Time Series Analysis for Device Usage in Electronic Classrooms

Joshua Williams[1], Yong Xu[2]
[1]Radford University, Box 7004, 801 E. Main Street, Radford, VA 24142
[2]Radford University, Box 6942, 801 E. Main Street, Radford, VA 24142

Abstract
With technology and multimedia devices becoming ever more present in classroom environments from Elementary Schools to Universities, understanding which devices and how often they are used is needed. The object of this current study is to develop a statistical model that tracks which and how often devices will be used in a classroom environment. We are using data collected from sixty-eight multimedia classrooms at Radford University from Crestron RoomView to develop time series models. Having the statistical models of classroom technology/multimedia will aid IT professionals, administrators, and teachers by allowing for more efficient classroom design and device selection.

Key Words: Time Series, Resource Management, Device Usage, Electronic Classroom ARIMA, Statistical Model

1. Background

The public university used in this study is Radford University located in Radford, Virginia. The university was founded in 1910 as an all-women's college which became coeducational in 1972. Radford University currently has 180 electronic classrooms, each using a Crestron control system, spreading across 18 academic buildings on the campus. Each classroom has source components consisting of a dedicated computer (PC), a document camera (DC), a laptop connection (LA), a Blu-Ray/DVD/VCR combination player (DVD), a cable TV tuner (CT) and a wireless projection system (WPS). In addition to the source components, each classroom has an output component consisting of at least one digital projector and can have a secondary output component consisting of either a second digital projector and/or one to three TV monitors. Central to these components is the Crestron MPS-300 and a Crestron DM-8X8 which controls and routes the A/V signals of the source components to the output component(s).

Beginning in October 2011 the University's IT department reprogramed the Crestron control systems to report data on how each classroom was used to the recording program, RoomView (version 7.2 also by Crestron). The data recorded by RoomView included in this study are the building and room number, the source component, and the date and time that a source component is used. By mid-January of 2012 sixty-eight of the Crestron controlled classrooms had been reprogrammed to send data to RoomView. The data collection continued through mid-September 2012 until RoomView was shut down to facilitate upgrading to a newer version. In late 2013, we decided to analyze the data which had been recorded to determine how the classrooms were being used and to alter any future classroom designs.

2. Literature Review

A large number of studies on In Classroom Technology (ICT) focus primarily on how equipment is used by instructors or the instructor's perceptions towards ICT. Usluel, Aşkar & Baş (2008) looked at faculty access to ICT through a survey of 814 Faculty from across 22 schools in Turkey. The survey asked what purpose the ICT was used and how long instructors had been using ICT, in year time frames. Usluel, Aşkar & Baş's survey found that 56% of the faculty listed the projector as the most common item available to them. Gülbahar (2008), also in Turkey, asked 25 faculty (24 of which responded) and 558 students (304 of which responded) about ICT usage in the classroom. This study focused primarily on how the ICT was being used, finding that only 58% of the faculty surveyed used computers for presentations. Brill and Galloway (2007) conducted a survey of faculty at a land grant university in the United States to investigate the usage of classroom based teaching technologies. They summarized the attitudes of college level instructors toward teaching technologies. Of the 53 surveys returned out of 180 they found overhead projectors and VCR's to be the two devices with the highest usage among faculty. These studies and others like them often ask what kinds of equipment are used and how often as the secondary question. To get large sets of accurate and detailed data on what equipment is used and how often, surveys such as these would be impractical. This is due to the time investment in designing the survey and the dependency on accurate responses in sufficient numbers to generate a model.

In studies looking at energy usage of equipment, questions of what kind equipment is used and how often are the primary focus. However the equipment observed in these surveys often include devices that are not found in a classroom environment. Webber, et al (2005) and Roberson, et al (2004) looked at the electrical use of office equipment after business hours during the weekdays and weekends at 12 buildings (including office buildings, schools and hospitals) for a governmental agency's energy estimation. These audits were to determine which devices were being left on, which were in standby mode and which were turned off. Webber, et al. (2005) explained the uncertainties that come from estimating energy usage, user's behavior and power management rates. These types of studies which conduct physical audits of equipment during and after business hours would prove impractical for most educational institutions that often have limited budgets.

With the available RoomView resources, we are able to conduct more accurate and efficient analysis to answer the questions that which ICT was used and how often these ICT are used. Using Autoregressive Integrated Moving Average (ARIMA) modeling for time series analysis and forecasting has been very popular with many researchers in the past decade. Bowerman, et al (2005) studied forecasting and time series in an applied way. Shenston and Hyndman (2005) used stochastic models to forecast intermittent demand. Shih and Tsokos (2008) applied ARIMA models to analyze global warming issues. Shih and Tsokos (2009) forecasted the future CO_2 amount in the atmosphere through ARIMA modeling. Hyndman and Khandkar (2008) developed a forecasting package for the R language. Abdullah (2012) applied ARIMA modeling for gold bullion coin selling prices. Vasileiadou and Vliegenthart (2013) studied dynamic social processes using an ARIMA model. In this study, we will use time series analysis to analyze the data which generated from RoomView system and construct ARIMA models for each ICT source: PC, DC, LA, CT, DVD and WPS.

3. Methodology

The data set for our study consists of 68 classrooms. The earliest classrooms began recording data on October 13th 2011 and the last recording was on September 14th 2012. The data was processed using SAS 9.2 (32bit version) to divide the time that the equipment was used into 2 distinct time frames. The first time frame (referred to as "Prime Time") covers 7:30 a.m. through 7:30 p.m. Monday through Friday. "Prime Time" is when the majority of instruction occurs at the university. The second time frame (referred to as "After Time") covers 7:30 p.m. through 7:30 a.m. Monday through Friday and all day Saturday and Sunday. "After Time" is when the students are using the classrooms for studying and/or recreation. After processing, the data set was then converted to a time series format and exported for further analysis by R. The ARIMA models and plots of the usage of the equipment were constructed in R version 3.02.

4. Statistical Analysis

4.1 Preliminary observations

An example plot of the data by source for a single classroom (Cook 317) gives us Figure 1, which shows the average daily usage in minutes for each device. A plot of the data for each device's average daily usage, also in minutes, for the whole of campus gives us Figure 2. Clearly seen in both of these figures is that the Document Camera and PC are the most used devices, with all the other devices' usage lagging behind by a significant amount.

Additional plots of the PC and DVD data are given by Figures 3 and 4, respectively. These figures, which plot the data in chronological order, also show the drastic difference in average minutes used between the PC and DVD during the time frame in which the data was collected. Additionally these figures closely mirror each other in the periodic rise and fall of the weekly usage and in the "gaps" of low usage. These "gaps" in both figures correspond to winter break (centered at 2012.0) and summer break (centered roughly at 2012.5). A closer look at the periods of heavy usage in both figures also reveals a pair of mini "gaps" which corresponds to Thanksgiving break (roughly at 2011.9) and spring break (roughly before 2012.2).

4.2 ARIMA Modeling of each Device

ARIMA is used for the modeling and forecasting of time-series data. Since our data set has a short time frame, we tested the data set and found that there are no significant seasonal effects. Therefore we chose to use ARIMA model without seasonal effects, given by Equation 1.

$$\left(1-\sum_{i=1}^{p}\Phi_i L^i\right)(1-L)^d Y_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t \qquad (1)$$

Where L is the lag operator $d \in Z$ and $d > 0$. Y_t is the response value, $\hat{Y}_t$ is the fitted value, and $\varepsilon_t = Y_t - \hat{Y}_t$ is the forecast residual for a given time period.

We will select the best possible ARIMA model with p and q with a range of $0 \le p \le 5$ and $0 \le q \le 5$.

For the PC source component, the model which best fit the data was found to be model (4, 1, 5), given by Equation 2. A plot of the data including a 30-day forecast using this model is given by Figure 5.

$$Y_t - 1.78Y_{t-1} + 2.17Y_{t-2} - 2.16Y_{t-3} + 1.7Y_{t-4} - .93Y_{t-5} = \varepsilon_t - 1.078\varepsilon_{t-1} + 1.44\varepsilon_{t-2} - 1.1\varepsilon_{t-3} + .84\varepsilon_{t-4} - 3.06\varepsilon_{t-5} \quad (2)$$

For the Document Camera source component, the model which best fit the data was found to be model (4, 1, 5), given by Equation 3. A plot of the data including a 30-day forecast using this model is given by Figure 6.

$$Y_t - 1.78Y_{t-1} + 2.17Y_{t-2} - 2.153Y_{t-3} + 1.693Y_{t-4} - .93Y_{t-5} = \varepsilon_t - 1.08\varepsilon_{t-1} + 1.44\varepsilon_{t-2} - 1.1\varepsilon_{t-3} + .84\varepsilon_{t-4} - .31\varepsilon_{t-5} \quad (3)$$

For the Laptop source component, the model which best fit the data was found to be model (2, 1, 1), given by Equation 4. A plot of the data including a 30-day forecast using this model is given by Figure 7.

$$Y_t - 1.64Y_{t-1} + 1.03Y_{t-2} - 0.38Y_{t-3} = \varepsilon_t - 0.85\varepsilon_{t-1} \quad (4)$$

For the DVD source component, the model which best fit the data was found to be model (3, 1, 4), given by Equation 5. A plot of the data including a 30-day forecast using this model is given by Figure 8.

$$Y_t - 1.34Y_{t-1} + 0.23Y_{t-2} + 0.96Y_{t-3} - 0.86Y_{t-4} = \varepsilon_t - 1.01\varepsilon_{t-1} - 0.08\varepsilon_{t-2} + 1.03\varepsilon_{t-3} - .7\varepsilon_{t-4} \quad (5)$$

For the Cable TV source component, the model which best fit the data was found to be model (0, 1, 1), given by Equation 6. A plot of the data including a 30-day forecast using this model is given by Figure 9.

$$Y_t - Y_{t-1} = \varepsilon_t - 0.89\varepsilon_{t-1} \quad (6)$$

For the WPS source component, the model which best fit the data was found to be model (3, 1, 3), given by Equation 7. A plot of the data including a 30-day forecast using this model is given by Figure 10.

$$Y_t - 1.77Y_{t-1} + 1.19Y_{t-2} + 0.022Y_{t-3} - 0.44Y_{t-4} = \varepsilon_t - 1.86\varepsilon_{t-1} + 1.72\varepsilon_{t-2} - 0.67\varepsilon_{t-3} \quad (7)$$

From these forecast plots (Figure 5 to Figure 10), we can tell that the future trends of the usage of the six sources are not significantly different from the current trends. We can tell the PC, Document Camera, and Laptop connection usage is much more than DVD/VCR player, Cable TV, and WPS devices on campus. Figure 11 combines the device usage of all six sources, scaled to a common X axis, for accurate comparisons. From Figure 11, it is clear that the PC, Document Camera are used most often, followed by Laptop. DVD, CT and WPS are the least used devices on campus.

4.3 Evaluation

The descriptive statistics for the total campus equipment usage of all six sources are given in Table 1. The unit of measurement is in minutes. We investigated the following statistics for each of the six sources: the sample mean (MEAN), the sample standard deviation (SD), the sample standard error (SE), the minimum (MIN) value and the maximum (MAX) value.

Table 1: Descriptive Statistics for the Total Campus Equipment Usage

Names	*MEAN*	*SD*	*SE*	*MIN*	*MAX*
PC	2739.53	3980.27	216.5	0	12724
DC	2839.22	3793.44	206.34	0	13644
LA	470.58	738.72	40.18	0	2531
DVD	123.94	213.5	11.6	0	1275
CT	32.93	93.2	5.07	0	807
WPS	24.43	55.85	3.04	0	244

We used the following criteria to evaluate the model: the Mean Error (ME), the Root Mean Square Error (RMSE), the Mean Square Error (MSE), the Mean Absolute Error (MAE), the Mean Absolute Percentage Error (MASE), and Pseudo R-square.

The Mean Absolute Error (MAE) is the average of absolute value of the residual and is given by Equation 8.

$$\text{MAE} = \frac{1}{\text{n}}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right| = \frac{1}{\text{n}}\sum_{t=1}^{n}\left|\varepsilon_t\right| \tag{8}$$

Where Y_t is the response value, Y_{t-1} is the fitted value, and $\varepsilon_t = Y_t - \hat{Y}_t$ is the forecast residual for a given time period.

The Mean Absolute Scaled Error (MASE) is a measure of accuracy of a method for forecasts and is given by Equation 9.

$$\text{MASE} = \frac{1}{\text{n}}\sum_{t=1}^{n}\left(\frac{\left|\varepsilon_t\right|}{\frac{1}{n-1}\sum_{i=2}^{n}\left|Y_i - Y_{i-1}\right|}\right), \tag{9}$$

where Y_i is the response value at time i and Y_{i-1} is the response value at time $i - 1$.

Pseudo R-square is defined as the square of the correlation between a models predicted values to the actual values. The evaluation results for all six sources are given in Table 2. The largest Mean Error is 22.6 minutes for the total campus PC usage. This result is less than one percent of the mean campus PC usage for any given day. The Pseudo R-square for the campus PC usage was found to be .73. This indicates that the magnitude of the correlation between the predicted values and the actual values is strong for campus PC usage. The Mean Error for campus Document Camera usage is the second largest with a value of 17.18. This is 0.6% of the mean for campus Document Camera usage. The Pseudo R-square for the campus Document Camera usage was found to be .78. The rest of the four sources have less Mean Error of the residuals along with a smaller Pseudo R-square.

From Table 2, we can tell that the models we constructed are of a decent quality. Considering the difficulties in constructing the models themselves, such as too many zeros due to low usage during breaks (i.e. Spring Break, etc.) and the weekends, we are satisfied with the quality of the six time series models.

Table 2: Evaluation of Time Series Models

Names	***ME***	***RMSE***	***MAE***	***MASE***	***Pseudo R-Square***
PC	22.6	2088.9	1228.79	.36	.73
DC	17.18	1778.23	1152.33	.37	.78
LA	9.96	491.7	298.84	.49	.56
DVD	1.74	162.45	100.97	.63	.42
CT	.19	88.11	41.43	.84	.12
WPS	.0083	43.39	24.27	.63	.4

5. Conclusions

From looking at the charts that have been generated for this data set, we can clearly see that the PC, Document Camera, and Laptop connections were the most used devices on campus during the data sets time frame, while the DVD/VCR player, Cable TV, and WPS were clearly used the least for this same time frame. The time series results attest to the fact that the DVD/VCR player, Cable TV, and WPS were used the least. The time series analysis used to predict the future usage patterns also did not reveal any obvious increase in the trend of the three resources. Judging from these results we should consider elimination of these least used devices from future electronic classroom designs. However, the WPS is a newer technology to many instructors and it might not be widely used simply due to the lack of awareness and training.

Therefore we suggest only removing the DVD and the Cable TV from the system in any future upgrades. It can save significant amount of money for the whole campus. We would like to have a larger scale of experimental units. If we can observe a similar pattern in other universities and governmental facilities, then we could provide useful information for decision makers.

6. Future topics

With a rudimentary model of how often classrooms are used and what is used, we can use this information to estimate how much electricity is being used and look for ways to reduce this cost to the schools. Another potential area of interest would be to compare how teachers are using the classrooms versus how the students are using them. Other possibilities would be to attempt to determine if there is any significant difference in student performances after upgrading the ICT compared to before the upgrade.

By early summer of 2015 our goal is to have a new data set that will include around 120 classrooms, along with new variables such as warning/error messages and occupancy sensor data. This new data set should allow us to further refine our model by comparing multiple semesters. These new data set variables should also allow us to determine how many problems are occurring in the classrooms during a semester and if they are truly being used or were simply left on.

References

Abdullah, L. 2012. ARIMA Model for Gold Bullion Coin Selling Prices Forecasting, *International Journal of Advances in Applied Sciences* (IJAAS),Vol. 1, No. 4, December 2012, pp. 153~158

Bowerman B.L., R. T. O'Connell, A.B. Koehler. 2005. Forecasting, *Time Series and Regression:An Applied Approach*. Thomson Brooks/Cole, Belmont CA.

Brill, J. M. and C. Galloway. 2007. Perils and promises: University instructors' integration of technology in classroom-based practice. *British Journal of Educational Technology*, Vol 38 No 1 2007 95–105.

Gülbahar, Y. 2008. ICT Usage in Higher Education: A Case Study on Preservice Teachers and Instructors. *The Turkish Online Journal of Educational Technology*, 1303-6521 volume 7 Issue 1.

Hyndman, R. J. and Y. Khandakar. 2008. Automatic Time Series Forecasting: The forecast Package for R, *Journal of Statistical Software*, July 2008, Volume 27, Issue 3

Roberson, J. A., C. A. Webber, M. C. McWhinney, R. E. Brown, M. J. Pinckard, J. F. Busch. (2004). After-hours power status of office equipment and energy use of miscellaneous plug-load equipment. Report number LBNL-53729.

Shenstone, L, R. J. Hyndman. 2005. Stochastic Models Underlying Croston's Method for Intermittent Demand Forecasting. *Journal of Forecasting*, 24, 389-402.

Shih, S. H. and C. P. Tsokos. 2008. A temperature forecasting model for the continental United States. *Neural, Parallel Sci. Comput*. 16, 1, 59-72.

Shih, S. H. and C. P. Tsokos. 2008. Prediction models for carbon dioxide emissions and the atmosphere. *Neural, Parallel Sci. Comput*. 16, 1, 165-178.

Shih, S. H. and C. P. Tsokos. 2009. A new forecasting model for nonstationary environmental data. *Nonlinear Analysis, Theory, Methods and Applications*, Vol. 71, No. 12. (2009), pp. e1209-e1214

Usluel, Y. K., P. Aşkar, & T. Baş. 2008. A Structural Equation Model for ICT Usage in Higher Education. *Educational Technology & Society*, 11 (2), 262-273.

Vasileiadou, E. and R. Vliegenthart. 2013. Studying dynamic social processes with ARIMA modelling, *International Journal for Social Research Methodology*, DOI: 10.1080/13645579.2013.81625

Webber, C. A., J.A. Roberson, M. C. McWhinney, R. E. Brown, M. J. Pinckard, J. F. Busch. 2005. After-hours Power Status of Office Equipment in the USA. Report number LBNL-57470.

Figure 1 Device Usage during "Prime Time" in Cook Hall 317

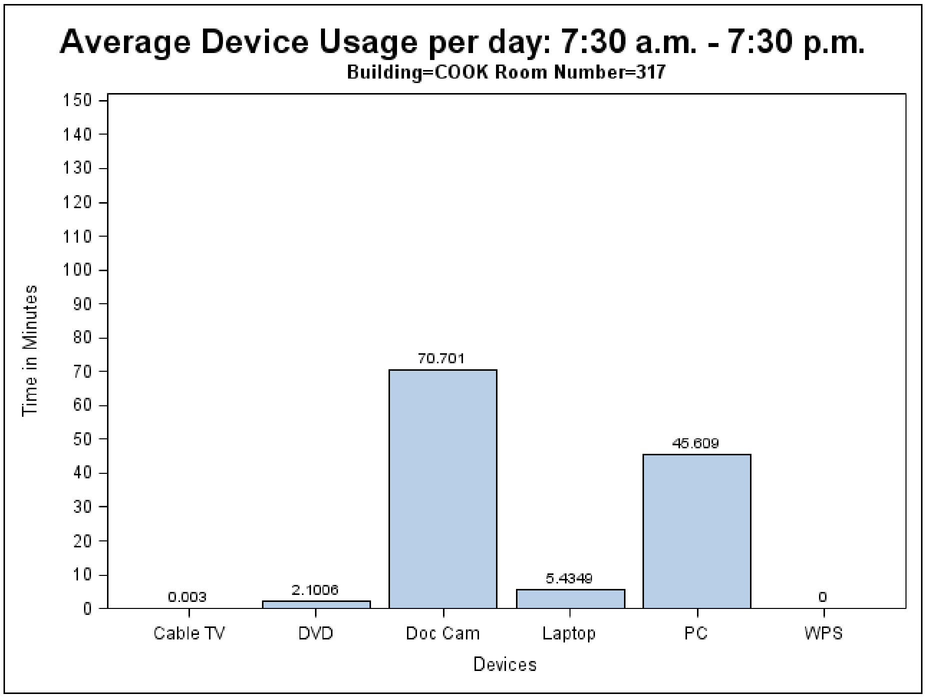

Figure 2 Device Usage during "Prime Time" for the Whole Campus

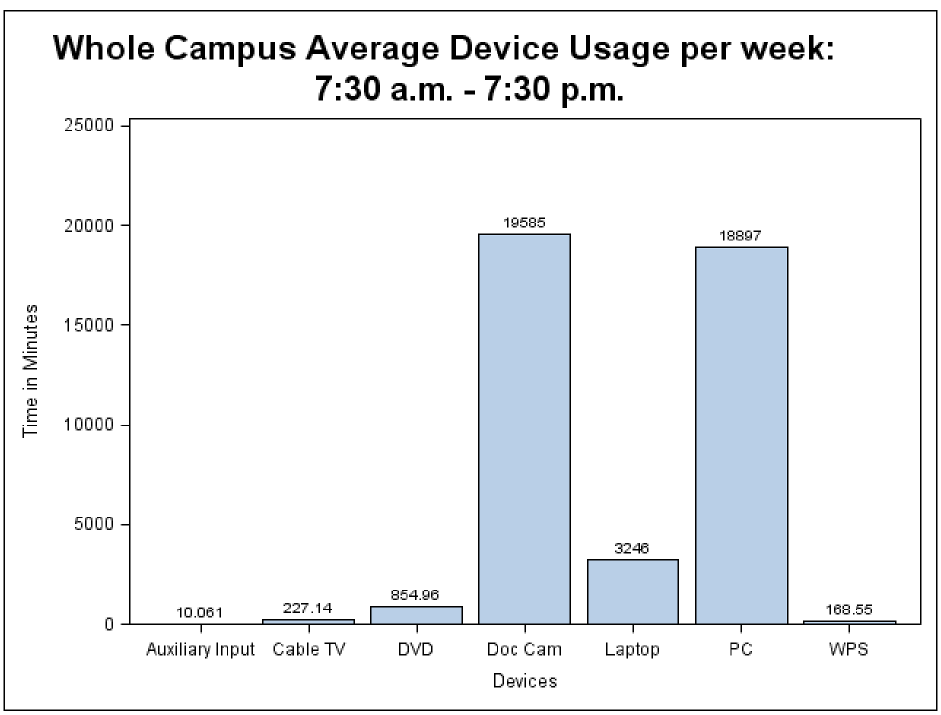

Figure 3 Plot of Campus Time Usage Data for PC

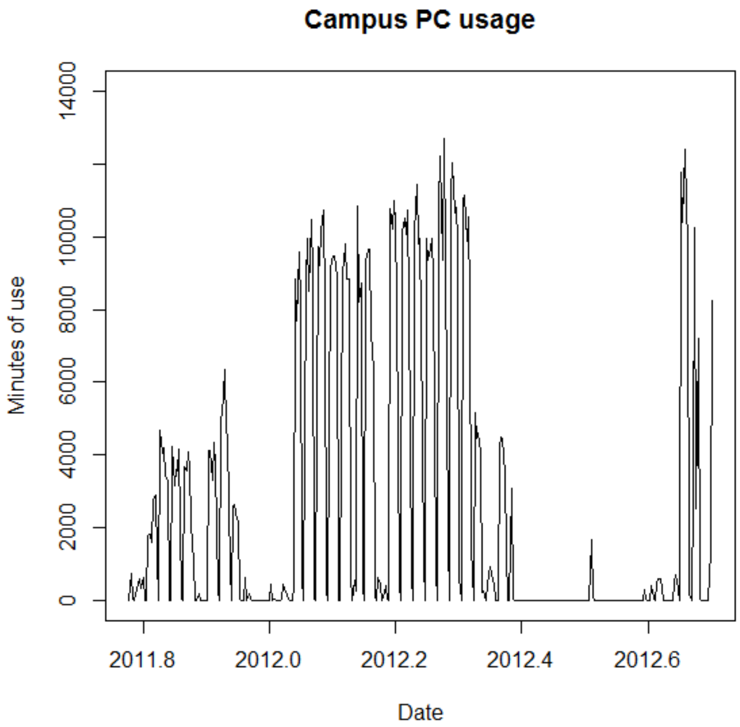

Figure 4 Plot of Campus Time Usage Data for DVD

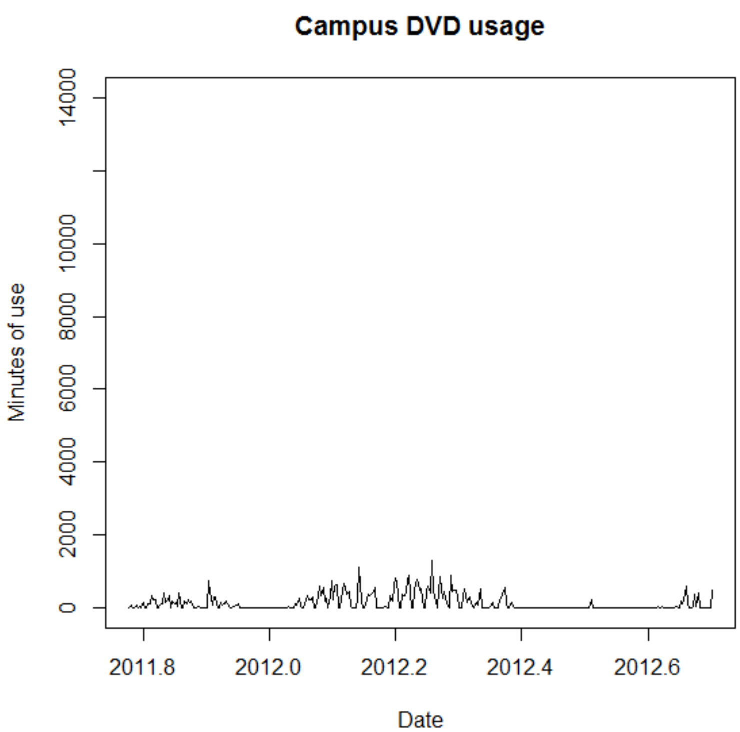

Figure 5 the 30-day Forecast of ARMIA (4, 1, 5) for PC

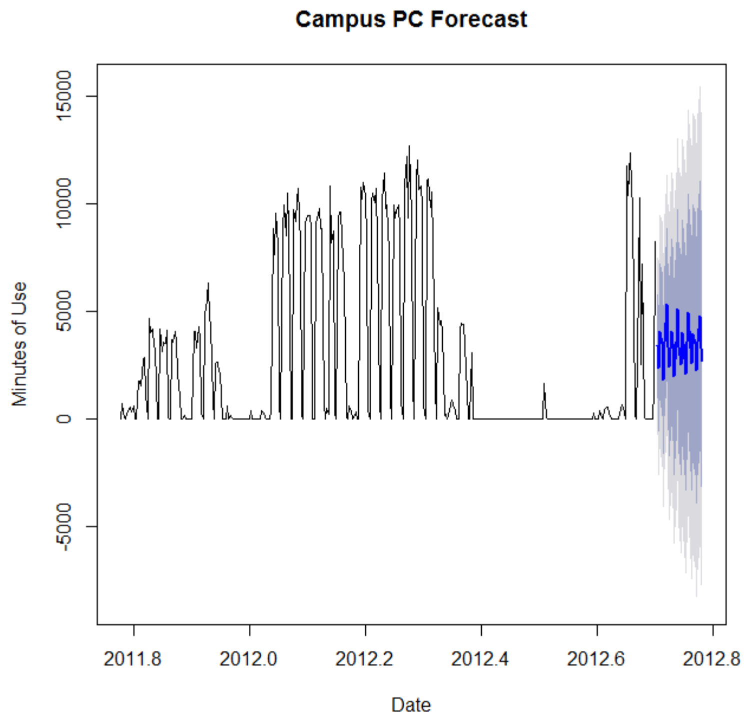

Figure 6 the 30-day Forecast of ARMIA (4, 1, 5) for Document Camera

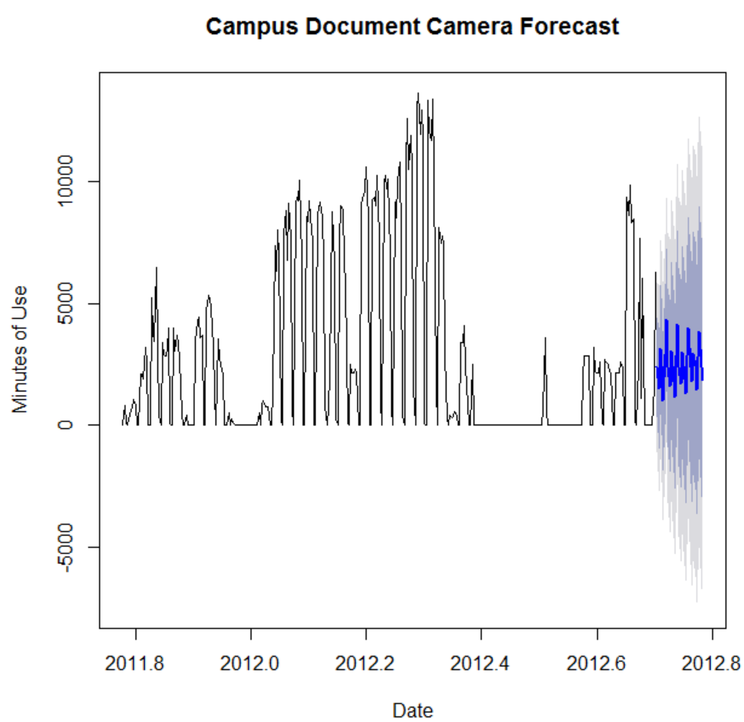

Figure 7 the 30-day Forecast of ARMIA (2, 1, 1) for Laptop

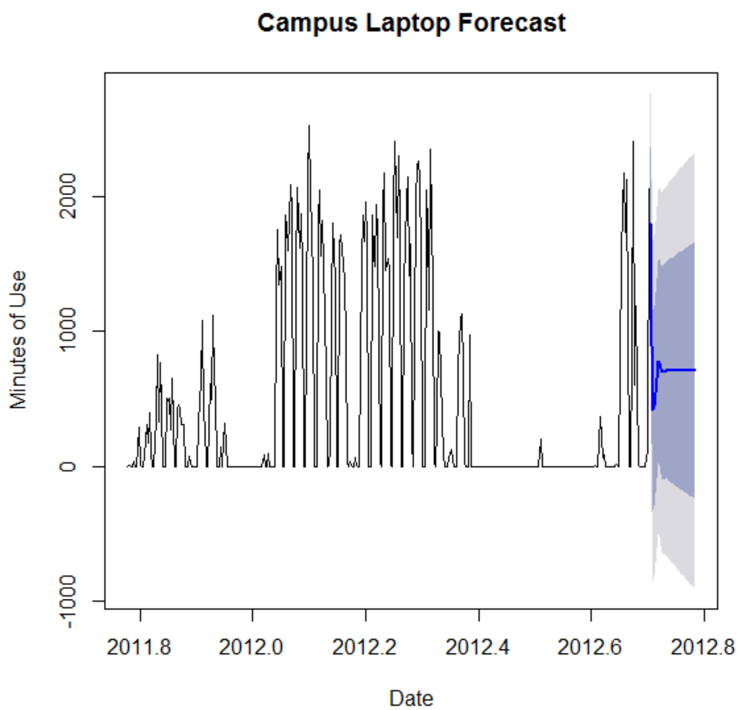

Figure 8 the 30-day forecast of ARMIA (3, 1, 4) for DVD

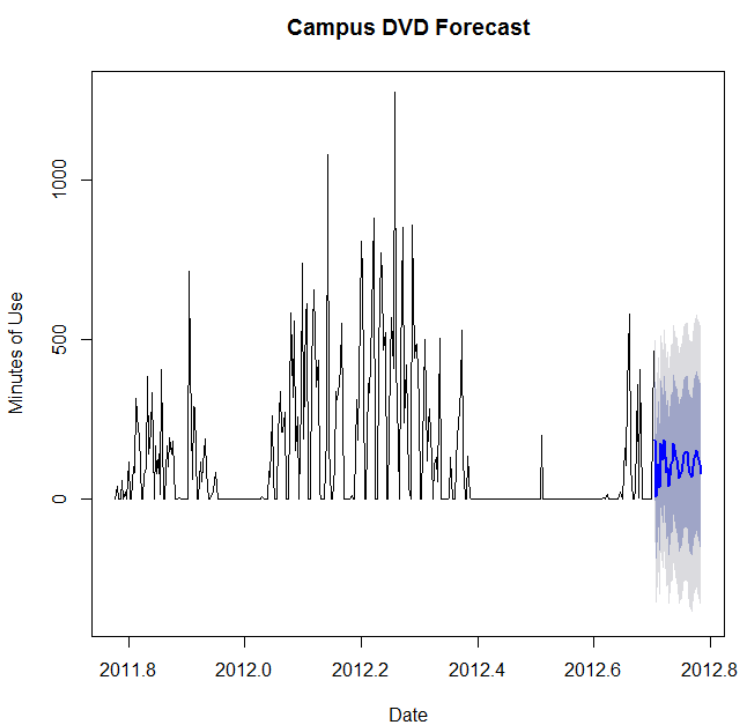

Figure 9 the 30-day forecast of ARMIA (0,1,1) for Cable TV

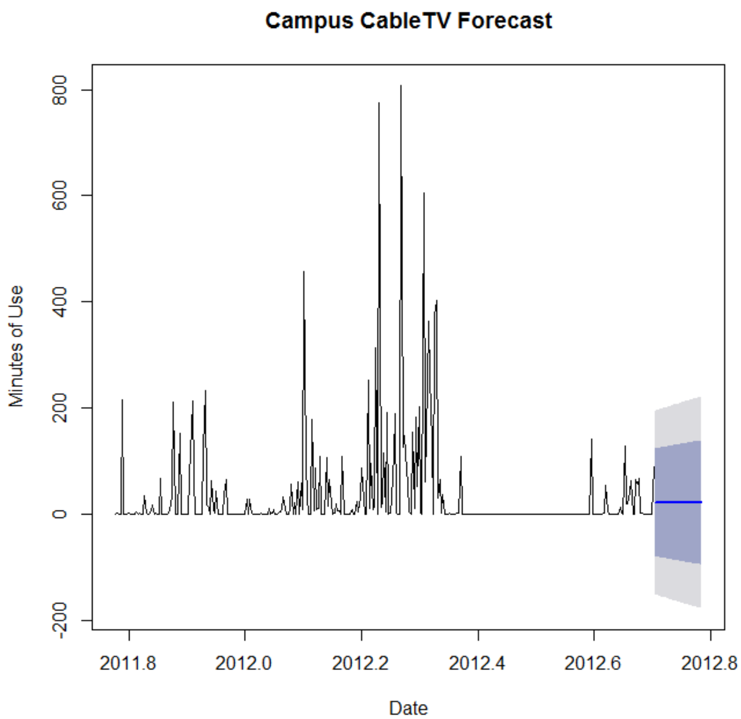

Figure 10 the 30-day forecast of ARMIA (3, 1, 3) for WPS

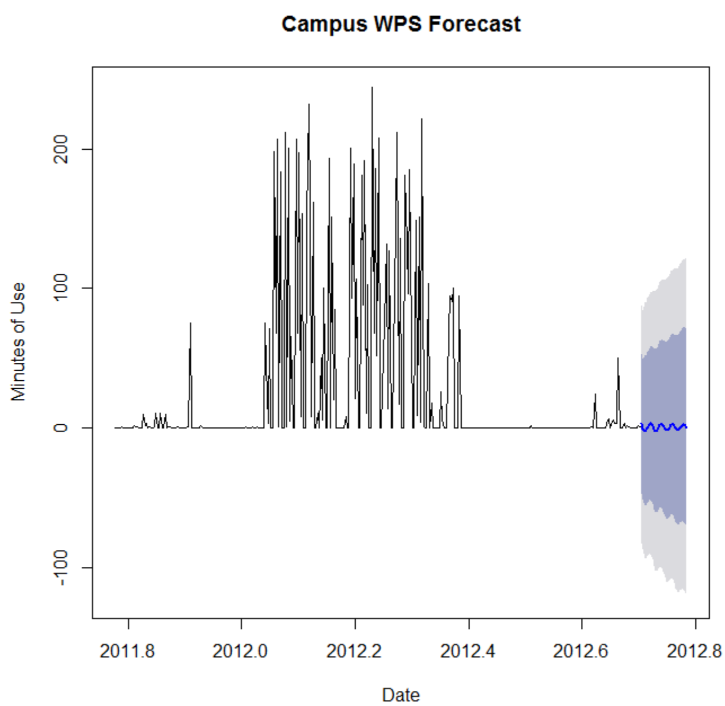

Figure 11 Different Sources of Device Usage for the Whole Campus