# Chapter 5: Results & Discussion

Draft version

This chapter presents and discusses the results of the experiment described in Chapter 4. It is structured as follows. The first section shows the clusters that resulted from the cluster loop, along with their main characteristics, and the chosen locations of the model points. Section two presents the structure of the models that were build in the model loop, and the residual diagnostics for each them. Then, the third section focuses on the accuracies of the forecasts, and their patterns in both space and time. Finally, in the fourth section, the limitations of DBAFS are discussed, and recommendations for possible improvements are given.

## 5.1 Clustering

Figure 5.1a shows the grid overlaying the JUMP Bikes system area in San Francisco, including the centroid of each grid cell. In total, the grid contains 249 cells, each 500 meter high and 500 meter wide. Figure 5.1b shows the calculated number of pick-ups per grid cell, during the training period. On average, there were 218 pick-ups per grid cell, wich corresponds to approximately eight pick-ups per day. The maximum number of pick-ups in a grid cell was 1985 (i.e. 71 per day on average), while in 25 of the 249 grid cells, there were no pick-ups at all. It can be seen that high counts of pick-ups occured in the grid cells along the diagonal axis from south-west to north-east. Mainly in the south-eastern corner of the system area, the usage intensity was very low.

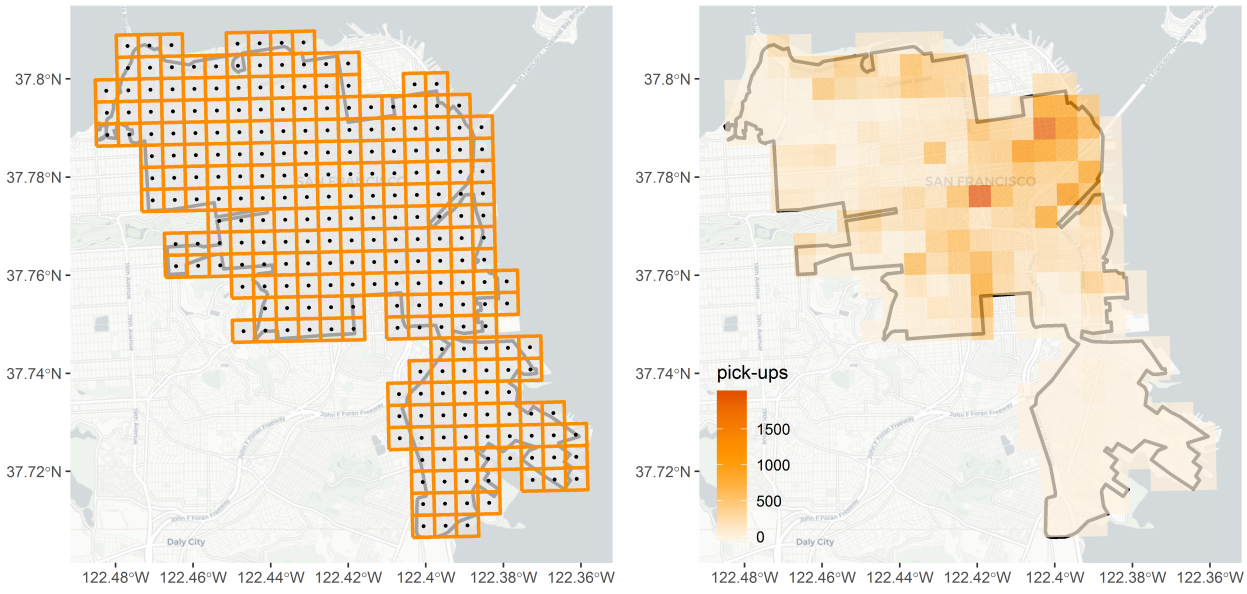


Figure 1: a) grid overlaying the system area; b) number of pick-ups per grid cell

Figure 5.2 shows the temporal patterns of the usage data, with the pick-ups per day of the week, and per hour of the day. Friday was the day with on average the most pick-ups, while Saturday and Sunday had the least. The busiest hours of the day, where 8:00 and 9:00, during morning rush hours, and 16:00 and 17:00, during afternoon rush hours. The lowest numbers of hourly pick-ups, as expected, occured during the night.

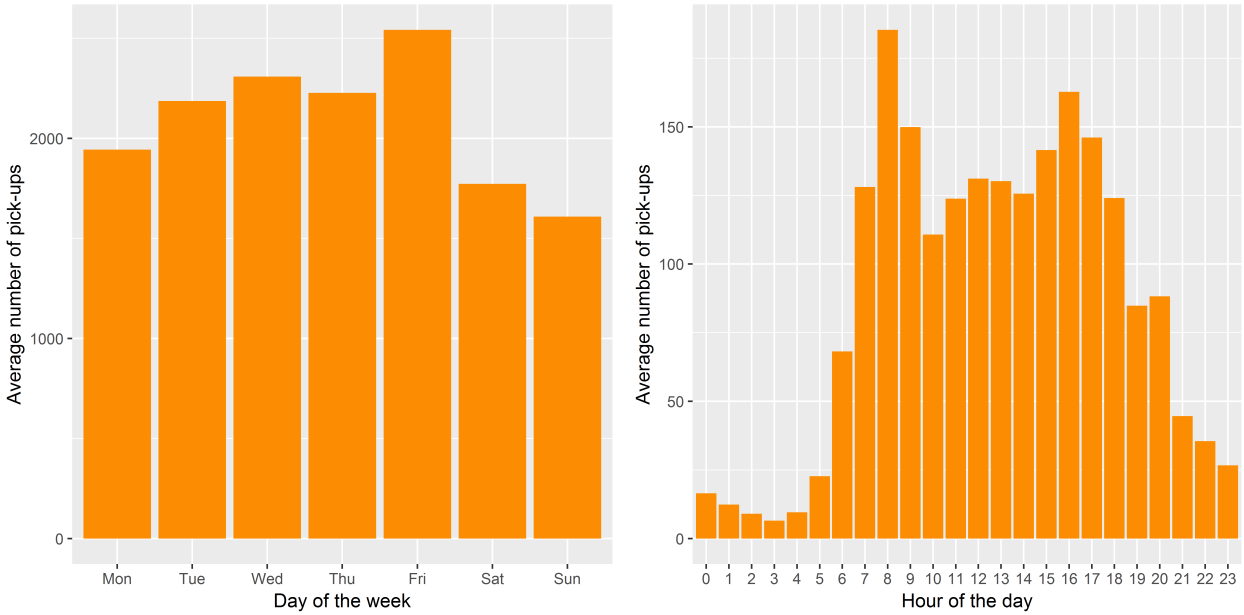


Figure 2: a) pick-ups per day of the week; b) pick-ups per hour of the day

Recall that for each grid cell centroid, a time series of historical distance data was queried, and that the normalized, average weekly patterns in these data were clustered using spatially constrained hierarchical clustering. The automatic procedure of defining the number of clusters $k$ and the mixing parameter $\alpha$, lead to a definition of $k = 4$ and $\alpha = 0.6$. This resulted in a partition containing four fully spatial contiguous clusters. The geographical outlines of these clusters are shown in Figure 5.2a. The centroid of each cluster, weighted by the number of pick-ups in the corresponding grid cells, are shown in Figure 5.2b. These weighted centroids serve as the model points in DBAFS.

Roughly speaking, and based on a large study of neighbourhood indicators in San Francisco (San Francisco Department of Public Health 2014), the four clusters can be characterized as follows. The orange cluster covers the Bayview/Hunters Point neighbourhood, which is a rather isolated area, with a high percentage of low-income households and relatively high crime rates. The blue cluster forms the city center of San Francisco, containing the neighbourhoods with the highest population densities, but also with a relatively high job density compared to the residential density, and large areas zoned for commercial usage. The purple cluster mainly contains neighbourhoods where the residential density is high compared to the job density, and the area zoned for commercial usage is relatively small. Finally, the green cluster covers the Presidio Park, a recreational area with few inhabitants, and a relatively high number of bike lanes. For the sake of clarity, the orange, blue, purple and green clusters are from now on referred to as the *Bayview*, *Downtown*, *Residential* and *Presidio* clusters, respectively. Consistently, the four corresponding model points will be called the *Bayview*, *Downtown*, *Residential* and *Presidio* model points, respectively.

Table 1 presents some descriptive statistics of the time series, averaged per cluster, and averaged over the whole system area. From the 249 grid cells, more than a hundred are located within the Residential cluster, while the Presidio cluster is by far the smallest of the four. During the training period, the nearest available bike was on average located 619 meters from the grid cell centroids. In the Bayview cluster, however, this was more than one kilometer, a difference of almost a factor two compared to the Downtown cluster, and even more compared to the Residential and Presidio clusters. The Bayview cluster also showed the largest variation in the data, with a high average standard deviation compared to the other clusters, and an average range that spanned more than four kilometers. This can possibly be explained by the low usage intensity of the bike sharing system in this part of the system area. When the number of bikes in an area is low, the nearest available bike and the second nearest available bike are more likely to be far away from each other.

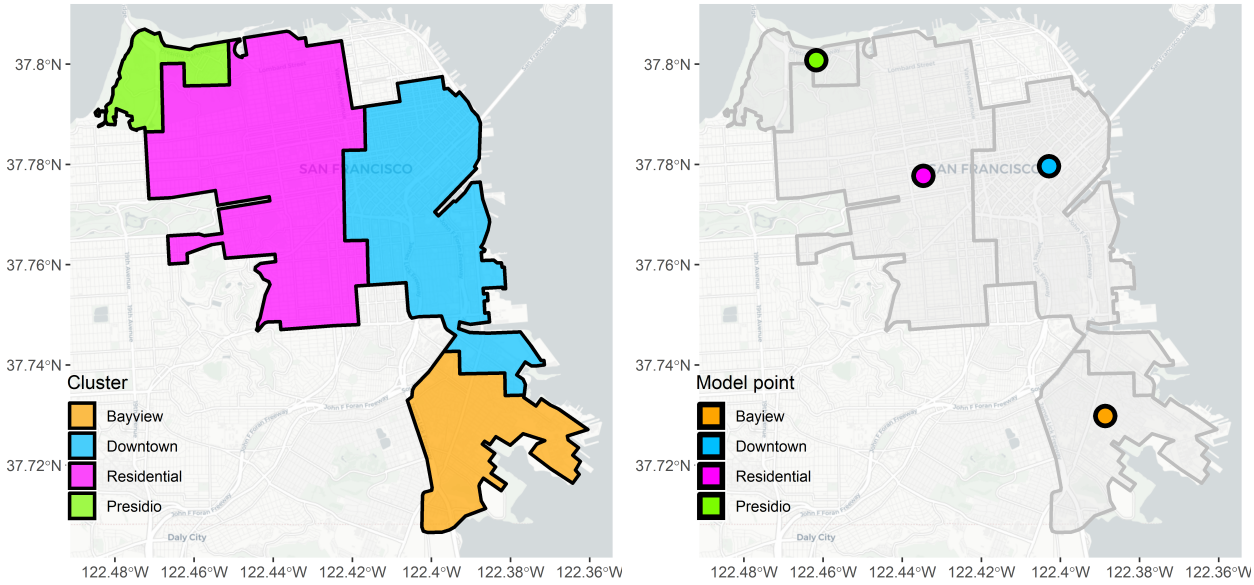


Figure 3: a) geographical outline of the clusters; b) geographical locations of the model points

In that case, when the closest of them gets picked-up, the distance to the nearest available bike will suddenly increase substantially. The other way around, when all available bikes are far away, and one bike gets dropped-off inside the area, the distance to the nearest available bike will suddenly decrease substantially.

Although not as extreme as the Bayview cluster, also the other clusters had on average high ranges when compared to the mean and standard deviation. However, the standard deviation itself turned out to be rather small relative to the mean. This implies either the presence of outliers, or population distributions with thin, but wide tails.

The first order autocorrelation measures the average dependency between data values at time $t$ and corresponding data values at time $t-1$. In the whole system area, this dependency was strong, especially in the Bayview and Presidio clusters. These high autocorrelation values are important, since they imply that it is reasonable to use past observations when forecasting future ones. However, the calculated spectral entropy values show that in general, the data are also very complex, and the forecastability is low. This mainly concerns the Downtown and Residential clusters, which contain, as could be seen in Figure 5.1b, the areas where the pick-up density is high. In such areas, the data are more dynamic, since bikes get picked-up and dropped off constantly, and the location of the nearest available bike will change often. In most cases, the more dynamic the data, the harder to forecast.

Figure 5.3 shows the normalized, average weekly patterns of the time series, averaged once again per cluster. The patterns can be explained intuitively. The Bayview cluster has a low usage intensity, and although there are peaks in the data every day, a clear and consistent pattern is absent.

The Downtown cluster has a high density of jobs and commercial activities. During working hours, the demand for bikes is low, which leads to a high number of available bikes, and consequently, short distances to the nearest available bike. In the afternoon, just after working hours, the demand starts increasing, and it gets harder to find an available bike nearby. This peak in the data continues during the evening, when the activity in the commericial zones is high. Later in the evening, the demand decreases again. A clear difference between weekdays and weekends, can not be seen.

The Residential cluster shows the exact opposite pattern. In the morning rush hours, commuters use the bike to get to work, and not many available bikes are left in the residential areas. Hence, in those areas,

Table 1: Descriptive statistics of the grid cell centroids distance data

| | $N$ | $\mu$ | $range$ | $\sigma$ | $\rho(1)$ | $H$ |
|---|---|---|---|---|---|---|
| **Total** | 249 | 619 | 2726 | 155 | 0.82 | 0.77 |
| **Bayview** | 46 | 1080 | 4021 | 155 | 0.95 | 0.67 |
| **Downtown** | 81 | 557 | 2551 | 77 | 0.77 | 0.81 |
| **Residential** | 103 | 490 | 2410 | 112 | 0.79 | 0.81 |
| **Presido** | 19 | 462 | 2057 | 137 | 0.92 | 0.68 |

*Note:*
Except $N$, all metrics are calculated for each time series seperately, and averaged afterwards.
[1] $N$ is the total number of grid cell centroids
[2] $\mu$ is the mean of the data, in meters
[3] $range$ is the difference between the maximum and minimum data value, in meters
[4] $\sigma$ is the standard deviation of the data, in meters
[5] $\rho(1)$ is the first order autocorrelation, see section 2.2.1
[6] $H$ is the normalized spectral entropy, see section 2.2.3

the distance to the nearest available bike is higher during working hours. In the afternoon, commuters come back from work, and leave the bikes in the residential areas, causing a decrease in distance to the nearest available bike. Hence, the distance data peaks during working hours. In the weekends, the peaks seem to be slightly lower, but this difference is not as large as might have been expected. They do happen later on the day, corresponding to the same periods as the Downtown cluster.

Finally, the Presidio cluster is mainly a recreational area. There are a lot of bikes, but during weekdays, they are used less, leading to small and relatively constant distances to the nearest available bike. In weekends, and mainly on Sunday afternoon, the usage intensity is high, and it takes longer to find an available bike.

## 5.2 Model building

Figure 5.4 shows the time plots of the distance data that were queried for each of the model points in Figure 5.2b, with the dark grey shaded areas representing weekends. The plots endorse the findings in the previous sections.

The data corresponding to the Bayview model point show large variation, interspersed with flat sections, and lack a clear repeating pattern. The data corresponding to the Downtown and Residential model points are most dynamic. A daily pattern shows for both of them. However, in both datasets, this pattern is far from smooth, and the daily peaks vary considerably in height from day to day. This underlines the high spectral entropies that were found for these clusters. A clear difference between weekdays and weekends, can not be seen. The Presidio model point shows the most constant data, with a low mean and long flat sections. Sunday afternoons stand out clearly in most of the weeks, but not in all of them. The last sunday, for example, shows only a minor peak in the data. In less extent, this also applies to the other clusters, with lower peaks than normal, in the last weekend. Finally, none of the datasets contain missing values, and clear evidence for non-constant variances is not present.

The structures of the models that were fitted on the historical data of the four model points, are shown in Table 5.2. The automatic seasonality detection resulted in a daily seasonal pattern for both the Downtown and the Residential model point. That is, including a weekly seasonality, did not increase the overal accuracy of the cross-validation process. As expected, a weekly seasonal pattern was found for the Presidio model point, an no seasonaity for the Bayview model point. The ARIMA($p$, $d$, $q$) models for the Bayview and Downtown model points, have a relatively high number of autoregressive terms, while for the Presidio model point, the number of moving average terms is high. For the Residential model point, the best fit was obtained by only including one autoregressive and one moving average term. All datasets passed the KPSS test for

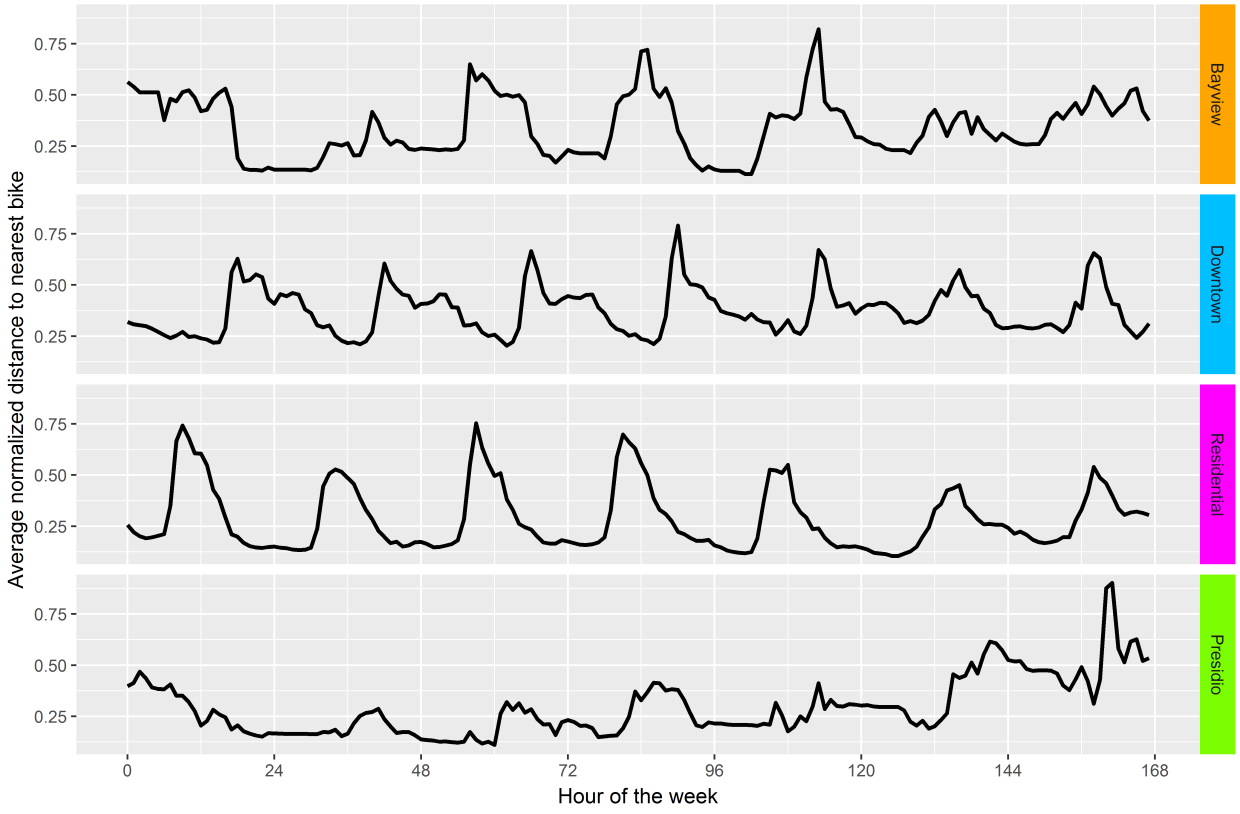


Figure 4: Normalized, average weekly patterns of the grid cell distance data, averaged per cluster

stationarity after one differencing operation. The full details of the components and fitted models, including parameter estimates and decomposition plots, can be found in Appendix B.

Table 2: Model structures

| | STL | ARIMA | | |
|---|---|---|---|---|
| | seasonality | $p$ | $d$ | $q$ |
| **Bayview** | none | 3 | 1 | 1 |
| **Downtown** | daily | 3 | 1 | 2 |
| **Residential** | daily | 1 | 1 | 1 |
| **Presidio** | weekly | 1 | 1 | 4 |

Figure 5.5 shows the residuals of each model, plotted over time. All models have residuals with an approximately zero mean, and the variances look approximately constant. Comparing the residual time plot with the data time plot of Figure 5.4, it can be seen that in for the less dynamic data of the Bayview and Presidio model points, the models struggle to find a good fit for the peaks and valleys in the data, while the flat sections are explained accurately.

The autocorrelations at several time lags in the residuals are shown in Figure 5.6. Since the data have a temporal resolution of 15 minutes, 96 time lags correspond to one day, and 672 time lags, the total span of the x-axis in the figure, to one week. The dotted orange lines form the lower and upper 95% confidence bounds, assuming a normal distribution. This means that the residuals are considered to be a realization of a white noise process when at least 95% of the autocorrelation values fall within these bounds. It is

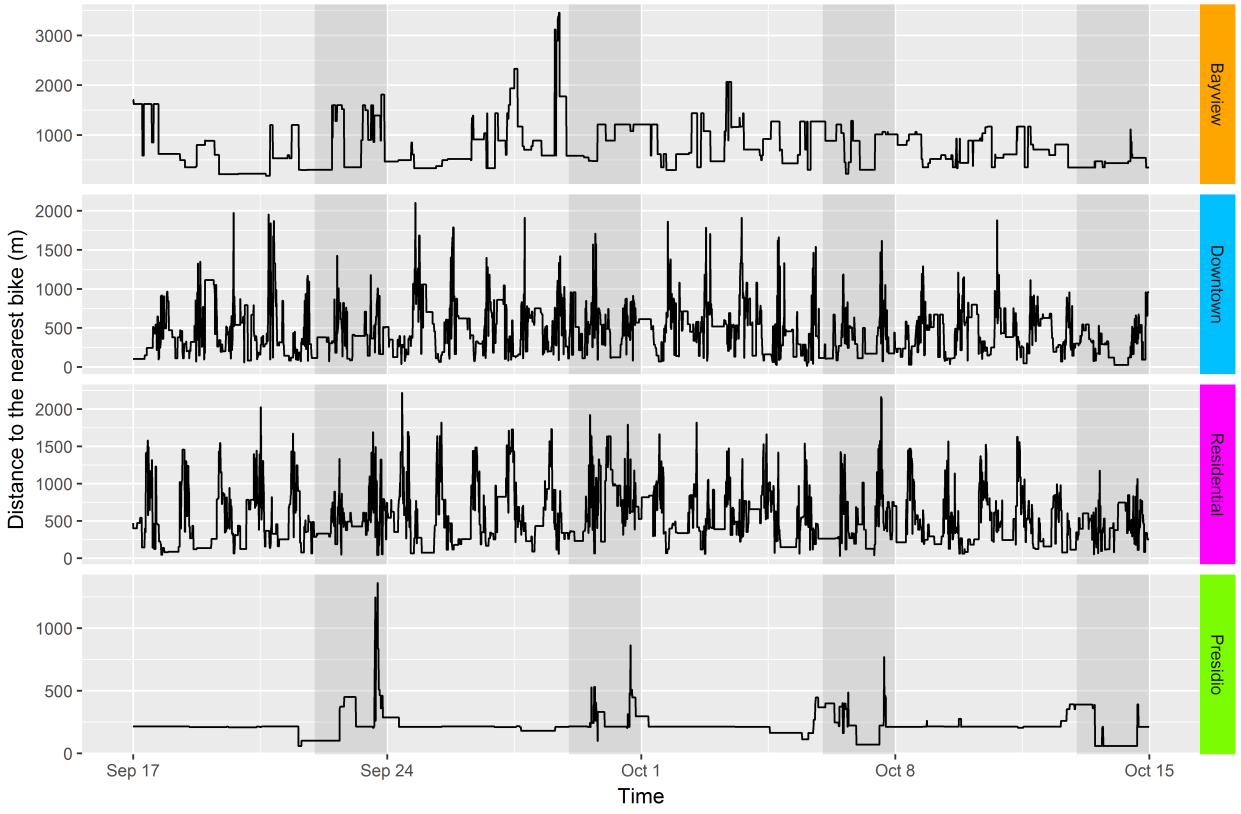


Figure 5: Time plots of the model points distance data

important to note here that when working with real-world data, finding perfectly random model residuals is an exception, especially when the data have a high entropy. Taking that into account, the autocorrelation plot of the Bayview, Downtown and Residential models look good, and their residuals seem to approximate white noise.

However, for the Presidio cluster, the residual autocorrelation has a strong peak at lag 672, corresponding to one week. Recall that the data of the Presidio model point was relatively flat during the weekdays, and spiky in the weekends. These spikes, however, varied considerably in amplitude from week to week. The weekly seasonal component that was substracted from the data, accounts for the recurring patterns, but can not completely capture the differences from week to week. Therefore, errors during the 'spiky' weekends, will still be higher than during the 'flat' weekdays, causing autocorrelation in the residuals. With just a stochastic time series model, it is hard to solve this. Including external variables that explain the variation, could be an option, and will be discussed in section 5.4.

Finally, Figure 5.7 shows the histograms of the model residual distributions. As expected, for the Bayview and Presidio models, most values are clustered closely around the zero mean, with the tails being extremely thin and long, especially for the Bayview model. The residuals of the Downtown and Residential models follow a distribution that comes closer to a normal one, but also here, the tails are wide.

As discussed in section 2.4.2.4, using Gaussian likelihood is sensible even when non-normally distributed residuals show up. Of course, it could be a possibility to try different likelihood functions, but this will make the process much more complex and much slower, for, probably, only a little gain in forecast accuracy. The non-normality of the residuals does have an effect on the validity of the prediction intervals, however. When 95% prediction intervals are calculated, assuming normality of the forecast distribution, they can not be interpreted as such. This issue will be covered further in the next section.

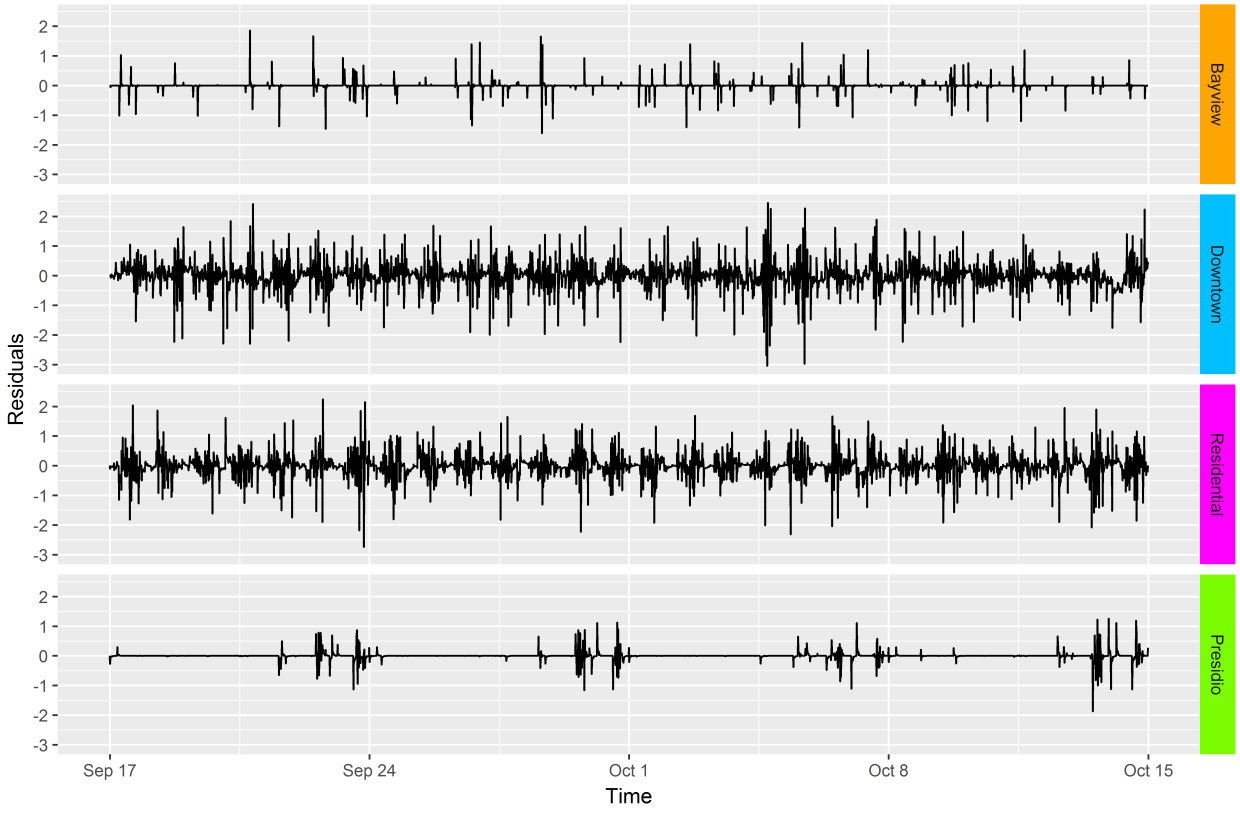


Figure 6: Time plots of the model residuals

## 5.3 Forecasting

Figure 5.8a shows the spatial distribution of the 500 test points. As planned, areas with high usage intensity have more test points, with 94% located in the Downtown and Residential clusters,and only the minimum of ten test points in the Bayview cluster. Figure 5.8b shows the temporal distribution test points. All days in the test week are well covered, with less test points during working times and in the night, and more during the morning rush hours and in the evening. On weekend days, there is only one strong peak, around noon. Furthermore, it can be seen that the morning peak on November 1st is somewhat lower compared to the other weekdays. This may be, because it is the morning of All Saint's Day, following the Halloween night. For the full information on the test points, with all unique location-time combinations, see Appendix A.

The first row of Table 5.3 lists the RMSE's, averaged over the whole system area, of the forecasts produced by DBAFS, and of the forecasts produced by the baseline system, NFS. DBAFS clearly outperforms NFS, by producing forecasts with errors that are on average 31% lower. Furthermore, the range of error values is much lower for DBAFS, than for NFS. The minima are comparable, but NFS produces forecasts with error values up to 1644 meters, while DBAFS never exceeds 1004 meters.

Regarding the spatial patterns of the forecast errors, the remaining rows of Table 5.3 show the RMSE's averaged per spatial cluster. With NFS, the lowest errors are obtained in the Bayview and Presidio clusters, where the data are less dynamic. In the Bayview cluster, NFS gives the same results as DBAFS, and in the Presidio cluster, DBAFS performs only slightly better than NFS. For DBAFS, however, the lowest errors are not found in those clusters, but in the highly dynamic Downtown cluster. Here, DBAFS gives errors that are 40% lower than those of NFS. In the Residential cluster, there are larger errors than in the Downtown cluster, but also here, DBAFS outperforms NFS with errors that are 23% lower. It shows the strength of

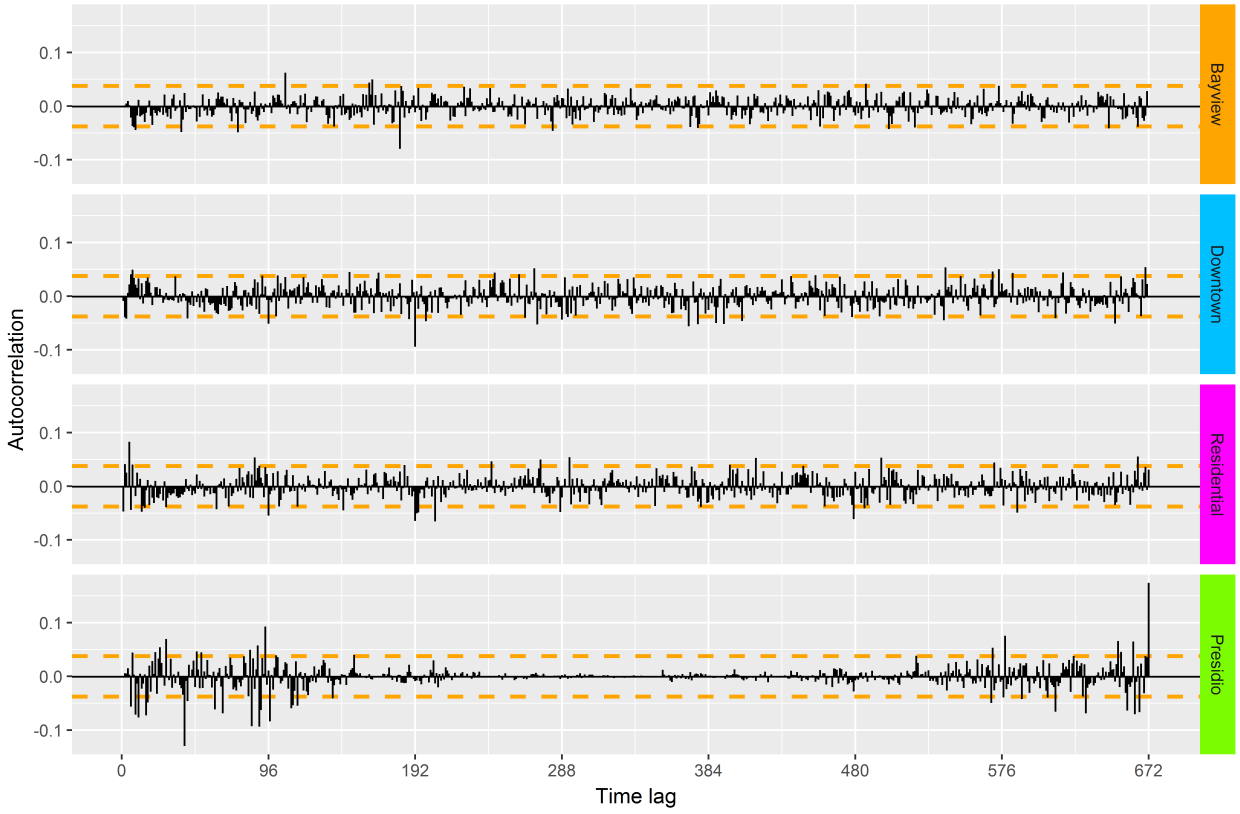


Figure 7: ACF plot of the model residuals

DBAFS in forecasting dynamic data, when compared to NFS.

Table 3: Forecast RMSE's, in meters

| | | DBAFS | | | NFS | | |
|---|---|---|---|---|---|---|---|
| | n | mean | min | max | mean | min | max |
| **Total** | 500 | 282 | 38 | 1004 | 408 | 37 | 1644 |
| **Bayview** | 10 | 389 | 38 | 1004 | 389 | 38 | 1004 |
| **Downtown** | 259 | 248 | 122 | 523 | 414 | 116 | 927 |
| **Residential** | 211 | 317 | 97 | 705 | 411 | 37 | 1644 |
| **Presidio** | 20 | 299 | 80 | 577 | 320 | 175 | 699 |

Regarding the temporal patterns of the forecast errors, Figure 5.10 shows the RMSE's averaged per hour of the day, and per forecast lag. The lowest forecast errors occur during the night, when the usage intensity of the system is low. During the day, higher errors occur, with peaks at the morning peak hour, around noon (i.e. the peakhour in the weekend) and after working hours. This patterns are similar for NFS, but with higher RMSE's at each hour. The forecast errors of both methods rise steeply directly after the first forecast lag, but for NFS, this increase is much larger than for DBAFS. What strikes, is that the RMSE does not increase constantly when the forecast horizon gets larger. From the forecasting lag of 12 hours, the errors for both DBAFS and NFS decrease again. Moreover, at a forecasting lag of approximately 18 hours, the RMSE of the DBAFS forecasts is, on average, back at almost the same level as the one at a forecasting lag of just 15 minutes. This conspicuousness can be explained as follows. Most of the simulated forecast requests are made at times with a high usage intensity, that are hard to forecast. The first forecast lags, will

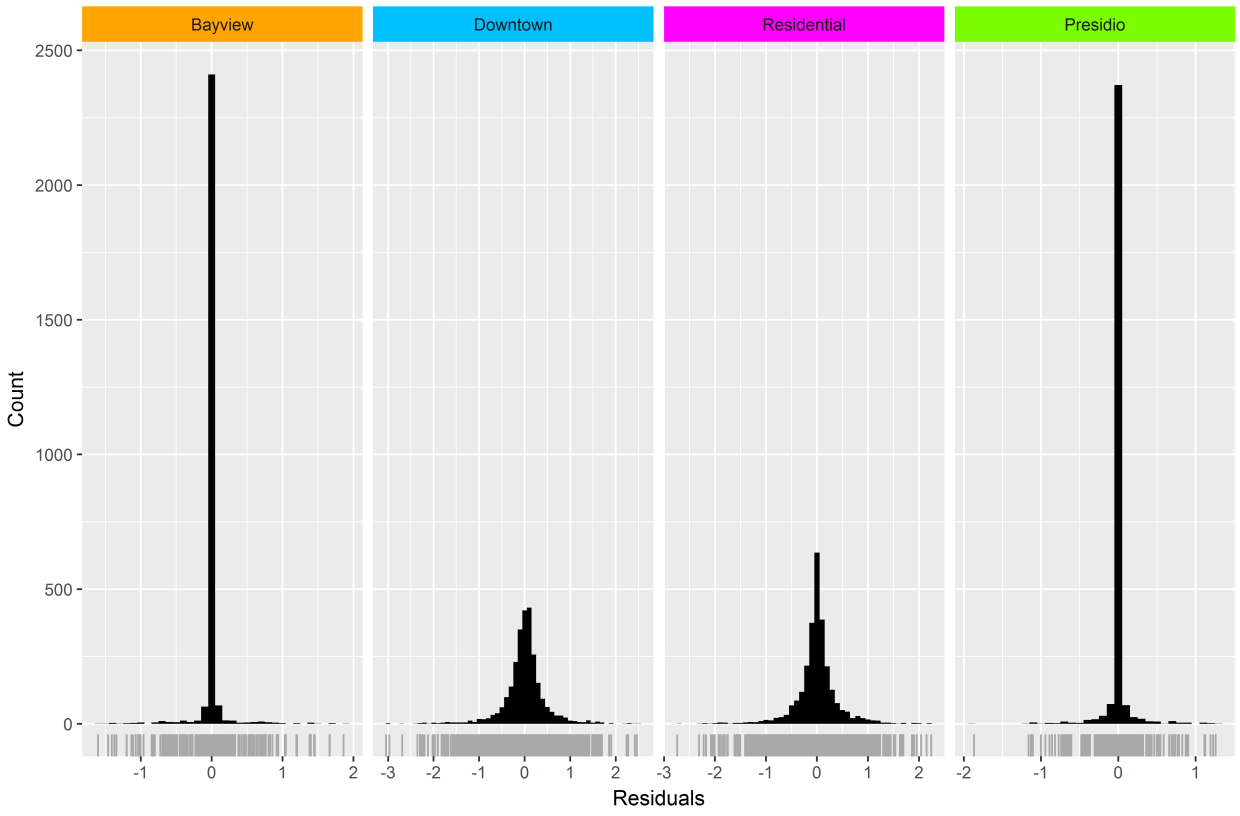


Figure 8: Histograms of the model residuals

still correspond to high usage times, but after a while, forecasts will be made during night time. As could be seen in Figure 5.10a, these night time forecasts have much lower errors. Therefore, it can happen that, despite the length of the forecasting window, 'far-ahead' forecasts have lower errors than 'close-by' forecasts.

## 5.4 Limitations & Recommendations

*NOTE: write about limitation of DBAFS, and give recommendation how things could be improved. Cover at least the following points: - the low forecastability of dockless bike sharing data in general, compared to station based bike sharing systems - methods that can include exogeneous variables. For example, the new forecast packages in R, that are still in development, like FASSTER, and a Dynamic Harmonic Regression with seasonality in Fourier terms and an ARIMA model for the errors - talk about possible exogeneous variables, based on the literature. Weather, special events like football matches, holidays, etcetera - machine learning possibilities - higher density of model points - clustering based on ARIMA parameters instead of on raw data*

### 5.4.1 Limits of forecastability

### 5.4.2 Exogeneous variables

### 5.4.3 Non-normal distributions

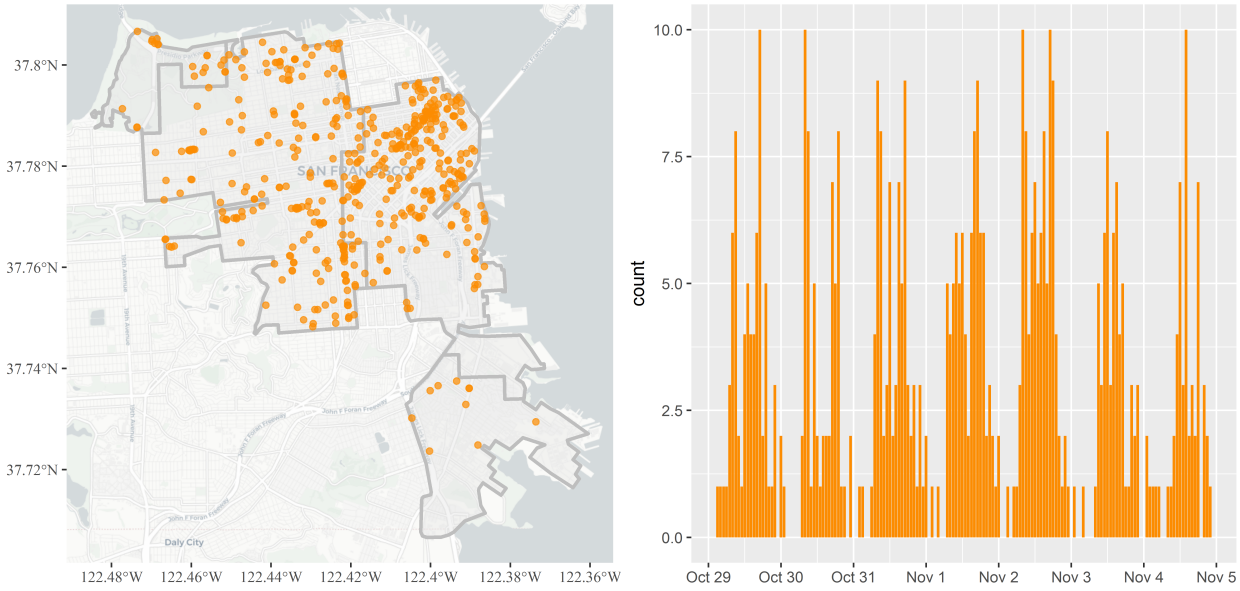


Figure 9: a) geographical locations of the test points; b) time stamps of the test points, counted per hour

# References

San Francisco Department of Public Health. 2014. "The San Francisco Indicator Project." https://www.sfindicatorproject.org/.

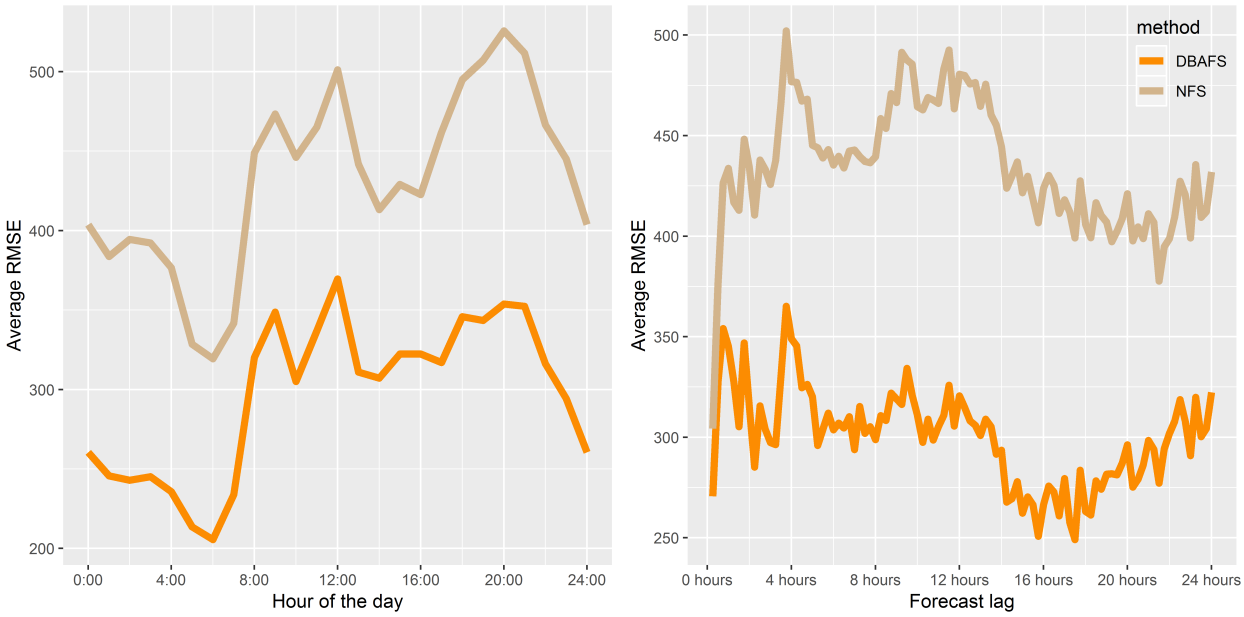


Figure 10: a) RMSE averaged per hour of the day; b) RMSE averaged per forecast lag