Location-Centric Analysis For Advanced Indoor Mobility Models: Nodes Temporal Density Predictions and Distribution

Mimonah Al Qathrady, Ahmed Helmy
Computer and Information Science and Engineering Department
Gainesville, Florida, USA
{mimonah,helmy}@ufl.edu

ABSTRACT
Building’s density, as its number of nodes at a specific period, is a significant parameter that affects smart and mobile applications performances and evaluations. Consequently, the buildings’ temporal density predictions and their nodes spatial distribution modeling have to follow real-world scenarios to a realistic evaluation. However, there is lack of real-buildings density studies that examine these aspects thoroughly. As a result, this work presents a data-driven study that investigates the temporal density predictability and spatial density distributions of more than 100 real buildings with ten different categories, over 150 days across three semesters. The study covers the buildings’ temporal modeling and predictions, and their spatial distributions in the building. Seasonal predictive models are utilized to predict hour-by-hour density for a variable length of consequent periods using training data with different lengths. The models include Seasonal Naive, Holt-Winter additive seasonal, TBATS, and ARIMA-seasonal. The results show that the Seasonal Naive model is usually selected as the best predictive model when using data of the shorter period of time to predict the following periods. For example, Seasonal Naive predicted with the least error in 73%, 63% and 57% of cases in Summer, Spring and Fall respectively when using only one week to predict its consecutive five weeks. However, when using five weeks of data to predict the sixth week, TBATS model predicted with the least error in 60%, 54% and 43% of cases in Fall, Spring and Summer respectively.

1 INTRODUCTION
Building density is defined as the number of nodes in the area during a specific period. It is an individual behavioral pattern of a location. The density is an important parameter in the indoor modeling scenarios since it affects many wireless network functions and characteristics such as capacity, connectivity, and routing algorithm. Moreover, many indoor operations depend on update information about the population density, for example, system management of pedestrian flow inside crowd buildings [16]. Therefore, understanding the different aspects of indoor density, and reproducing them is requisite for realistic modeling and smart service designing and evaluation.

The density has always been set as a static parameter that does not change during the simulation. However, in real life, buildings densities are dynamic and change from one period to another as nodes keep coming to the buildings and then leaving from one period to another. It is necessary to measure the population number change over time as it is reflected in the real traces. The dynamic change of density is essential to be considered when modeling the user mobility. This is the first time, to the best of our knowledge, where the building level multi-temporal dynamic spatial density of population has been investigated. The study processes the trace data in days and hours respectively. From time-series density data, we observed clear peak hours during the day for all of the buildings. Then, the spectrum analyses were applied to investigate the frequency domain of the population density. Discovering the dominant density cycle facilitates reconstructing of the building density in indoor mobility models. The analysis of this paper involves more than 98 million mobile records from more than 100 buildings during three different semesters: Spring, Summer, and Fall. It covers several buildings categories including academic, museums, libraries, labs, administrations offices, sports facilities, dining, theaters, housing and health-care facilities. Moreover, population density prediction is an important mechanism when designing future indoor services or evaluating them. For example, predicting the crowd density in a critical situation such as emergency evacuation helps to implement the right and efficient evacuation plan for the building. Consequently, we have examined four seasonal models to predict hour density in the building: Seasonal Naive, Holt-Winter Seasonal Additive, TBATS, and ARIMA-seasonal. More than 5 million hour prediction operations are performed to establish the validity of these models and their ability to predict future population density at the building level. After that, building level density distribution is studied across different buildings. Previously, the nodes are usually distributed uniformly in the mobility models. However, this does not represent the reality of nodes distributions in real buildings. Density distributions for outdoor environments have been analyzed before, and it has been found to follow a power law distributions [9, 15]. In this work, we investigate the density at the building level where the number of nodes among different access points is examined.

In summary, the contributions of this work are as follow:

- Multi-temporal dynamic densities of more than hundred buildings during different semesters are studied for the first time. The Fourier Transform analyses are used to analyze the density frequency domains and discover the cycle. The study shows 24-hour as the dominant cycle for most of the buildings.
- The prediction of density per hour is studied extensively for all possible combination of modeling and testing windows. Several seasonal modeling is investigated: Seasonal Naive, Additive Holt-Winter, TBATS, and ARIMA. The result shows the seasonal naive prediction error is the least in most cases,
Various researchers have taken Data-driven mobility models approach. Therefore, real data traces have been collected and analyzed to build the models. Traces analysis are categorized as user-centric or location-centric [5].

2.1 User-Centric Vs. Location Centric
User-centric approaches to study the mobility behavior of the user, by tracking a user history, and their statistical patterns. Most of the mobility studies are user-centric where they focused on the mobile user individual, pairwise or collective behavior [17]. Individual behaviors such as temporal and spatial preferences, a user average speed and duration have been analyzed [11]. Pairwise patterns include encounter studies and modeling [10], while collective behaviors detect communities and model them [17].

Location-centric analysis approaches focus on the location and summarize its statistical characteristics of its visitors regardless of the identity or mobility histories of their users. Examples are studying the pair-wise characteristics of location based on the flow of users between locations [5]. Another study is the location density and density distributions as an individual location pattern characteristics.

2.2 Density Analysis and Modeling
Since they are significant factors in mobility models, several previous studies analyzed the spatial density distributions over outside area from different traces environment, and a mobility model that preserve density distributions is proposed [9, 15]. Most of the previous studies were not designed to be used for modeling at the building level. They usually cover a wide range of area and track pedestrian movement outdoors. On the other side, this paper has focused on the spatial distributions of users at the building level.

Also, the study tracks the density change over time at the multi-granularity temporal level. The change of density from one period to another is necessary to be captured and modeled in realistic indoor mobility.

Moreover, this study considers the building categories, compare and contrast between buildings based on their category.

3 DATA-DRIVEN BUILDING LEVEL DENSITY ANALYSIS AND MODELING FRAMEWORK

The analysis framework is presented in figure 1. The process starts with data processing and building categorization. Then, the density analysis consists of two parts: temporal and spatial. The multi-temporal analysis tracks the hour by hour and day by day density, while spatial analysis focuses on the distributions of the density in the buildings.

3.1 Data
This section consists of two parts: description of the wireless data that is used in this study, and the data processing.

3.1.1 Wireless data. They are collected from more than 100 buildings on the university campus. The mobile records include when a user starts and ends an association with an access point. The buildings belong to ten different categories. The categories include museums, academic, libraries, labs, administration, sports facilities, dining, and housing. Each access point is tagged with the buildings’ code and the room number. Information about the buildings categories data such as the number of buildings, users, and records are presented in table 2. The traces are for 150 days for three semesters: Spring, Semester and Fall.
3.1.2 Records Filters. We filtered out the mobile records for two purposes: reducing the effect of ping pong, and filtering based on the device type.

Filtering phone devices. Our traces contain the mobile data of both phones and laptops. Most of the people always have their Smartphone at their side day and night [1] and they are more likely to use it while they are moving than laptops users. As a result, we concentrate on phone mobile data and filter out the laptop data. The device mac address in the traces and their website visitation are used to identify if the device is a laptop or phone. More detail about device type filtering can be found in [6, 14].

Filters to reduce the ping-pong effect. The ping-pong effect occurs when the wireless user is at the edge of the two access points and hope between them. Here the user does not change its location. Indeed, the user is being directed back and forth between the access points. Then, the user appears to be repeatedly associated with a fixed number of access points. Due to incomplete information and the ambiguity in its interpretation [18], there is no perfect solution to this problem. To reduce the ping-pong effect, we filtered out the records that have association back and forth between two access points in less than Theta. In our filters, we assign Theta = 10. Increasing Theta values risks deleting records that do not result from ping-pong effects. This process filtered out less than 0.01 of the records.

Filtering very short session duration. Since the analysis done at the building level, and the people that are suspected to be passing outdoor are to be discarded. As a result, records with session duration that are less than Theta are to be ignored. In this part, Theta are selected to be one second. We are cautious here to not throw records that happens inside the building.

3.1.3 Assigning Buildings Category. Our data includes several buildings categories. The Access points are linked to their buildings using online information about buildings code. Then, the buildings have been categorized. The categories are academic, administration, labs, dining, housing, sports facilities, museum, libraries, theaters and auditorium and health care. To identify which buildings belong to what category, we use the following means:

(1) Campus Map: We used online university campus map [4] to identify several buildings categories. The map offer answers to queries such as the dining, libraries or housing buildings.
(2) Academic and classrooms buildings: pictures and information about classrooms on the campus can be found in [2].
(3) Specific Building Information: there are web pages that introduce a building history and what the building have been used for, especially the named facility, as in [3]. We were able to categorize more than 102 buildings into a specific category. The categories include Academic, Administration, Sports, Labs, Libraries, Museums, Dining, Theaters and Health Care. Some buildings have several functionalities such as a building consist of classrooms on one floor and a museum on the second floor. Hence, we do not consider them to belong to a specific category.

3.2 Density Analysis and Modeling
The density analysis covers two parts: one is analyzing the temporal change, and the other is to study the density distribution inside the building.

3.2.1 Temporal Density Analysis and prediction. We targeted two temporal density analysis: day by day and hour by hour. In each building, the number of population per hour is computed. The analysis focuses on two parts: the spectrum analysis of density data, and correlate the density with encounter events.

Spectrum analysis of density. The frequency of the density data are analyzed to answer the following questions:

(1) Is the population density has a dominant cycle that keeps repeating again and again?
(2) If there is a dominant cycle, what is it?

The frequency magnitude in the spectrum. The Fast Fourier Transform is used to calculate the periodogram. If y_{h} is the autocovariance function for the data at time lag h, then at frequency ω, the spectral density f(ω) can be calculated as:

\[ f(\omega) = \sum_{h=-\infty}^{\infty} y(h) e^{-2\pi i oh} \]

We concerned about the frequency between 0 and 1/2. If the frequency f corresponds to the maximum spectrum, then our dominant cycle will be 1/f. The periodogram graph shows the dominant cycle.

3.3 Temporal Density Prediction
This section describes the training and testing period windows, prediction models and assessment.

3.3.1 Prediction Testing and Training Windows. We have tried all possible combinations of training modeling weeks and testing predictions. For example, One week of training model, and the next one week of predictions. For example, the next figure is the first-week training models; it will be tested for its ability to predict the next week, 2weeks,...5 weeks.

After that, the training data become two weeks, and its ability to predict the next period is tested for the next week, the next 2weeks. The number of prediction operations for each building for each semester will be: We will have 15 tries when we use one week for modeling, and 10 when using two weeks. In total, we will have 35 tries for each building for each semester. We have 103 buildings in fall and summer and 100 buildings in Spring. In fact, there are 32130 predictions operations for each model that have been completed. Each operation predicts hour by hour for a period range from one week to five weeks. Since we are comparing four forecasting prediction models, 128,520 predictions operations are evaluated. They contain density predictions for 539,784,000 hours (4200 hours for each model for each building for each semester).

3.3.2 Prediction Models. Four seasonal predictive models are used to model hour by hour density and predict the future density: Seasonal Naive, Holt-Winters (additive seasonal), auto ARIMA and TBATS. The following paragraphs will describe them briefly. More detail can be found in [12].

Seasonal Naive: the predicted value is equal to the last value from
the same season.

\[ y_{T+h|T} = y_{T+h-km} \]  

where \( m \) is the seasonal period and \( k = \left\lfloor \frac{h}{m} \right\rfloor + 1 \). In case of density study, a power spectrum analysis to hour by hour density reveals that one day is the dominant cycle in most of the buildings. As a result, one day is considered as a season in the model.

**Holt-Winter Seasonal Model**: It comprises the forecast equation and three smoothing equations: one for the level \( \ell_t \), one for trend \( b_t \), and one for the seasonal component denoted by \( s_t \). For smoothing parameters \( \alpha, \beta \) and \( \gamma \), \( m \) is used to denote the period of the seasonality. In this study, we used the additive method since the seasonal variation is roughly constant. The seasonal component is expressed in absolute terms in the scale of the observed series, and in the level equation, the series is seasonally adjusted by subtracting the seasonal component. The additive method components can be expressed as:

\[ \begin{align*}
\hat{y}_{t+h|t} &= \ell_t + h b_t + s_{t-m} + h^* m \\
\ell_t &= \alpha (y_t - s_{t-m}) + (1 - \alpha) (\ell_{t-1} + b_{t-1}) \\
b_t &= \beta (\ell_t - \ell_{t-1}) + (1 - \beta) b_{t-1} \\
s_t &= \gamma (y_t - \ell_t - b_{t-1}) + (1 - \gamma) s_{t-m}.
\end{align*} \]  

where \( h^* m = \left\lfloor (h - 1) \mod m \right\rfloor + 1 \). The level equation shows a weighted average between the current seasonally adjusted observation \( y_t - s_{t-m} \) and the non-seasonal forecast \( \ell_{t-1} + b_{t-1} \) for time \( t \). The seasonal equation shows a weighted average between the current seasonal index, \( y_t - \ell_t - b_{t-1} \), and the seasonal index of the same season (i.e., \( m \) time periods ago). More detail about this model can be found in [12].

**TBATS Model**: It is Trigonometric regressor with Box Cox Transformations, ARMA errors, Trend and Seasonality. It is used to model series exhibiting multiple complex seasonality [8]. It uses a combination of Fourier terms with an exponential smoothing state space model and a Box-Cox transformation, in a completely automated manner. As with any automated modeling framework, there may be cases where it gives poor results, but it can be a useful approach in some circumstances. A TBATS model differs from the dynamic harmonic regression in that the seasonality is allowed to change slowly over time in a TBATS model, while harmonic regression terms force the seasonal patterns to repeat periodically without changing. One drawback of TBATS models, however, is that they can be slow to estimate, especially with long time series

**ARIMA Model**: It stands for Autoregressive Integrated Moving Average. It aims to describe the autocorrelations in data. A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA models. It is written as follows: ARIMA \((p, d, q)(P, D, Q)_m\), where \( m \) = number of periods per season. The uppercase notation is used for the seasonal parts of the model, and lowercase notation for the non-seasonal parts of the model. The seasonal part of the model consists of terms that are very similar to the non-seasonal components of the model, but they involve backshift of the seasonal period. For example, \( d \) is the order of first differencing, and \( D \) is the order of seasonal differencing. Autoregressive AR\((p)\) implies current values depend on its \( p \)-previous values. Moving average MA\((q)\) means the current deviation from the mean depends on \( q \)-previous deviations, where \( q \) is the order of MA process.

In this part, we used auto.arima model [13], which return the best ARIMA model according to information criterion (AIC), but it is not necessary to be the best in terms of prediction error. The order of differencing \( d \) is based on the KPSS test, while the order of seasonal differencing is based on OCSB test.

### 3.3.3 Prediction Assessments

The mean absolute error have been used widely for prediction assessment. MAE can be computed as:

\[ MAE = T^{-1} \sum_{t=1}^{T} |\hat{y}_t - y_t|, \]  

where \( y_t \) denote the \( t_{th} \) hour density and \( \hat{y}_t \) denote its prediction values, where \( y_t \geq 0 \). However, MAE is scale dependent. Therefore, the mean error is not a useful metric to compare between different buildings with different density. To make MAE scale independent, the MAE is then divided by the mean density per hour in the testing period.

### 3.4 The density and encounter events

The encounter is critical events since they give the opportunity for sedimentation of messages or even infections. To demonstrate the importance of considering density, we computed the number of encounter events that occurred at each hour. Then, the relation between encounter and density are computed for each daily and hourly time series data.

**Figure 2**: Spectrum Analysis of day by day data in CISE

**Figure 3**: Spectrum Analysis of hour by hour data in CISE

### 3.5 The density distributions

To investigate the density distributions, we investigate the number of users that have associated with an access point through a day. Several distributions are used to fit the data. Then, eleven distributions
are used to fit the density using the maximum likelihood methods. The distributions are: Power-law (Pl), Weibull(W), gamma (G), log-normal (Ln), Pareto(Pr), Normal(N), Exponential(Ex), Uniform(U), Cauchy(C), Beta(B) and log-logistic(LG). The Kolmogorov-Smirnov statistic is used to evaluate the fitted distributions. The three best fit is selected. Also, the percentage of distributions that have KS-stat less than 0.075 or 0.10 are reported.

4 RESULT AND DISCUSSION
This section shows the spectrum analysis result, hour-by-hour predictions, density distributions and temporal density and encounter correlations.

4.1 Spectrum analysis of temporal density
To discover the cycle in the density, we analyze its frequency domain as prescribed in section 3. This work used [7] to analyze the spectrum of buildings density. The daily cycle is shown to be one week in most of the building, with exception to a few buildings. However, all kinds of buildings have a 24-hour cycle. Figure 2 shows the daily spectrum of the building CSE and figure 3 shows the hourly spectrum of the same building. Most of the other buildings have similar spectrum analysis result, where 24 hour is the dominant cycle in hour by hour data and seven days for daily density data.

4.2 Prediction Models
This section discusses the density predictions from several dimensions: The prediction models, the predicted period length (testing data size), the training model length. Figure 4 shows the mean Normalized MAE for predicting density per hour for one, two, three and four weeks using different lengths of training models during Fall Semester, using different density algorithms, Spring and Summer show similar trends too. Interestingly, the mean error across all buildings shows the seasonal naive is the best to model and predict the temporal hour by hour data, since it predicts with less error percentage less than 0.3 on average. Also, the training period length does not affect the seasonal naive result as much as other models. TBATS, Seasonal ARIMA, and Holt-Winter Additive Seasonal improved when the training data period increases. For instance, TBATS specifically improved when using five weeks to predict the next week and slightly outperforms the Seasonal Naive. Figure 5 shows the result of predicting the next five weeks using one week of data in the training models. Also, seasonal naive outperformed other models when predicting the hour by hour density for a week.

The result shows it represents the most cases. However, there are exceptional cases when seasonal naive is not the best, especially when having complex seasonality where TBATS outperform Seasonal Naive. The result in this section summarizes the case in most of the buildings. In future, we will examine the exceptional cases and categorize the buildings based on their best prediction models.

Note that, to remove outliers, we exclude the errors that are more than 1, which turn out to be rare cases.

4.3 Density Distributions
To investigate the density distributions, we consider buildings that have more than 20 Access point during a day in summer where buildings are occupied the most. The distributions of users among the access points in a day are investigated. The buildings have a dense population in that day, more than most of the days in the trace. Several distributions are used to fit the density using the maximum likelihood methods. The distributions are: Power-law (Pl), Weibull(W), gamma (G), lognormal (Ln), Pareto(Pr), Normal(N), Exponential(Ex), Uniform(U), Cauchy(C), Beta(B) and log-logistic(LG). The Kolmogorov-Smirnov statistic is used to evaluate the fitted distributions. The three best fit are selected and presented in table 4. Interestingly, the power-law is only selected as the best fit from 29% of the buildings in spring to 35% of buildings in summer, other distributions such as log-logistic have been selected as the best fit distribution for several buildings. Besides, the percentage of buildings that have KS-stat less than 0.10 in their fitted log-logistic distributions are 53% in Spring, while 29% only are power-law. This result is significant since the previous investigation at the campus level or park level have found that the power law density distributions are the best fit distribution for the population density. However, the data at the building level do not follow the same result. Example of the buildings density distribution is Florida Gym. It consists of 48 Access points and more than 2.6K reports. The three best-fit distributions do not include power-law distribution.

4.4 Correlation Between Temporal Density and Encounters
The encounter per hour and density per hour data are measured. As expected, in most cases, the encounter rate is following the density rate as their correlations are strongly positive. Figure 7 and 8 show the result of density and encounter correlation in different buildings categories.

5 CONCLUSIONS AND FUTURE WORK
We have analyzed a real trace data at the building level toward seeking a better indoor mobility model and smart applications. We have selected buildings from various categories and have focused on density, as a critical parameter for indoor modeling and investigated hour density prediction and spatial density distribution in the buildings. The density is covered as an individual location pattern. The paper findings are beneficial for IoT applications that require information about density, relies on the crowd and affected by the crowd number. This study is built with the purpose of creating an advanced indoor mobility model that represents the most important observations, and enable accurate simulation of smart indoor applications by recreating real scenarios of population density. The mobility model will combine the statistical data from real traces with other contextual information such as buildings layout, constraints and vertical movement between floors.

ACKNOWLEDGMENTS
This work was partially funded by Najran University, Saudi Arabia, and NSF 1320694.
Table 2: Percentage of the Best Prediction Model Per Semester, S: Seasonal Naive, T: TBATS, Ar: Arima, H: Holt-Winter

<table>
<thead>
<tr>
<th>Training Data Length</th>
<th>Semester</th>
<th>One</th>
<th>Two</th>
<th>Three</th>
<th>Four</th>
<th>Five</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>Spring</td>
<td>S[52%], T[36%]</td>
<td>S[57%], T[32%]</td>
<td>S[53%], T[43%]</td>
<td>S[58%], T[40%]</td>
<td>S[57%], T[38%]</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>S[47%], T[32%]</td>
<td>S[58%], T[32%]</td>
<td>S[66%], T[27%]</td>
<td>S[68%], T[27%]</td>
<td>S[73%], T[20%]</td>
</tr>
<tr>
<td></td>
<td>Fall</td>
<td>S[56%], T[37%]</td>
<td>S[59%], T[36%]</td>
<td>S[59%], T[37%]</td>
<td>S[56%], H[41%]</td>
<td>S[63%], H[35%]</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>T[34%], S[28%]</td>
<td>S[33%], T[33%]</td>
<td>S[33%], T[33%]</td>
<td>T[40%], H[37%]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Fall</td>
<td>T[40%], A[27%]</td>
<td>T[44%], S[28%]</td>
<td>T[47%], S[31%]</td>
<td>T[52%], S[33%]</td>
<td>-</td>
</tr>
<tr>
<td>Three</td>
<td>Spring</td>
<td>T[58%], A[17%]</td>
<td>T[58%], S[18%]</td>
<td>T[66%], S[12%]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>T[39%], S[24%]</td>
<td>T[37%], S[35%]</td>
<td>T[40%], S[33%]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Fall</td>
<td>T[50%], S[26%]</td>
<td>T[56%], S[24%]</td>
<td>T[58%], S[23%]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Four</td>
<td>Spring</td>
<td>T[66%], S[15%]</td>
<td>T[72%], S[16%]</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>T[40%], S[25%]</td>
<td>T[47%], S[33%]</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Fall</td>
<td>T[61%], S[18%]</td>
<td>T[57%], S[25%]</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Five</td>
<td>Spring</td>
<td>T[54%], H[22%]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>T[43%], S[23%]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Fall</td>
<td>T[60%], S[15%]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Semester</th>
<th>1st best Fit</th>
<th>2nd best Fit</th>
<th>3rd best Fit</th>
<th>≤7.5% KS statistic</th>
<th>≤10% KS statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer</td>
<td>Pl[35%], Li[29%], Ln[18%]</td>
<td>Li[24%], C[18%], Pl[18%], W[12%], C[18%], Ln[12%]</td>
<td>W[29%], Li[29%], G[12%], Ln[12%]</td>
<td>Li[18%], Pl[12%], Ln[12%]</td>
<td>Li[51%], Pl[29%], Ln[29%], G[29%], W[24%], C[11%]</td>
</tr>
<tr>
<td>Spring</td>
<td>Pl[29%], Li[18%], Ln[18%]</td>
<td>G[12%]</td>
<td>Li[41%], Li[31%], Ln[18%]</td>
<td>Li[24%], Li[24%], Pl[24%], W[12%], G[12%]</td>
<td>Li[11%], Li[33%], Ln[33%], G[29%], W[21%], C[11%]</td>
</tr>
</tbody>
</table>

REFERENCES
Figure 5: Normalized MAE of predictions for hour by hour density for five weeks using one week. For Fall, Spring and Summer.

Table 4: Parameters and KS-statistics for Florida Gym best fit density distributions.

<table>
<thead>
<tr>
<th>Distributions</th>
<th>Parameters</th>
<th>KS-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st fit</td>
<td>Log-logistic</td>
<td>Shape=2.382299, scale=19.298108</td>
</tr>
<tr>
<td>2nd fit</td>
<td>Log-normal</td>
<td>meanLog=2.9636519, sdLog=0.7599539</td>
</tr>
<tr>
<td>3rd fit</td>
<td>Gamma</td>
<td>Scale=14.215362, Shape=1.833448</td>
</tr>
</tbody>
</table>

Figure 6: Florida Gym three best population density fit distributions: 1st: Log Logistic, 2nd: Log Normal, 3rd is Gamma. Models parameters and KS-stat are listed in Table 4.

Figure 7: Correlation Between Daily Density and Daily Encounter Events

Figure 8: Correlation Between Hourly Density and Hourly Encounter Events