in order to provide accurate predictions.

## 3.1 Experiment I: The effect of the decomposition scheme

This section introduces the results from the clustering procedure, meaning clustering map and average prediction model.

The aggregation and clustering phase estimated six typical day-types (Figures 2).

Concerning the clustering, results clearly identify two main groups of observations. Cluster (2) contains observations about typical behaviour during working days (Monday to Friday), while Cluster (1) mostly represents weekends and public holidays. Most importantly, Clusters (1-2) are consistent in terms of parameters not-included in the clustering procedure but significant to mobility, such as the total number of trips and temporal profile. Clusters (0,3,4,5) represent outliers, typical for agglomerative clustering technique used in the paper. They cannot be used for prediction purposes, as this approach captures only systematic behaviour.

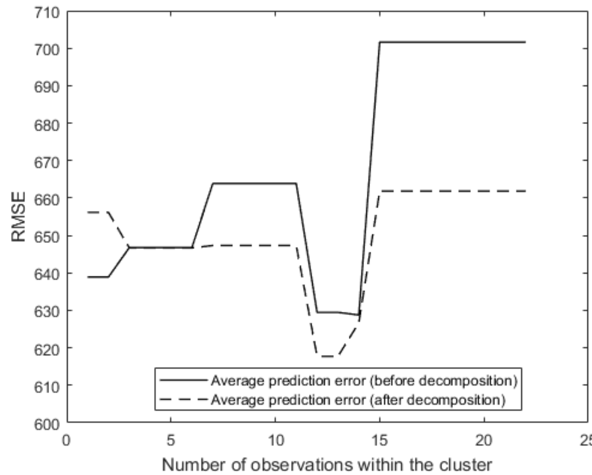


Figure 3: Average prediction error as a function of the number of observations

While Figure 2 supports the claim that vectors of movements are sufficient to infer average human mobility, it also shows that demand profiles within a cluster largely vary. This means that the average prediction model - marked with the red line in Figure 2 - could provide a biased prediction of the daily demand patterns. Additional information is required to reduce the variance within each cluster and provide accurate predictions.

We thus use the decomposition scheme proposed in Section 2 to reduce results variability without introducing additional parameters. In order to generate the refined cluster $C^*$, temperature data have been divided into six classes, which generated a total of 22 refined clusters $C^*$.

Figure 3 shows the results and, specifically, the average prediction error in terms of Root Mean Square Error (*RMSE*). The $x$ axis represents instead the precision of the model. To each refined cluster $C^*$ corresponds, in fact, a specific prediction model. However, when few observations for a certain temperature class $\Theta_c$ or for a certain day type $C$ are available, the decomposition process can return a refined cluster $C^*$ with one or zero elements. This means that the prediction model will also be calculated only on a very limited amount of data.

In Figure 3, *Number of observations within the cluster* equals to one means that even refined clusters $C^*$ with one single observation have been used to forecast the demand for BSS. By contrast, a *Number of observations within the cluster* equals to 5 means that only refined clusters $C^*$ with more than 5 observations have been accepted within the improved prediction model, while the others have been discarded.

We can observe that, when the average error over all days is computed (*Number of observations within the cluster is 1*), the average RMSE is lower when weather data are not considered within the estimation process. However, this is reasonable, as this means that the historical database has limited or no information for that combination of day type and temperature. Thus, when weather data are not available, the normal average prediction based on all available data is providing a better estimation.

On the other hand, for clusters with more than 5 observations, the improvement becomes systematic. This suggests that, if enough weather data are available, the decomposition step can largely improve the prediction. In this specific case study, results show that the reduction is particularly effective for larger errors. When the average error is calculated for refined clusters with more than 15 elements, the largest improvement is achieved.

### 3.2 Experiment II: Sensitivity

The model discussed in the previous sub-sections is a one-parameter model. The only parameter is in fact the number of clusters, which was assumed equal to 6. In this section, we evaluate model performances for different number of clusters. Additionally, we include weather information within the clustering phase to investigate the difference between using the decomposition step and relying on the clustering to directly infer refined clusters $C^*$. We thus study the average prediction error (RMSE) in three cases: (i) only vector of movements are used to create the average prediction model [*Cosine similarity*], (ii) only weather data are used [*Temperature*] and, finally, (iii) their combined effect [*Cosine+Temperature*]. Results are shown in Figure 4

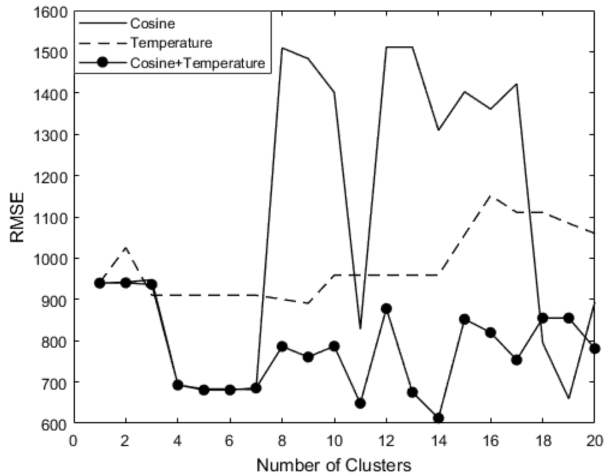


Figure 4: Average prediction error as a function of the parameters

Concerning the base-case, the first conclusion is that - for the database adopted in this study - a number of clusters between four and seven is convenient when *cosine similarity* is the sole similarity measure. Since vectors of movements mostly identity day-type and systematic mobility patterns, when too many clusters are created, the model over-fits available observations losing its predicting capabilities.

Weather data, on the other hand, provide a relatively poor estimation. For a small number of clusters, the error is significantly higher than for both other approaches and it slowly increases for an increasing number of clusters. This suggests that weather data alone are not suited to predicting BSS demand.

Finally, when weather data and vectors of aggregated movements are combined together, the model provides the best performance. *Cosine* and *Cosine+Temperature* show almost the same performance for a number of clusters equal or less than seven. The reason is that a higher number of clusters is needed to leverage weather data. In the previous experiment, starting from an initial set of six clusters, 22 refined clusters have been identified during the decomposition phase. As a consequence, when the number of clusters increases, the combined approach takes advantage of the additional information to avoid overfitting the data.

## 4 Future work

This paper introduces a Low Dimensional (*LD*) model for Bike Sharing demand forecasting. The proposed model leverages a two phases approach to exploit available information while keeping the number of parameters to be tuned low.

There are two main innovative elements. First, we propose a new method to synthesize big, multidimensional trip data sets. This synthetic description of daily mobility through vectors of aggregated movement (morning and afternoon) significantly reduces the problem size and allows to introduce pairwise distance measure and, consequently, to apply clustering methods, which was not possible on the raw trip data. Second, we combined said vectors of movements with temperature data - which is assumed to be a proxy for weather conditions in this study. We show that the combined effect of these two elements can substantially improve the accuracy of the prediction model.

Preliminary results showed in this paper provide useful insights on the potential of the proposed methodology. However, in this study, data about the average temperature in New York Central Park area have been assumed to be representative of the entire study area. Moreover, despite having positive results, the average temperature only partially captures weather dynamics. Thus, in the full paper, a more complete model evaluation will be presented. First, a richer database - 1 or 2 years - will be used in order to account for seasonal effects, which are underrepresented in this study. This will allow us to study the following elements:

1. **Spatial Granularity:** While in this short paper we exploited the LD-BSS to forecast the daily demand at a city level, in the full paper, we will investigate the possibility of estimating the demand at a traffic zone level (i.e. clustering only a sub-set of docking station).

2. **More contextual data:** Average Temperature is one of the many different contextual data that can be used withint the proposed framework. The next step is to test the LD-BSS approach with different data, such as precipitation data and snow depth.
3. **Improving demand predictions:** Preliminary results support two points. First, combining weather data and cosine similarity within the clustering phase can enhance the prediction capabilities of the LD-BSS. Second, the decomposition phase can explain the variance within a cluster and further improve predicted demand. Thus, we aim at investigate which data provide the maximum improvement during the *Aggregation and Clustering* phase and the *Prediction and Disaggregation* phase in order to enhance the LD-BSS framework.

## References


[1] C. Bullock, F. Brereton, and S. Bailey, ‘The economic contribution of public bike-share to the sustainability and efficient functioning of cities’, Sustain. Cities Soc., vol. 28, pp. 76–87, Jan. 2017

[2] L. Lin, Z. He, and S. Peeta, ‘Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach’, Transp. Res. Part C Emerg. Technol., vol. 97, pp. 258–276, Dec. 2018

[3] L. Caggiani, R. Camporeale, M. Ottomanelli, and W. Y. Szeto, ‘A modeling framework for the dynamic management of free-floating bike-sharing systems’, Transp. Res. Part C Emerg. Technol., vol. 87, pp. 159–182, Feb. 2018

[4] L. Chen, D. Zhang, L. Wang, D. Yang, X. Ma, S. Li, Z. Wu, G. Pan, T. M. T. Nguyen and J. Jakubowicz (2016, September). Dynamic cluster-based over-demand prediction in bike sharing systems. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 841-852). ACM.

[5] F. Pedregosa, G. Varoquaux, A.Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830.