

Locational Demand

Tanner Fiez

1 Spatial and Temporal Heterogeneity

Parking occupancy across the Belltown neighborhood in Seattle displays both temporal and spatial heterogeneity. Figure 1a demonstrates the spatial heterogeneity by looking at the average occupancy load for each block in the neighborhood, and figure 2a demonstrates the temporal heterogeneity by looking at the average occupancy over the entire neighborhood at each time of day and day of week that paid parking is available.

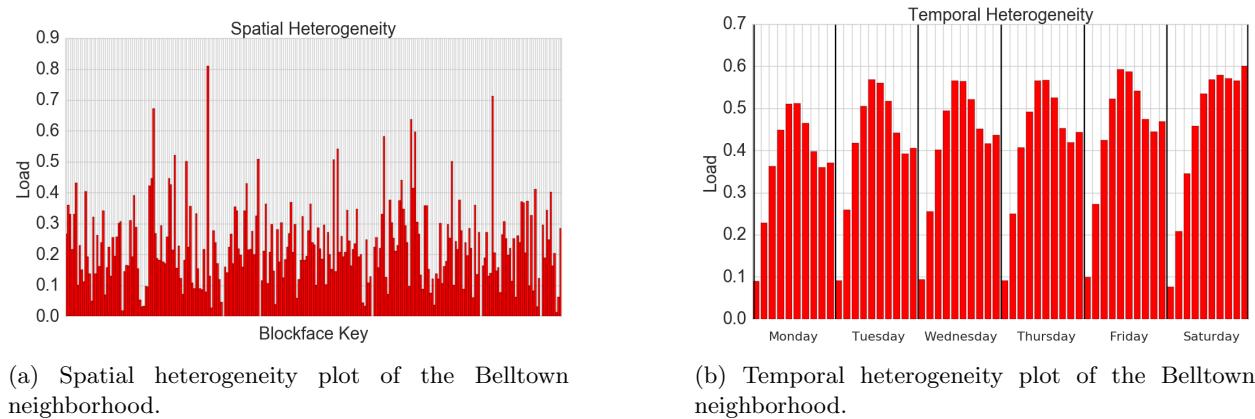
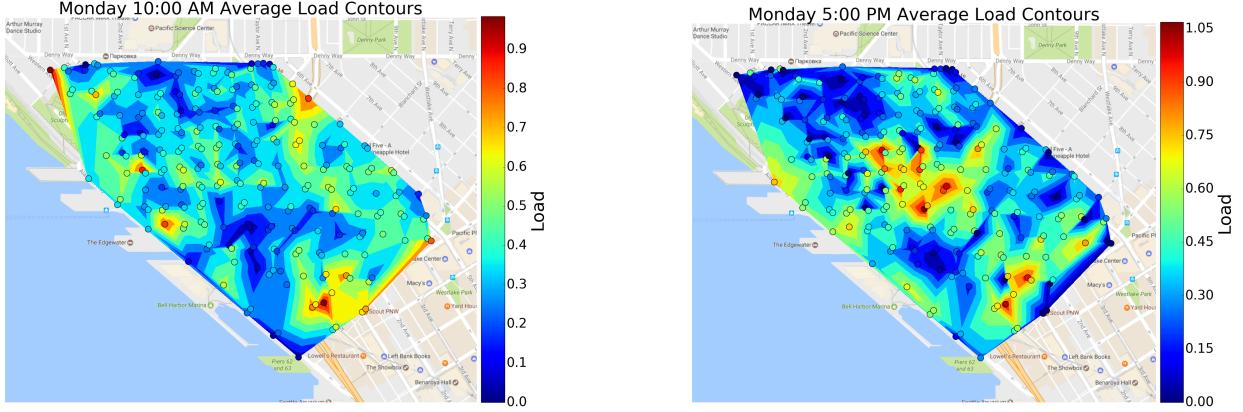


Figure 1

The spatial heterogeneity demonstrated in figure 1a, is one of the fundamental reasons that using average occupancy as the sole factor in policy decisions is an insufficient metric when considering how occupancy impacts congestion. The temporal heterogeneity demonstrated in figure 1b, shows that parking behavior is similar during weekdays, while behavior on Saturdays is very different. These properties of parking behavior are key when designing control strategies to mitigate congestion. We further observe that within the Belltown neighborhood there are centers of demand that dictate a majority of the congestion. The centers of demand are repeatable at given time of day and day of week over time, and we also find that the centers of demand change with time of day and day of week. We model locational demand by using a Gaussian Mixture Model (GMM) with spatial and occupancy data.

2 Gaussian Mixture Model for Locational Demand Analysis

As discussed previously, the occupancy of parking exhibits spatial heterogeneity within the Belltown neighborhood. By using a GMM we can identify centers of demand based on spatial information and occupancy data. By first considering only the occupancy data we can begin to get an idea for the distribution of parking occupancy and how parking behavior changes over the course of a day.



(a) 10 Contours of the average load in the Belltown neighborhood at 10:00 AM on Monday.

(b) 10 Contours of the average load in the Belltown neighborhood at 5:00 PM on Monday.

Figure 2

Although in figure 1a it can be observed that across Belltown there is a similar average load on Monday at 10:00 AM and 5:00 PM, the contour plot reveals that locational demand changes over a course of a day. In the morning parking is spread more uniformly across the neighborhood as people arrive to work. In the evening the parking becomes much more concentrated in the center of the neighborhood where many of the popular restaurants and bars are located.

Using a GMM we can model the centers of demand and give an estimate the influence the centers of demand have on neighboring blockfaces. A GMM can be used as an unsupervised clustering method to find partitions in a population containing similar samples. We consider a Multivariate GMM in which each sample \bar{x}_i is a vector containing spatial and occupancy data as $\bar{x}_i = [x_{\text{latitude}}, x_{\text{longitude}}, x_{\text{load}}]$. The purpose is to find which of the K possible mixture components each sample belongs to with the highest probability. Each mixture component has a prior on the probability of a sample belonging to it, $P(C_j) = \pi_j$. The likelihood of a sample belonging to a mixture component is then given as a Multivariate Gaussian Distribution

$$P(\bar{x}_i | C_j) = \frac{1}{(2\pi)^{\frac{3}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2} (\bar{x}_i - \bar{\mu}_j)^T \Sigma_j^{-1} (\bar{x}_i - \bar{\mu}_j)}. \quad (1)$$

We assume that the features (x_{latitude} , $x_{\text{longitude}}$, x_{load}) are conditionally independent given the component, i.e. each covariance matrix Σ_j is diagonal. The GMM is a classic chicken and the egg problem. We want to find the component that each sample belongs to, but to do so we need to know the distributions of each of the components. This problem is solved by the expectation-maximization (EM) algorithm which maximizes the log likelihood of the data. In the EM algorithm we initialize each of the component parameters, i.e. the prior, mean, and covariance, randomly. Until convergence is reached, which occurs when the complete log likelihood given by $L = \sum_{i=1}^n \log \sum_{j=1}^k P(C_j) P(\bar{x}_i | C_j)$, increases by an amount smaller than some ϵ , two steps are repeated at each iteration of the algorithm. These are the E step, in which the expected values of the unobserved component labels are calculated with the current parameter values, and the M step, in which new parameter values are calculated to maximize the likelihood of the data. The E step for all samples, x_i , is:

$$P(C_j | \bar{x}_i) = \frac{P(C_j) P(\bar{x}_i | C_j)}{\sum_{i'} P(C_{i'}) P(\bar{x}_i | C_{i'})}, \quad (2)$$

and the M step for all components, C_j , is:

$$P(C_j) = \frac{1}{n} \sum_{i=1}^n P(C_j | \bar{x}_i) \quad (3)$$

$$\bar{\mu}_i = \frac{\sum_{i=1}^n \bar{x}_i P(C_j | \bar{x}_i)}{\sum_{i=1}^n P(C_j | \bar{x}_i)} \quad (4)$$

$$\Sigma_j = \frac{\sum_{i=1}^n (\bar{x}_i - \bar{\mu}_j)^2 P(C_j | \bar{x}_i)}{\sum_{i=1}^n P(C_j | \bar{x}_i)}. \quad (5)$$

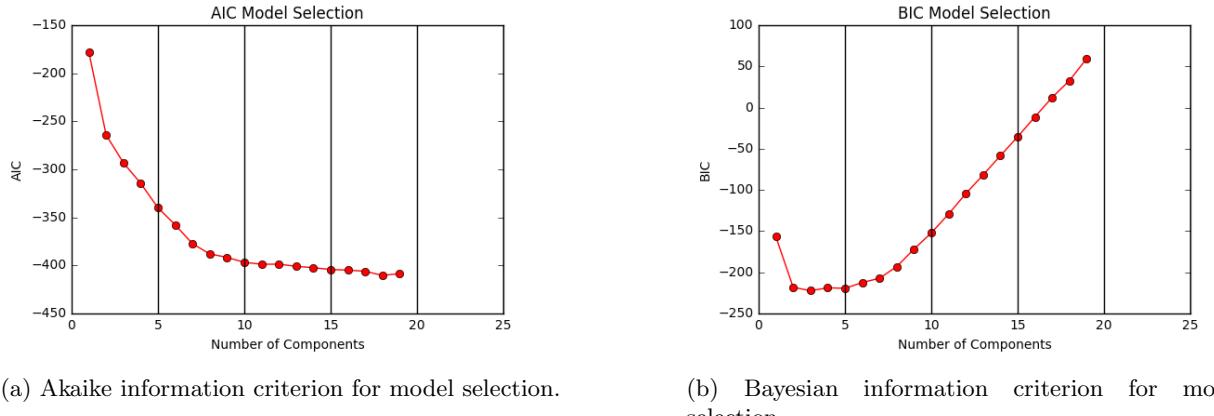
At convergence each sample x_i is assigned the component label which maximizes the posterior probability $P(C_j | \bar{x}_i)$, and the component parameters are recorded.

The GMM also involves a model selection problem in selecting the appropriate number of mixture components. We used both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) when choosing the model, whose equations are given as follows.

$$AIC = -2\ln(L) + 2k \quad (6)$$

$$BIC = -2\ln(L) + \ln(n)k \quad (7)$$

In each of these models, L is the maximized log likelihood of the GMM, n is the number of data points used in the GMM, and k is the degrees of freedom in the GMM. In a GMM containing C components with three dimensional data, there are seven degrees of freedom for each component. Three degrees of freedom come both the diagonal covariance matrix and the mean vector, and the final degree of freedom comes from the mixing weight. In total the number of parameters for the complete GMM is then $k = 7 \cdot C$.



(a) Akaike information criterion for model selection.

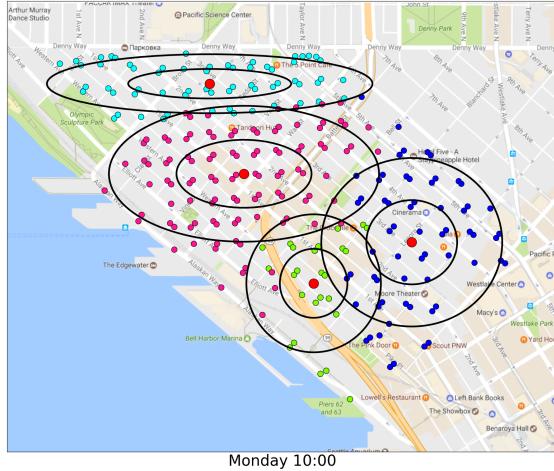
(b) Bayesian information criterion for model selection.

Figure 3

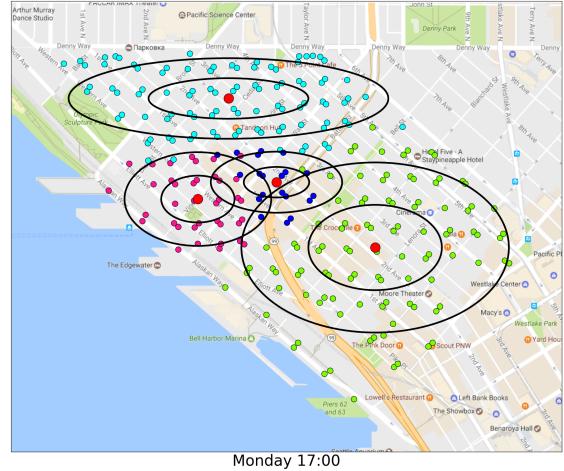
The AIC is known to often select a more complex model than the BIC. For this reason we considered both criterion and chose to use four components in our model.

We use the GMM to demonstrate three things: the spatial characteristics, including the locational demand, of parking depend on the time of day and day of week, parking behavior is repeatable at a given time of day and day of week, and to show there is spatial correlation in the network.

We begin by demonstrating the GMM applied to the Belltown neighborhood in the morning and evening on a Monday using the average loads from [Specify Dates](#). From the plots we can observe that both the mean and the shape of the distributions changes significantly. Further, we can observe that some spatial correlation is present since generally samples near each other are included within the same mixture component. We will quantify this correlation later on. These observations make it clear that parking behavior is dependent on the time of day. In this case, this reasons for this are quite clear, in the morning drivers are arriving to work, while in the evening drivers have different motives causing components to be more concentrated in the locations of popular restaurants and businesses.



(a) GMM applied to the Belltown neighborhood at 10:00 AM on Monday.



(b) GMM applied to the Belltown neighborhood at 5:00 PM on Monday.

Figure 4: GMM mixture applied to the Belltown neighborhood. Red circles indicate the mean of the component with respect to the latitude and longitude, and the black circles are the first and second standard deviations of the mixture component with respect to the latitude and longitude.

Although we have shown that on average locational demand and parking behavior depends on time of day, as well as day of week, we find that parking behavior at a given time of day and day of week is predictable from week to week. To demonstrate this we used occupancy information from [Specify Dates](#) and split data by the hour of the day and day of the week. For each day of the week and hour of the day, we fit the GMM on the occupancies of each block and the GPS coordinates, using the occupancy from a specific date. We then selected a new date at the same day of week and hour of day and made predictions for each blockface which component it belonged to. We assess the error of the prediction to be the number of predictions in which the blockface component prediction does not match that of the model we fit divided by the total number of blockfaces. This is the following:

$$\text{error} = \frac{\# \text{ Component labels changed}}{\# \text{ blockfaces}}. \quad (8)$$

We proceeded to make predictions in this manner on all dates in the date range with the same day of week and hour of day as in the model we fit. We average these errors, to give the average error for the model we fit. We then select a new date from the set with the same day of week and hour of day and repeat the procedure. Averaging over these errors gives us the average prediction error for the day of week and time of day. Finally, we do this for all days of week and hours of day with paid parking.

Table 1: Average prediction error of the GMM over all paid days of the week and hours of the day.

	8:00 AM	9:00 AM	10:00 AM	11:00 AM	12:00 PM	1:00 PM	2:00 PM	3:00 PM	4:00 PM	5:00 PM	Average Daily
Monday	39.99%	20.13%	10.74%	7.01%	6.13%	7.34%	11.21%	13.8%	14.86%	15.82%	14.70%
Tuesday	41.11%	20.96%	9.76%	7.21%	6.41%	5.23%	7.80%	11.28%	11.56%	15.33%	13.66%
Wednesday	37.32%	20.77%	9.45%	7.70%	5.63%	4.47%	4.33%	9.75%	10.19%	10.39%	12.00%
Thursday	37.29%	25.35%	10.12%	7.28%	6.96%	6.12%	7.32%	12.31%	11.99%	12.88%	13.76%
Friday	34.09%	22.17%	10.97%	6.34%	5.78%	7.07%	5.80%	9.74%	12.81%	13.28%	12.81%
Saturday	42.35%	32.87%	17.06%	12.26%	12.01%	10.64%	12.14%	11.53%	9.79%	9.58%	17.02%
Average Hourly	38.69%	23.71%	11.35%	7.97%	7.15%	6.81%	8.10%	11.40%	11.87%	12.88%	

The results in Table 1 illustrate that the locational characteristics of parking are similar at a day of week and hour of day from week to week. With the exception of 8:00 AM and 9:00 AM where the errors are much higher since occupancy is very low during these times causing anomalous behaviors, the average error at an hour of a day is very respectable with the range between 7.15 and 12.88 percent.

TODO: Look at how the location of the mean changes at each day of week and hour of the day - almost done

TODO: Show spatial correlation using standard spatial correlation measures like Morans I