Spatial Energy Visualization and Nighttime Anomaly Detection for the University of Virginia's Living Link Lab

Natalie Bretton[*]
khc6du@virginia.edu
University of Virginia
Charlottesville, Virginia, USA

Seth Chase[*]
wma7uc@virginia.edu
University of Virginia
Charlottesville, Virginia, USA

Abigail Corish[*]
rdt5su@virginia.edu
University of Virginia
Charlottesville, Virginia, USA

Taran Gupta[*]
vxx6yw@virginia.edu
University of Virginia
Charlottesville, Virginia, USA

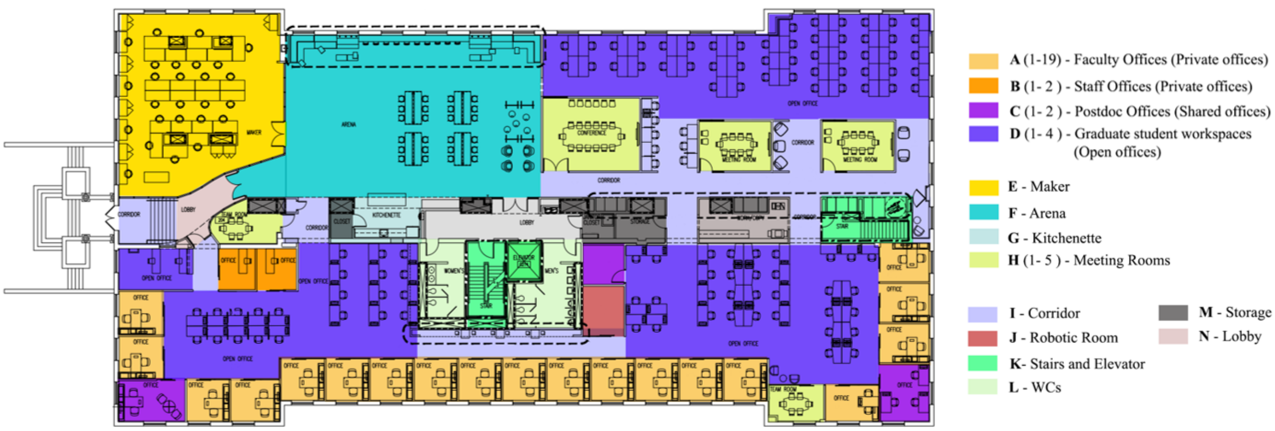


Figure 1: Floor plan of the Link Lab with labeled occupant spaces [3]

Abstract

This paper presents an integrated workflow for spatial energy visualization and nighttime anomaly detection within the University of Virginia's Living Link Lab (LLL), a cyber-physical research environment equipped with extensive sensor infrastructure. The system aggregates minute-level electricity consumption data from InfluxDB, processes them through Python and GeoPandas, and then links them to a georeferenced CAD/QGIS floorplan of the LLL to produce an intuitive room-level energy map. Traditional methods require a building manager to manually analyze thousands of data points collected from dozens of different sensors. This spatial representation enables rapid identification of high- and low-consumption zones and supports pattern recognition that is not readily accessible through raw time-series data. By creating a visual, color-coded, power-based map, occupants and facility managers can have a better understanding of how their spaces utilize energy and if there are sensor errors or rooms that are energy-intensive. Building upon this spatial mapping, the project implements a multi-model anomaly detection framework that combines rolling Z-score, IQR outlier analysis, and K-means clustering to identify potentially unnecessary nighttime loads at the room scale. The workflow is fully integrated into Microsoft Power BI, providing adaptable dashboards for comparative and near-real-time visualization. Together, the system reduces analytical barriers for building managers and occupants while facilitating targeted interventions to reduce wasteful energy use. Additionally, it establishes a scalable foundation for the future expansion into multi-variable analysis and ML-based automated building management.

Keywords

Smart Buildings, Energy Visualization, Spatial Mapping, Cyber-Physical Systems, Sustainability, Anomaly Detection, Nighttime

1 Introduction

The Link Lab is a University of Virginia (UVA) multidisciplinary, cyber-physical systems research environment, opened in January 2018, that supports work ranging from robotics to smart and connected health [1]. The Living Link Lab (LLL), a dedicated platform within this space, operates as a long-term, sensor rich environment for studying human-building interaction and the promotion of sustainability [7]. More than 100 occupants use the 17,000-square-foot facility, which contains sensors monitoring power, temperature, humidity, air quality, noise, and more. A diagram of the Link Lab is

[*] All authors contributed equally to this research.

shown above in Figure 1. The LLL's sensing infrastructure allows for the monitoring of occupants' states, along with human behavior. It can be used for energy optimization and sustainability studies or to gain insights into occupant health, comfort, and productivity. Although this extensive infrastructure produces large volumes of time-series data, meaningful interpretation is often difficult. The patterns of energy use that matter most to building managers or sustainability researchers are typically buried beneath thousands of raw readings. This paper addresses this challenge by developing a workflow that converts raw sensor data into an accessible spatial map of energy use and establishes an anomaly detection framework capable of identifying nighttime irregularities. To achieve this, all of the data is gathered and stored in InfluxDB, an open-source database platform specializing in the storage and retrieval of time series data [6]. This work advances the LLL's goal of building a robust, user-friendly foundation for ongoing research into smart, energy-efficient environments.

2 Problem Statement

Despite the availability of detailed energy data, there remains a gap between the volume of information gathered and the ability of non-technical users to derive actionable insights. Building managers and occupants often encounter sensor outputs in complex time-series formats, where meaningful relationships are not immediately intuitive. The examples of weekly power usage show how easily important trends become obscured by raw numerical data. Figure 2 demonstrates this point. This project will focus on power data, a direct indicator of the electricity use in different spaces, which, as shown, can prove difficult to analyze in its complex, raw format.

A second challenge involves unnecessary nighttime power use. This can include such loads as lights or equipment left on, equipment restarting, or phantom detections from sensors and sensor based systems. These energy expenditures provide no utility during unoccupied hours yet contribute to carbon intensity and operational cost. As universities nationwide face rising utility expenses and budget constraints, reducing such waste becomes increasingly important.

To address these challenges, this project focuses on two key tasks. First, developing an intuitive spatial visualization of room-level energy use and, second, establishing a nighttime anomaly detection method that can guide corrective action.

3 Motivation

Understanding how energy is consumed within indoor environments is essential for promoting sustainability and optimizing building function and operation while also enhancing occupant comfort. However, many building occupants lack *accessible and meaningful* energy data and visualizations that connect their local decisions to broader energy patterns, while managers must sift through dense data sets to identify issues. While there is some existing research involving broader energy planning and efficiency using Geographic Information Systems (GIS) [4] and complex systems utilizing sensors, Internet of Things, and AI within building management [2], this project is quite novel.

The core motivation of this project is to improve energy literacy and operational efficiency by connecting spatial visualization with

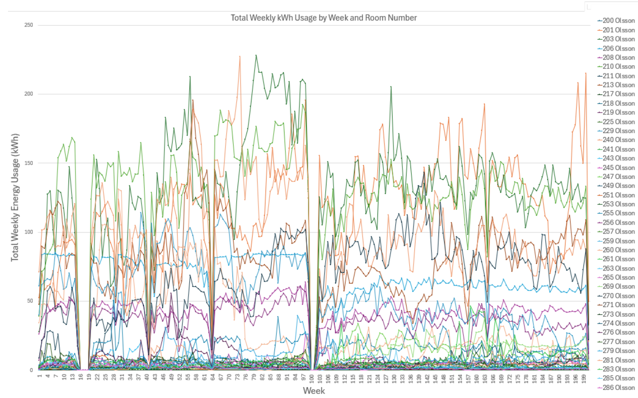


Figure 2: Sensor by room number since 12/30/2021 (204 total weeks) showing the sum of total kWh for each room on a weekly basis. Demonstrates the volume of data the sensors collect, how it can be difficult to understand in its raw format, and why only one week of data was mapped.

anomaly detection. By placing energy data in a spatial context and emphasizing nighttime irregularities, the system enables users to identify wasteful patterns quickly. The work contributes not only to cost savings and emissions reduction for the School of Engineering and Applied Science (SEAS) but also to the Living Link Lab's broader mission of supporting human-centered and sustainable building research. Building energy represents a major operational expense and emissions source. Both occupants/users and owners/managers should have a shared sense of responsibility in reducing excess energy usage, even if to achieve different end goals. Yet, many functions of a building that rely on energy are needed for the health and comfort of its occupants, and some energy usage is determined by the building itself (e.g. Insulation efficiency). This project will address two main goals that improve energy monitoring and usage:

(1) *Providing a more intuitive way to understand energy use:* A floor-plan based visualization tool will help bridge the gap between raw sensor output and intuitive human understanding. A spatial map allows occupants and managers to better understand where power usage spatially occurs throughout a building, and helps identify historical patterns and trends in power, along with enabling data comparison.

(2) *Identifying areas of action:* The nighttime anomaly detection tool utilizes energy mapping in a way that provides financial and environmental incentives for spatial visualization. If unnecessary power usage can be identified and addressed, the University can save money on utilities that would have otherwise gone to waste. The development of a visualization tool that allows for the correction of energy waste furthers the LLL's goal of being a testbed for cyber-physical systems research with a particular focus on fostering human-building interaction and contributing towards sustainable development.

4 Methodology

The integrated system that links spatial visualization, comparison, and automated anomaly detection is composed of a few distinct parts: Data Aggregation and Spatial Mapping, Comparative Visualization, Nighttime Anomaly Detection, and Error Injection. The sections below will detail the processes and steps taken to make each of these parts of the system functional.

4.1 Data Aggregation and Spatial Mapping

The InfluxDB for the LLL was connected to VS Code (Python) using a secure connection. Only sensors that collected wattage power data in the Link Lab were selected for analysis. A list of relevant room numbers was compiled so that a 2D floor map could be produced accurately and appropriately labeled. Only sensors in rooms within the LLL were utilized as these sensors were able to be tested, and their exact location was known.

When collecting data, it was decided to aggregate sensor data by location (room number) and not by device ID. An initial sum of data by room number produced a result with a 400% faster runtime and produced results equal to those obtained when sensor data was summed based on device ID. Data was aggregated over 1 week, and each sensor utilized collected data once a minute, meaning that each sensor produced 10,080 individual lines of data. It was decided to utilize a one-week block of time, as how both an academic and office building are used, which tends to vary day to day but is consistent on a week-to-week basis. A weekly total accurately accounts for weekends when it is expected that power usage is at its lowest. The use of one week as the study block allowed for all sensor readings to be used without bogging down and prolonging the analysis period. Further, if the spatial map is to be used as a predictive tool, the events of the past week are more predictive of future behavior than what occurred years ago. It was chosen to take a room-by-room total, as how a room is used is more intuitive compared to an individual outlet. Most building occupants and building managers know how a space is generally used or who is occupying the space, but they do not necessarily know how each outlet is being utilized. Further, how an outlet is utilized is fluid; an individual may not always plug their device into the same set of outlets, but how a space is used tends to be less fluid.

Based on the parameters chosen above, the total energy usage of the room was found per week. A CSV (Comma Separated Value) file was created to organize the data and allow for future integration into a 2D floor plan of the space. This was chosen because its universality would allow for seamless integration between the data set and the floor plan and they can be constantly written over, eliminating the need to rewrite code whenever an updated analysis is desired. Next, Computer-Aided Design (CAD) software was used to make a diagram of the floor plan that could be exported to a .dxf file. This was done by importing an image of the Link Lab floor plan [3], and drafting the perimeter of the lab and rooms that were equipped to track power use (W).

While the drawing file was a crucial step to create a format that Python would be able to understand and map, the floor plan had to be converted to a GeoPackage (.gpkg) file. The spatial map is ultimately created using a geospatial library called GeoPandas, which is not built to read drawing file formats. Instead, the file must

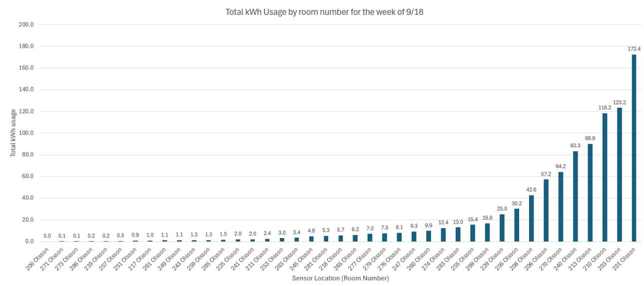


Figure 3: Bar Graph of energy usage in kWh by sensor location (room number) for the week of 9/18/2025

be changed into a more structured format that can define the spatial relationships between different objects, or in this case, rooms. To convert the .dxf into a .gpkg file, a free and open-source application called Quantum GIS (QGIS) that specializes in spatial visualization was used [8]. Using QGIS, the drawing file can be imported, and the lines that were used to define the various rooms of the lab in CAD can be labeled as different polygons with a corresponding Room ID. It is critical to note that the name assigned to each closed polygon in QGIS must match the variable name used to identify the rooms in the CSV file. If these names do not match up, the code will not be able to correlate the energy value from the CSV file to the correct spatial location on the visual map.

Once an accurate floor plan was created and could be read by GeoPandas, it was able to be referenced alongside the CSV file. The CSV was utilized to create a bar graph of kWh for the week of 9/18/2025 so that the resultant map could be crossed checked to ensure accuracy (Figure 3). The spatial map and CSV file could be cross-referenced based on the shared parameter, room number, and the final spatial map was created using the GeoPandas Python package. An overview of the overall framework for this approach, along with the main design features that make it work, is shown in Figure 4. These steps are what allow the raw sensor data to be easily visualized throughout the Link Lab.

4.2 Comparative Visualization

A Microsoft Power BI model of the data was created and published. This methodology was chosen as it allows for a third-party user to have easy access to the data in a purely visual format, its use is very intuitive, and a user can customize what time period of data and which room they want to analyze. Power BI can be easily updated by the developer, allowing for the time period of analysis to be easily altered. Further, it allows a user to select which rooms they want to visualize. This allows occupants of various spaces to compare the energy use of their office relative to those occupying spaces that are of similar size and function. The model was developed to visually analyze sensor data for a period of 104 weeks (2 years). This block of time was chosen since any larger period proved to have produced too much data for Power BI to analyze in a reasonable amount of time. Power data was summed over one day so that accurate slicing of the data could be integrated into Power BI. The dashboard allows the user to select the time period they want the visual analysis to be conducted on (Link Lab Power BI Dashboard).

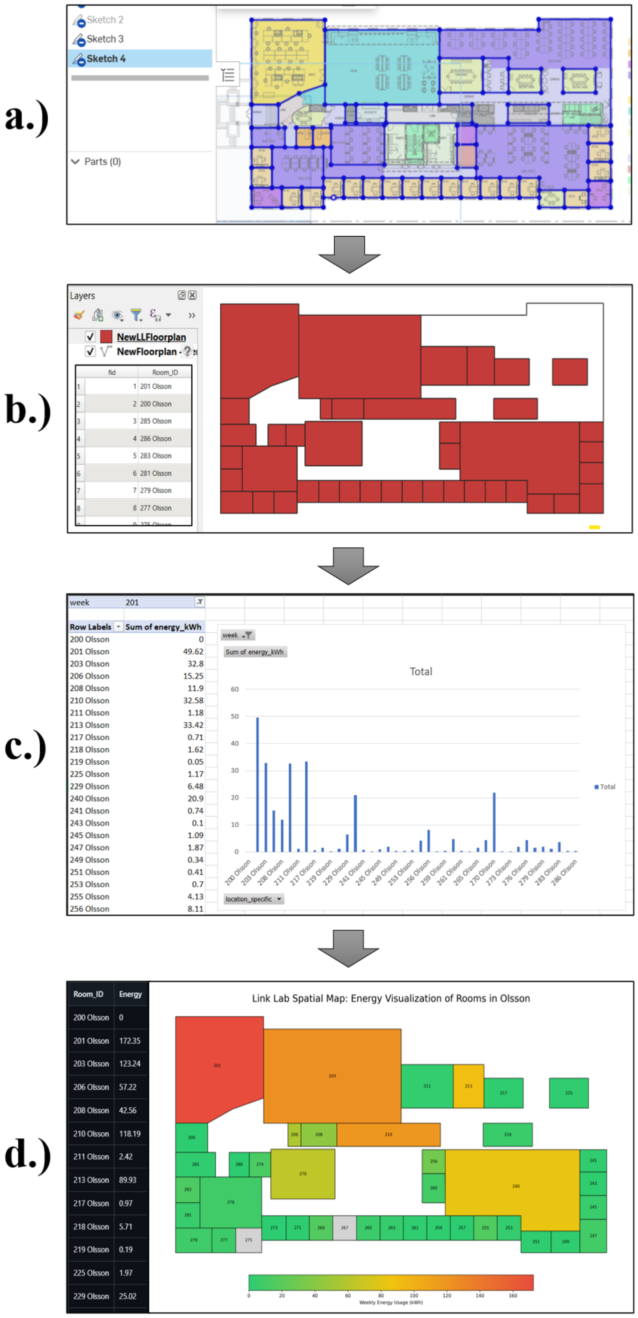


Figure 4: This flow chart summarizes the steps involved in visualizing the raw sensor data spatially. Part a.) depicts the development of a .dxf CAD file for the floorplan, part b.) shows the drawing imported into QGIS where each of the rooms is assigned a feature ID number (FID) that corresponds to a room ID, part c.) shows the use of an Excel pivot chart to gather specific weeks of data for further analysis, and part d.) shows the corresponding CSV data from the Excel (with columns labeled accordingly) being cross-referenced against the .gpkg file to create the final spatial diagram.

Figure 5: Power BI dashboard analyzing the daily average kWh usage from 12/3/2023 to 11/30/2025. The model automatically color codes the data bars and the map based on average energy usage. A stacked bar chart and floor plan show average daily kWh usage for the selected period, while the line chart shows the daily usage for the selected period.

The model was created by selecting Python as a data source and copying and pasting in the code developed above, adjusting the summing interval and the time period analyzed as described above. To integrate the floor plan model into the dashboard, so that it can easily be updated by the user, the map developed above was then converted into a TopoJSON file. Power BI is then able to convert this file into a shape map, which allows the map's room coordinates to be linked to the data sets' room labels, and subsequently color-coded to be assigned based on average power usage. This then allowed the data on power usage to be applied to the floor map, enabling users to easily see how power usage varies from room to room in the Lab (Figure 5) .

The stacked column chart and the line chart were also helpful, as they allow a user to better understand how the power use of a selected room and the Living Link Lab as a whole have changed over time. Moreover, they help provide a connection between the floor plan and the exact energy usage.

4.3 Nighttime Anomaly Detection

This project focuses on a room-level anomaly detection of Olsson Hall's Living Link Lab, which is equipped with circuit power sensors that pick up loads from all of the electric outlets and lighting. The detection model is a proof-of-concept, and many steps remain between this paper and functional deployment; these steps will be discussed in the Future Work section and throughout this section. The model itself is broken up into two main algorithms, a detection pipeline and an error injection testing pipeline. The accompanying code can be found in the linked GitHub Repository [5]. Given the time-constrained nature of this project, the two algorithms largely act independently. The next major step for model advancement would be linking the algorithms and using ML to tune various parameters.

A core concern of this model approach is accuracy. By nature, collecting "true" data is extremely intensive. To know if a light or computer is on in the middle of the night for a whole building floor would require a human present to monitor the building's state throughout the night. Not only is this impractical, but having an occupant present may skew data from normal usage, especially for sensor-based systems. Recordings could also be taken, but this is once again a labor and storage-intensive process. Because of this, it was decided that the model be manually tested by inserting artificial power anomalies to determine whether the model can detect them or not. This method is not as effective as "true" training data, nor does it assess the validity of the data itself, but it can at least improve model performance, assuming the injected errors model the anomalies that are desirable to detect. In other words, if the anomalies can be simulated, they can be detected.

A key tenet, along with accuracy, is false positives versus false negatives. False negatives fail to reduce excess energy usage, whereas false positives likely increase the amount of human intervention that may take place in the evening or during the night, for example, by a building manager or security staff. Because of this, our detection algorithm uses three different models and only flags an anomaly if two out of three agree. The steps for the detection pipeline are listed below explained in further detail:

4.3.1 Step 1: Querying Power Data. Power data is summed over the selected time frame (default = 15 min) by room and converted to kilowatt-hours. The data is then stored in a pandas data frame.

4.3.2 Step 2: Helper Functions. Three helper functions are then defined, which convert the data into more usable formats for analysis. For example, they may filter the data frame to just a single selected room, add time features for data categorization and time zones, or create a new date column where AM hours of the night are classified as the previous day for analysis.

4.3.3 Step 3: Run Models. Three models are used, all based on different statistical principles. As mentioned above, the goal is to reduce the number of false positives and increase model robustness given the lack of training data associated with the "truth state" of the building.

(1) *Rolling Z score:* The first model is a rolling Z-score model. The model takes in only the past two weeks' worth of data, which is assumed to be representative of normal building use. The average and standard deviation are calculated for that time frame, and any values that are above two standard deviations above the mean are flagged as anomalies. The function then returns a data frame with a sigma column for each observation as well as a boolean column that returns if the observation is flagged as an anomaly.
(2) *IQR Outliers:* Similar to the previous model, the inter-quartile range (IQR) is computed for night-time hours. Any values that are above the IQR are flagged as anomalies. Unfortunately, this is one model where the current deployment does not match our intended goal. The model is intended to subset the data and create different IQR for different time categories, such as weekdays versus weekends and semesters versus breaks, but time limitations prevented us from developing it thoroughly. This will be expanded upon further below.
(3) *K-Means Clustering:* This machine learning algorithm classifies all observations into one of two clusters, intended to represent points that are "normal" and those that are considered "outliers", or anomalies. Five features are included in the model: mean, standard deviation, max value per night, percent above 5 W (arbitrary baseline), and number of power step-changes. If the model believes all the points are similar and clusters them together, it is possible for it to only return one cluster. This does not represent a problem with the model, but rather that no anomalies were detected.

4.3.4 Step 4: Combining Model Outputs. The outputs from the previous models are all combined in this function. Every model produces a data frame with a Boolean column representing whether a time step is predicted as an anomaly. This function takes those outputs and produces an additional column with a boolean value, set to trigger by default as true if two of three models agree that a time step is an anomaly. The function can also output if any one model flags a time step or if all three agree.

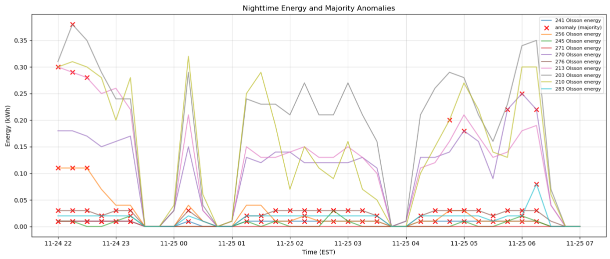
4.3.5 Step 5: User Evaluation. Three metrics were chosen as final indicators for anomaly detection. The first is the anomaly rate per night. The total number of nighttime data points, as well as the number of anomalies, are counted for a given room. Then, a simple ratio is calculated to determine the number of anomaly points. A high number of points likely indicates something being left on that is not seen in the training data, or that a room was genuinely occupied overnight. The second metric is the persistence of anomalies. Data are gathered every 15 minutes; any sequential anomalies are grouped as one event. The frequency of these event lengths is outputted for examination. The last metric is the Jaccard Matrix. The Jaccard Index is a statistic that represents the similarity of data by dividing the area of intersection between two datasets by their total area of union. Since there are three models, a matrix is computed, representing how similar the models flagged anomalies.

4.3.6 Step 6: Plot Anomalies. This function graphs the energy usage of a selected room overnight. Any anomalies flagged by the three models, plus the combined output, are overlaid on the graph to provide the user with numerical and graphical insight as to what the anomaly event may have been and when it occurred.

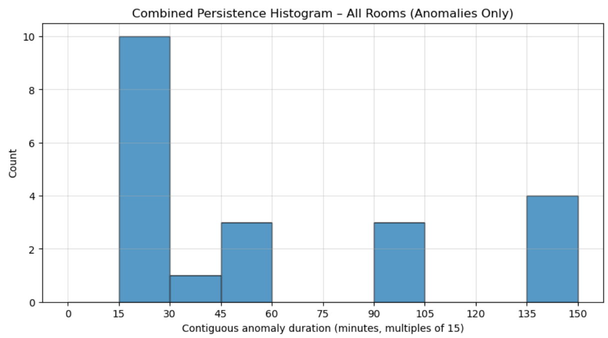
4.3.7 Step 7: Entire Floor. Metrics are recomputed for the entire floor, as opposed to just one room. For example, another persistence histogram is created, this time combining anomalies across the entire floor for a single night. An overall anomaly ratio is also calculated. Together, these two metrics could also calculate the amount of potential energy savings available for a given night. The last output is the same anomaly plot over an energy usage graph, but now includes all rooms that have at least one anomaly.

4.3.8 Visualizations and Results. An example of the pipeline's results is included in Figure 6. The overall computed metrics are included in the output along with the two graphs shown. In addition, the number of anomalies per night per room are saved in an external file so the data can be mapped spatially in PowerBI. The spatial mapping is powerful for gaining quick insight into which rooms may have excess energy usage and provides the most direct utility to end users who may have limited technical knowledge and time.

Power BI mapping also shows how the room's current and historic power usage could be directly linked to the average number of anomalies seen over the indicated period. The CSV file that resulted



(a) Energy anomaly outputs.



(b) Histogram anomaly outputs.

Figure 6: Two graphical outputs of the anomaly pipeline algorithms, also shown spatially in PowerBI.

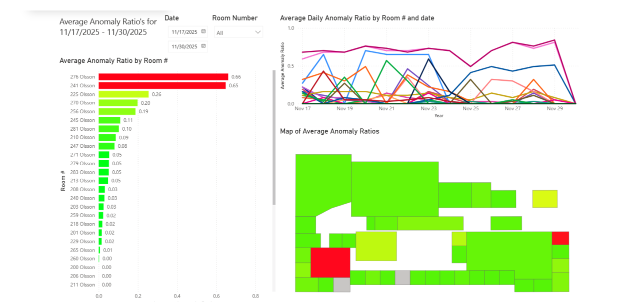


Figure 7: Mapping of Anomaly Outputs

from the above calculations was imported into Power BI to allow for plotting based on the indicated room number (Figure 7).

A nighttime anomaly detection dashboard was developed in Power BI, allowing users to quickly identify potential anomalies before utilizing the above code to better understand if anomalies are common in that room. The Power BI model looks at the past seven days of data and sums over 5-minute intervals. A seven-day interval was chosen so that a user will always be able to compare their current time usage to nighttime usage on another weekend/weeknight. 5-minute intervals allow a user to get close to real-time data without bogging down the system, leading to slow analysis. However, it should be noted that the current power dashboard requires the user to download the model and manually reload it so that the data can be updated.

4.4 Synthetic Error Injection

Since the data is unlabeled, it is difficult to directly measure the efficacy of the models in detecting anomalies. The goal of the synthetic error injection pipeline is to manually insert random power values to see if the model can detect if they are anomalies. When combined with the detection pipeline, models and global parameters can be automatically adjusted through ML to get the highest performing predictions. Examples of adjustable parameters include: necessary deviation from the mean to be considered an anomaly, types of features to include in clustering, and the number of days to consider for model training. The pipeline is described in greater detail below:

4.4.1 Step 1: Anomaly Injection. The data is injected with N number of anomalies, lasting a random amount of time between minimum and maximum bounds, and representing a random positive delta power between bounds. The data is saved to a new data frame. This process will be expanded upon in the Future Work section. It is important that these injections represent the anomalies we seek to detect, which is no simple task.

4.4.2 Step 2: Detection Functions. The data frame is passed through the same three detection models with the same parameters. The outputs are also combined by majority vote into a fourth additional boolean column. The rate of anomaly detection determines model performance, and model parameters can be changed accordingly to maximize detection while minimizing false positives.

4.4.3 Step 3: Pipeline Outputs. The mean is taken of each model's output, including the combined output. Since the output is one boolean column representing the existence of anomalies and only values with injections are preserved by this step, every row should be flagged as true for all models. In other words, the ideal mean value is 1, whereas the worst possible value is 0.

5 Results

The results from the integrated system are divided into several sections that detail the validity of the initial spatial visualization, Power BI work, and anomaly detection.

5.1 Spatial Visualization

The spatial mapping workflow successfully produced a 2-D representation of energy consumption within the Living Link Lab, as shown in Figure 8. A week in the middle of September 2025 (week 195) was used, and power data was mapped onto the floor plan, with a continuous color scale ranging from green (low consumption) to red (high consumption). This gradient enables rapid visual identification of energy-intensive and energy-efficient spaces. The spatial map used the total value of 10,080 different readings.

Across the analyzed period, the map revealed discernible spatial trends in energy usage. Collaborative and shared work zones, specifically Rooms 201, 203, and 210 exhibited higher weekly energy consumption relative to individual offices, likely reflecting greater equipment density and occupancy levels. Conversely, the peripheral, personal office spaces displayed lower and more stable consumption patterns. There was some slight differentiation between individual offices, indicating that most occupants utilize

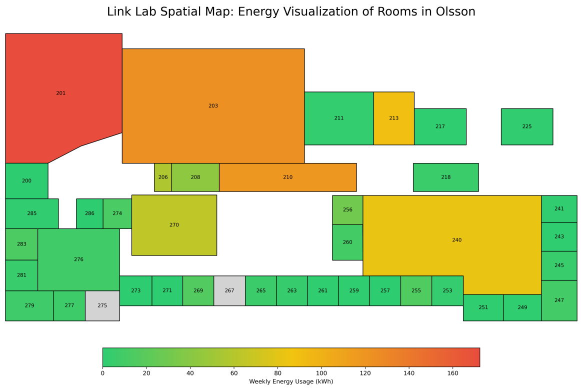


Figure 8: Link Lab Spatial Map: Total Energy Usage (kWh) aggregated by room number for the week of 9/18/2025

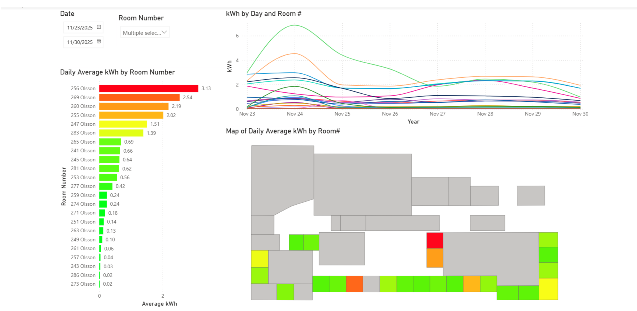


Figure 9: Power BI Visualization of 11/23/2025 to 11/30/2025 for a select group of rooms with a smaller spatial footprint. Rooms were chosen purely based on their relative size.

their space similarly and for approximately an equivalent amount of time during the week.

These preliminary results validate the functionality of the data aggregation and visualization process, demonstrating that the workflow can translate large quantities of raw sensor data into interpretable spatial insights. The resulting map provides a foundation for subsequent analyses, including anomaly detection, energy optimization, and comparative evaluation across time periods. While currently retroactive, by gathering more information on building use patterns, we plan to create a more predictive model.

5.1.1 Utilizing Power BI. The Power BI model allows for any single day of Living Link Lab energy usage to be easily visualized. This allows a user to easily see potentially energy latent spaces that could undergo retrofitting. Such a use includes looking at the small rooms in the Lab that have somewhat similar functions. By only selecting these rooms and setting the time period to one week (11/23/2025 - 11/30/2025), it is easier to visualize rooms that have a similar function and square footage (Figure 9).

The preliminary results displayed above illustrate how the 2D floor plan can be transformed into a user-friendly interface that allows users to focus on the time frame and space types that concern them. It illustrates how thousands of pieces of data can be aggregated to be analyzed and understood at a glance, optimizing usability but not sacrificing the quality of data.

5.2 Anomaly Detection

The anomaly detection pipeline was successfully run on the last two weeks of November 2025. The results are included in Power BI for adaptable visualization. When examining the graphical output of anomaly detection across various days, two main anomalies surface. The first is large power spikes that often last a short time, which the models flag. The second is very low power values, generally close to zero, that stay flagged for over two hours. Both cases could represent excess power usage at night, and examination of these events in real-time could lead to power and, thus, cost savings.

Communal rooms and large spaces were more likely to have flagged anomalies when compared to individual office spaces. Most nights have overall anomaly ratios under 0.15, showing that are models are not outputting many, if any, false positives. Examination of Jaccard matrices for most rooms and nights show general agreement in flagged anomalies across models. As previously mentioned, without labeled data, it is impossible to know if these anomalies represent truly superfluous energy use. Furthermore, the model cannot predict whether the energy use can actually be decreased, but it does provide robust detection given the minimal human labor required. In practice, either a person, such as a building manager, or a camera would have to be present to be able to ultimately make conclusions on the model's efficacy and reduce energy usage.

The models handled the uniform random positive delta error injections well. With 50 injections across all 2025 data, all model means were above 0.95, meaning they missed at most 2 of 50 injected errors. The highest performing model was the IQR Outliers, which detected every injection. The combined majority output detected every injection but one. When the two pipelines are connected for training, the number of injections should likely increase significantly to match the size of the dataset; however, just the 50 injections took about 15 minutes of runtime, so model performance is an area of improvement as the data scales.

5.2.1 Power BI. The Power BI anomaly detection integrates real-time power detection, anomaly detection, and 2D floor plan mapping to optimize usability. By allowing for the retroactive daily summary, 5-minute nightly summary, and anomaly averages over a set time period, a user can gain a complex understanding of what "typical" power usage is, how often variance from this baseline is seen, and what current nighttime usage is. When combined, a user can quickly see if their room is seeing an unnecessary peak in power usage due to a device being left on accidentally when it is not in use, allowing them to take adequate action both in the moment and going forward. Olsson 241 was analyzed for both the average power usage between 11/24 and 11/30 as well as the average anomaly ratio during this period. It can then be understood that while there are multiple peaks in power usage in this room, they occur at night when the room is unoccupied. The anomaly detection dashboard further proves that these peaks in power usage are anomalies and should be investigated to decrease overall power usage (Figure 10).

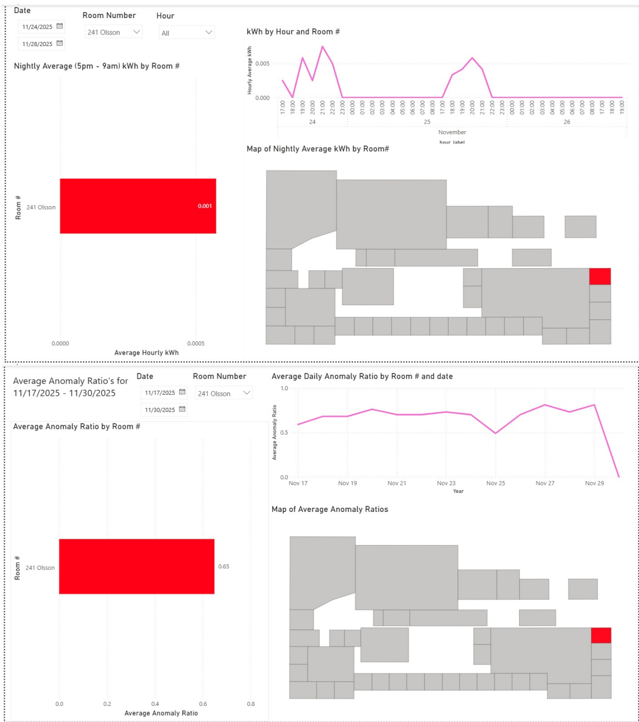


Figure 10: Olsson 241 nightly power usage for 11/24 - 11/28, showing multiple peaks in power usage. Anomaly detection for Olsson 241, showing that anomalies were occurring between the 24th and the 28th

6 Discussion

The development of the integrated energy spatial map system demonstrates the potential of geospatial visualization and advanced analysis to enhance understanding and optimization of complex building energy data. By integrating InfluxDB with Python-based data processing, geospatial visualization tools (GeoPandas, QGIS. etc.), and comparative analysis software (Power BI), the workflow establishes a replicable method for linking time-series energy consumption data to specific spatial locations within a building.

The resulting visualization offers a bridge between quantitative data and qualitative interpretation. It enables researchers, building operators, and occupants to identify patterns of use, inefficiencies, and opportunities for conservation without the need for specialized data-science expertise. Beyond descriptive insight, this framework could inform operational decision-making. For example, the map would allow occupants and building managers alike to review the past week's power usage to see how occupants are using the space. This knowledge can then be used to preemptively prime building systems, such as scheduling the HVAC system for increased use only when high occupancy is predicted, or allowing it to sit idle during expected periods of low usage. Such predictive actions directly contribute to decreasing operational costs by minimizing system run time when certain services are not needed.

Minimizing unneeded energy consumption directly connects to the main contribution of the project, which is providing actionable data by detecting nighttime anomalies. By identifying unnecessary power usage at nighttime and being able to identify and display the spatial location of such an anomaly, it becomes very easy for users to make positive changes. Whether this is an occupant looking to make sure they are not consuming more power than their fellow neighbors or a building manager trying to find areas to lower costs, easy, intuitive visualization benefits everyone. The automated pipeline detailed in the paper limits the human effort that is needed to sort through thousands of data points and connect them back to their real-world meaning. Instead, the anomaly visualization allows non-useful energy to be easily identified and addressed to help limit utility costs that the School of Engineering and Applied Science will have to pay the University. The successful implementation of this detection capability furthers the goal of the Link Lab to act as a testbed for human-building interaction and increasing sustainable development practices.

6.1 Areas of Risk

Since this project is complex and has several different parts, there are a few risks that should be mentioned. First off, one of the main difficulties faced was the lack of labeled training data. The models have to simulate and detect whether or not power discrepancies are considered normal or anomalies based on statistical analysis, not actual behavior. Since labeled data is often not an option, the model pipeline was built to train itself. Any additional training is beneficial and reduces uncertainty; however, the model performs best when the injected errors mirror typical patterns of excess energy use as closely as possible. Also, since there are several platforms involved in the current workflow, including Python/Geopandas, QGIS, and Power BI, there could be some difficulty in the implementation of the tool. This makes it harder for future researchers to address any potential issues with the tool, as there are multiple pieces of software involved. However, this paper has a very detailed methods section that should help alleviate some of this risk. Another potential risk of the project is that there may not be many nighttime anomalies to address. This would limit the amount of financial benefit that could come from visualizing unnecessary power usage at night. Although this is a definite risk, the project still enhances the important feature of data visualization, which in and of itself is helpful. Additionally, even if sufficient power anomalies are not found in the Link Lab, in the future, if there are ever more smart buildings at UVA, this tool could be extrapolated to visualize those spaces and optimize costs there as well.

6.2 Future Work

As previously mentioned, the nighttime anomaly detection model can be expanded upon greatly for greater efficiency and reliability. First, data should be categorized automatically based on a space's normal usage patterns. For example, the LLL is part of a University Building, mostly occupied by graduate students, researchers, and professors. As such, most occupancy occurs during normal 9-5 hours and on weekdays. There are also seasonal variations in occupancy depending on whether the University is on break. Having a model that automatically categorizes for these different time variables would improve the accuracy of the models, as different time frames can be analyzed separately. Power use anomalies are expected to be different depending on building use, so segmenting the data could lead to better-fitted models. Code was developed for time categorization; however, due to a lack of time to properly implement and run on the data, this feature was withheld.

Another vital step is connecting the error injection pipeline with the key parameters of the detection pipeline. This would allow the models to learn on their own, and while it cannot guarantee the validity of the data, it can ensure the models are trained to detect the desired events. Connecting the pipelines is not necessarily difficult to achieve. Instead, the difficulty lies in simulating a variety of undesirable power anomalies correctly. Currently, the injection is any random value between the minimum and maximum recorded power values; however, these injections should become more tailored to represent real-world events, such as lights staying on.

As alluded to previously, the models contain the capability of calculating the amount of potential power saved per night, based on the anomaly detection length and power value. This could be a helpful metric for the end user, as wasted energy use can be directly tied to cost savings, and would provide a better understanding of the scope of nighttime energy utilization.

Future work should also expand the system's analytical scope by incorporating additional variables such as temperature, occupancy, and air quality, thereby enabling multi-parameter correlations between environmental conditions and energy consumption. This would allow for a more complete picture of how occupants utilize a space and would allow for more accurate predictions on how a space is going to be utilized. Implementing temporal interactivity and real-time data synchronization would further enhance its applicability for continuous building performance monitoring. Despite current limitations—such as incomplete data coverage and static temporal resolution—the workflow demonstrates a scalable and adaptable model for advancing smart-building research and sustainable facility management.

7 Conclusion

This project successfully developed and validated an integrated system that helps visualize energy use spatially and detect nighttime power anomalies within the University of Virginia's Living Link Lab. The system directly addresses the challenges of making energy data more accessible and intuitive to everyone while avoiding unnecessary utility costs and environmental impacts.

The integrated system allows a high volume of time series data from InfluxDB, which would normally be difficult and time-consuming to analyze, to be easily translated into a highly visual and actionable format. By successfully mapping aggregated energy data on a 2D floor plan using CAD, QGIS, Python, and Geopandas, an intuitive interface was created to provide context on how energy-intensive different spaces in a building are. The ability to compare the energy consumption of various spaces historically and throughout various chosen time periods was further simplified through the use of Microsoft Power BI as a visualization platform. Furthermore, the anomaly detection pipeline using rolling Z-score, IQR, and K-means clustering was successfully implemented to identify and flag potential nighttime anomalies where unnecessary energy use occurs.

When this integrated system is used altogether, it moves beyond passively using sensors to collect data. By utilizing visualization and anomaly detection in conjunction, this project enables both everyday occupants and building managers to easily identify and address potential wastes of energy. This process is highly visual and eliminates any coding or data black box that non-technical users may not be able to understand. Consequently, the energy visualization and nighttime anomaly detection have the potential to initiate corrective action that can directly contribute to lowering utility costs and emissions. These are both outcomes that tie into the University and Link Lab's goals for sustainable development and lowering operational costs. While this project was mainly a proof of concept, the further work section provides a strong foundation for future research, and the detailed methodology section should make it easy to understand the system structure. Finally, this system can potentially be extrapolated to include other University buildings and help reduce environmental and financial impact on a larger scale.

Acknowledgments

The authors would like to acknowledge the guidance and support provided by course instructors Arsalan Heydarian and Brad Campbell throughout the duration of this project. In addition, we would like to thank Saja Alnusair and Hamidreza Nabaei for their guidance and extensive knowledge relating to InfluxDB and the LLL. The team also extends its appreciation to the Link Lab staff and the Living Link Lab research group for granting access to the InfluxDB database. Their contributions were instrumental in the successful completion of this work.

References

[1] 2025. Link Lab. https://engineering.virginia.edu/labs-groups/link-lab
[2] Tasdelen K. Atilgan A. Malinowski M. Cosgun A. Senol R. Akbulut A. Akbulut, L. and A. Petryk. 2025. A Systematic Review of Building Energy Management Systems (BEMS): Sensors, IoT, and AI Integration. *Preprints* (2025). doi:10.20944/preprints202509.1268.v1
[3] Smart and Healthy Buildings. 2025. Living Link Lab and Influx Database.
[4] Ferrari, S., Zagarella, F., Caputo, P., and Dall'O', G. 2021. Mapping Buildings' Energy-Related Features at Urban Level toward Energy Planning. *Buildings* 11 (2021), 322. Issue 8. doi:10.3390/buildings11080322
[5] Taran Gupta. 2025. SaHB-Energy-Mapping. https://github.com/Taran-Gupta/SaHB-Energy-Mapping
[6] InfluxData. 2022. InfluxDB | Real-time insights at any scale. https://www.influxdata.com/home/
[7] Living link Lab. 2024. Living Link Lab. https://livinglinklab.github.io/
[8] QGIS. 2025. Spatial without Compromise. https://qgis.org/