SAS® GLOBAL FORUM 2020 #SASGF

Paper 4714-2020

Achieving Net-Zero: Forecasting Power Usage to Improve Sustainability on Clemson University's Campus

David L. White, Watt Family Innovation Center, Clemson University; Tim Howard, Watt Family Innovation Center, Clemson University; Snowil Lopes, Clemson University, Facilities

ABSTRACT

Clemson Energy Visualization and Analytics Center (CEVAC) is helping lead Clemson University toward a more sustainable future by using analytics and modeling of campus utility data. Leveraging access to multiple data stores and web-based data streams, the CEVAC team is deploying visualization and analytic solutions at a building and campus level. CEVAC is working with an initial ten buildings across campus, using SAS® Visual Analytics™ 9.4 (VA) reports for each building. The CEVAC student team developed data pipelines and databases for a diverse data environment that ingests data from campus facilities databases, daily and hourly email attachments, and third-party APIs. Metrics include power, temperature, water, indoor air quality, and occupancy, which are available in each VA building report. SAS® LASR™ (LSR) Server tables are updated and are subsequently available in VA building reports. A python middleware solution pushes data in LSR tables at regular and irregular intervals. Included in each building report is a custom alert system designed for the Clemson campus. Each dashboard reports alerts for building sensors that include issues varying from failure to report, to a minimum and maximum value check. Current models include carbon footprint analyses and a 24-hour deep learning neural network forecast of total power use for a campus building.

INTRODUCTION

Clemson University is a land-grant institution located in Clemson, South Carolina, USA, with a combined undergraduate and graduate enrollment of approximately 25,000 students. SAS® and Clemson University entered into a collaboration in the spring of 2017. A comprehensive SAS® software suite supports education and research opportunities based out of the Watt Family Innovation Center located on the main campus. Clemson University has identified an ambitious Net-Zero goal of becoming carbon neutral by 2030. This goal will require reducing the campus carbon dioxide (CO_2) emissions through the use of renewables and reduced energy use. To realize this ambitious goal, the University will need to transform across all major sectors, including power usage and generation, technology investments, building construction and operations, and transportation. Building operations and construction account for approximately 40% of annual global carbon emissions (GlobalABC, 2018). Of that percentage, approximately 28% is attributed to building operations, which includes all utilities such as heating, cooling, receptacles, and lighting. A university campus can be thought of as a compact town or city, thus building operations likely account for an even higher percentage. The overarching goal of the Clemson Energy Visualization and Analytics Center (CEVAC) is to support campus sustainability objectives and identify opportunities for improved efficiencies at a building and campus level.

CEVAC is working towards a more sustainable campus and supporting a healthy working environment through the use of analytics and modeling of campus utility data. Leveraging access to multiple data stores and web-based data streams, the CEVAC team is deploying advanced visualization and analytic solutions. CEVAC is developing data sets and products for ten buildings across campus. The Watt Center and Clemson University Facilities entered into a collaboration to form CEVAC that allowed for the hiring of student interns during the Spring of 2019. The CEVAC student team has since developed data processing tools, pipelines, and databases that enable utility views for each building using SAS® Visual Analytics™ 9.4 (VA) for streaming and historical datasets. Building metrics include power, temperature, water, indoor air quality, and occupancy that are available through tabs for each VA dashboard. At a campus level, the CEVAC team is monitoring campus steam and chiller plants. In addition to building metrics, each VA dashboard reports notifications for building sensors that include issues varying from failure to report, to a minimum and maximum value check. Our vision for CEVAC is to build an educational, research, and operations platform that will change how a community can use building and campus utility data for domains varying from computer science, analytics, and supporting sustainability goals.

Clemson University owns and maintains the utility distribution systems on the main campus, including electrical, domestic water, sewer, natural gas, district steam, and chilled water. The campus mechanical distribution network contains around 41,560 linear feet of steam and condensate lines and 78,520 linear feet of chilled water lines. Clemson Facilities is responsible for approximately 8,000,000 sq.ft. of conditioned space. There are over 120 buildings on the main campus that have an annual energy use of about 167,000 MWh per year that equates to approximately 35,000 households per year. Building and utility systems communicate using the BACnet protocol for Building Automation and Control networks (ISO 16484-6). BACnet enables communication of building automation and control systems for utilities and sensing, such as HVAC, power, and lighting.

THE DATA PLATFORM

Facilities data and information systems have been in operation for many years and primarily served as a two-year database for building utility information with older data rolling off and replaced with the latest observations. These data systems were not designed for real-time and advanced visual analytics and modeling. However, after initiating this project, it became

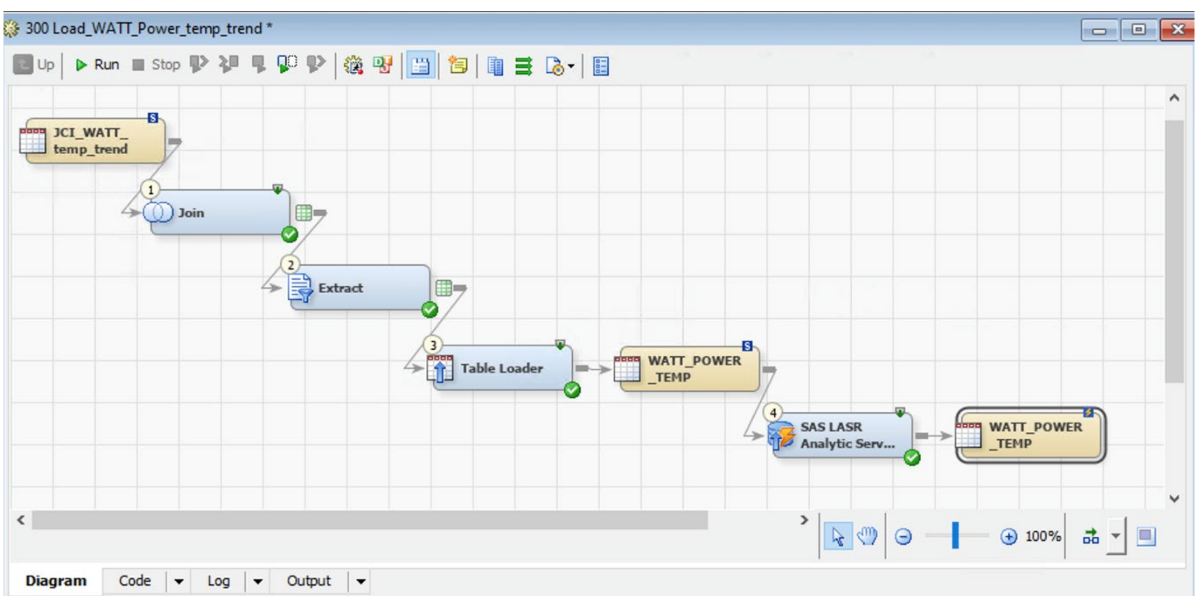

Figure 1. SAS Data Integration Studio Data Pipeline for Room Temperature at the Watt Center

apparent that the 4.6 Billion records for campus buildings could support an in-depth data analytics program. Two confounding issues were the lack of data dictionaries and well-documented data models for the facilities database that impacted the ability to perform correct SQL queries and limited the understanding of data records. Following initial work that spanned over a year, Watt Center staff entered into a collaboration with Clemson Facilities that realized CEVAC in the spring of 2019 and the hiring of six undergraduate students. Students developed several IT components, including data stores and pipelines, visualization, and modeling tools.

The first data platform focused on the Watt Family Innovation Center. The Watt Center was chosen as a test site as it was Clemson's newest building at the time (2017), and the authors were familiar with the data and sensing systems. Initial access was provided to the primary facilities data store using Standard Query Language (SQL). A SAS® ODBC library was created in SAS® Management Console™, exposing the database to other SAS® technologies.

The first data pipelines used SAS® Data Integration Studio™ (DIS) for near real-time data processing leveraging the SAS® LASR™ (LSR) Server. The primary objective was to build a functional data platform that regularly updated LSR tables, enabling the development of VA reports. DIS was an adaptable data exchange, transform, and load tool (ETL). The initial queries and resulting tables were automated to run every hour, resulting in multiple data sets, including building power usage, room temperature, building occupancy, and historical and forecast weather. The flexibility of DIS integrated with LSR and VA reports was pivotal for getting the project started. Afterward, analytic efforts increasingly relied upon the use of SAS® Studio™ for data mining and the development of building occupancy models (PROC REG). Ultimately, DIS data pipelines consisted of a data pull from historical facilities SQL data sources followed by join functions, data cleansing, manipulation, and model runs with the resultant hourly output tables written to disc and LSR (Figure 1). Two table concepts were developed 1) historical or trend that supported daily, weekly, monthly, yearly plots; and 2) live tables that stored the latest observation for any given sensor (Figure 1). The CEVAC data platform has grown more sophisticated since students took over day to day management in the spring of 2019. The platform no longer uses DIS as a management tool for data pipelines. The first major version update resulted in a series of SQL scripts that wrote CSV files to the SAS AUTOLOAD™ (AUTO) folder. AUTO scripts ran every 15 minutes resulting in updated LSR tables. The use of AUTO lead to a

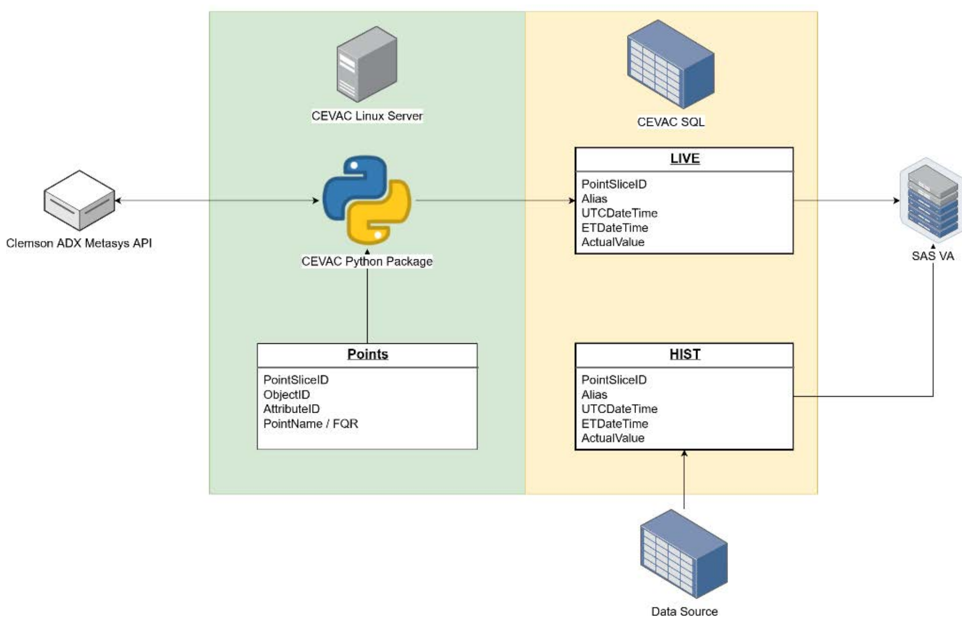

Figure 2. Latest CEVAC Data Pipeline Leveraging SASPy and CEVAC Python Package

more manageable data pipeline, given the large number of tables. This updated architecture was quickly implemented; however, two issues limited system performance 1) all data tables were uploaded at fifteen-minute intervals, it was an all or none solution, and 2) increasing the frequency of AUTO resulted in a degradation of server performance; thus there was little room for growth beyond the current data pipelines.

Version 3 of the CEVAC data platform uses the concept of pipes whereby data streams are a collection of objects defined in a custom CEVAC Python package. Each data stream is managed by a Pipe object, which balances several caches, including in-memory datasets on the LSR server (Figure 2). Transactions are efficient, and in a matter of seconds, the latest observations from the BACnet network are ingested, aggregated, and append to the LSR datasets supporting VA reports. For example, the following code appends the contents of a Pipe object to its corresponding LSR historical dataset (dest):

```
data {lasrlibref}.{dest_int};
  set {sqllibref}.{src};
   run;

 proc imstat batch;
  table {lasrlibref}.{dest};
   set {dest_int} / drop;
  run;
```

In summary, by leveraging the open-source SASPy™ module, data pipelines enable efficient ingestion, transformation, and visualization of CEVAC data in VA.

The current CEVAC platform supports a flexible and dynamic processing environment. Multiple data services enable access to SQL databases, cloud services, wifi devices, and even email attachments (Figure 3). Using the Watt Center as an example, each room has a wired thermostat, and many rooms have carbon dioxide (CO_2) sensors. CO_2 is considered a measure of Indoor Air Quality (IAQ) in combination with temperature and humidity. Clemson facilities has made a concerted effort to add power meters that enable monitoring of power use for lights, outlets, heating, and cooling. These building sensors communicate

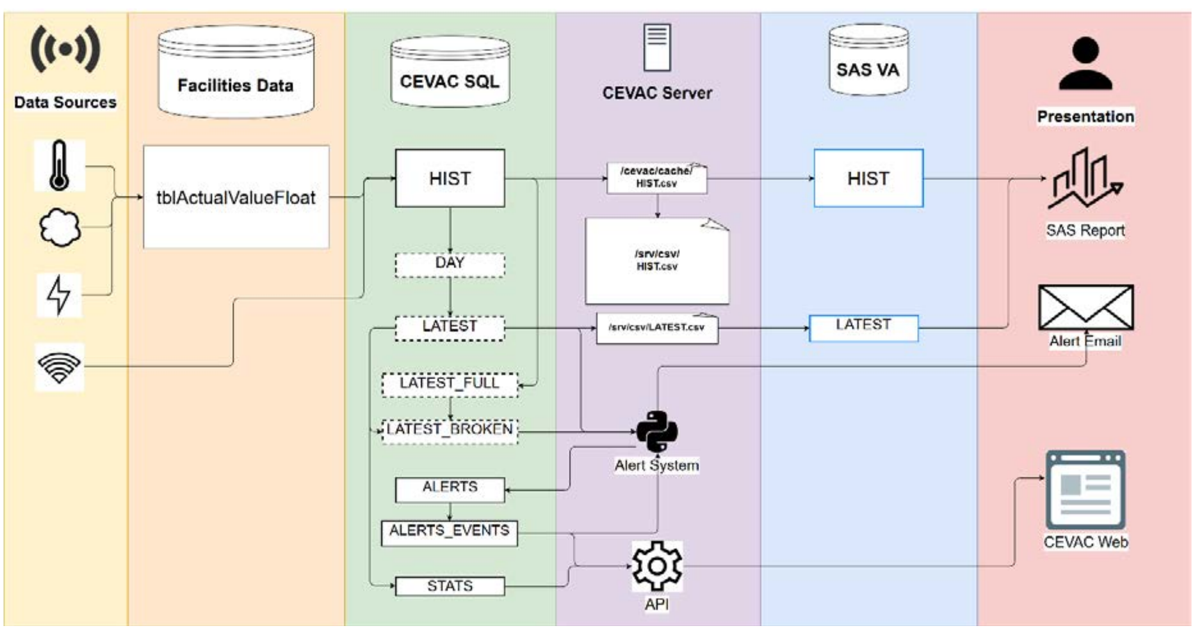

Figure 3. CEVAC Data Platform including Data Sources, Processing, and Presentation

over the BACnet network. Campus infrastructure, including chilled water and steam production, and building occupancy is ingested into CEVAC data systems as comma separated value (CSV) email attachments while a third party Application Interface (API) consumes weather forecast data.

The CEVAC data platform is not only focused on processing and presenting physical measures. It is essential to enable the review of informational data sets that can display the performance of a building in a more holistic manner. The integration of the Clemson Facilities ticketing system is currently in development. Any authenticated Clemson user can submit a ticket to request maintenance or identify an issue such as a cold or hot space. This system is also used by Facilities to schedule regular building maintenance. Although a system exists to submit tickets, there is no system to review and manage a ticket for the non-administrative user. An ODBC connection to the ticketing system enables the ingestion of these data into the CEVAC platform, although, currently, it has read-only functionality. Potential upgrades include allowing the authenticated user to interact with the ticket, such as editing and updating status directly.

Our work with the facilities database found inconsistencies such as missing data, incorrect data, and suspect data. An alert system monitors for discrepancies in a building sensor network. Typical reporting irregularities include missing data, sensors not reporting for extended periods, and suspect readings that are either too low or high. The alert system supports emails and VA reporting where the alerts give detail to the issue, sensor name, and a priority level. The alert system events script is designed to process recent sensor readings and document readings that deviate from the expected value. There are two primary quality control checks for sensors 1) a min and max value check, and 2) a failure to report.

VISUAL ANALYTICS AND MONITORING

Data exploration and monitoring are vital tools for CEVAC. The overarching goal is to develop what is known as a Dive Report (dive) for building security coordinators (BSC). The BSC is responsible for multiple aspects related to day to day management of a building on the Clemson University campus. Each building can have one to several BSCs depending on the size of the building and overall use by the student community. Currently, BSCs do not have a standard tool to monitor multiple aspects of building performance. VA reports are designed to support utility monitoring, data exploration, and ticket management. VA dives

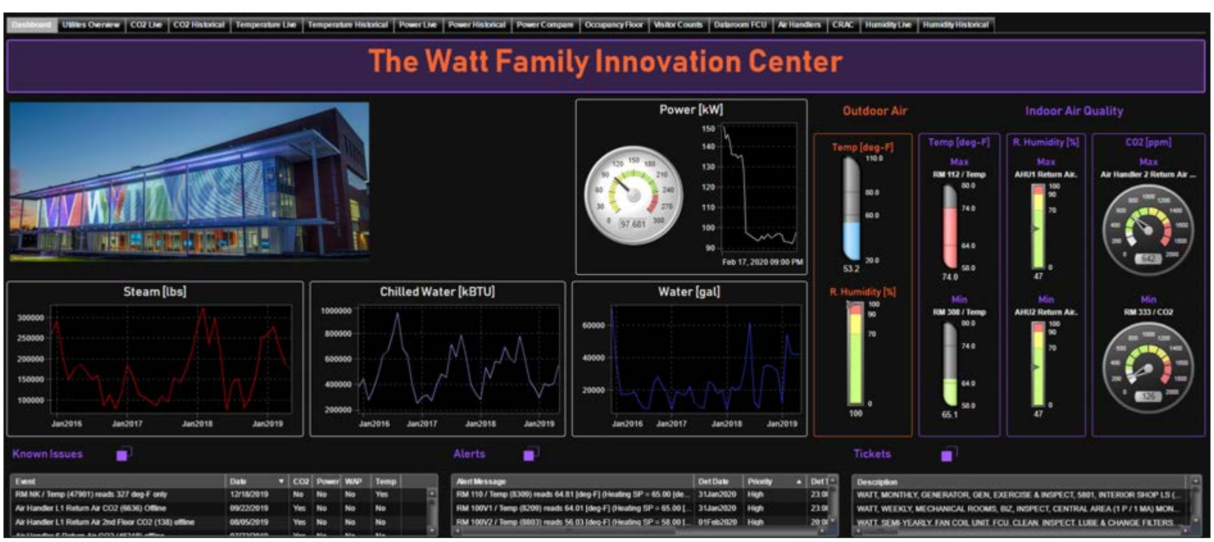

Figure 4. Watt Family Innovation Center SAS Visual Analytics Dive Report

have also resulted in improved building performance through the identification of building systems that were not running optimally.

Dive Reports initially consisted of just a few tabs and parameters. The design was simple, and functionality limited. Dives have advanced substantially in design and function due to 1) the CEVAC team learning more of the features of VA, 2) continued advancement of the data platform supporting more parameters, and 3) increasingly more real-time data, allowing for the use of the dials and other visualization elements. The Watt Center dive is the most data-rich of all ten buildings due to the number of sensors that are currently present in the building (Figure 4). Watt Dive has 16 tabs with historical and live data for temperature, CO_2, power, humidity, 30 minute-occupancy by floor, daily visitor counts, and air handler output. Overall, the first tab of a building dive report presents historical trends, including chilled water, steam, and power, and the latest observations of minimum and maximum room temperature, humidity, and CO_2. This design was the result of developing a presentation that enables a BSC to view the status of a building quickly. At the bottom of the dive, tabular data report building tickets and alert system notifications (Figure 4).

The CEVAC team is exploring the use of virtual reality (VR) technologies as a solution for visualization and remote access to an operational data center. The VA dives are the primary tool for visualization and data exploration, but the role of VR is an exciting aspect of this work. The Watt Center has a dedicated student space for VR use and exploration, which formed a collaboration with the company 3Data that builds AR/VR operation centers. CEVAC has worked with 3Data to develop real-time APIs that allow for the visualization of time series data in a themed CEVAC operations center using VA screenshots (Figure 5).

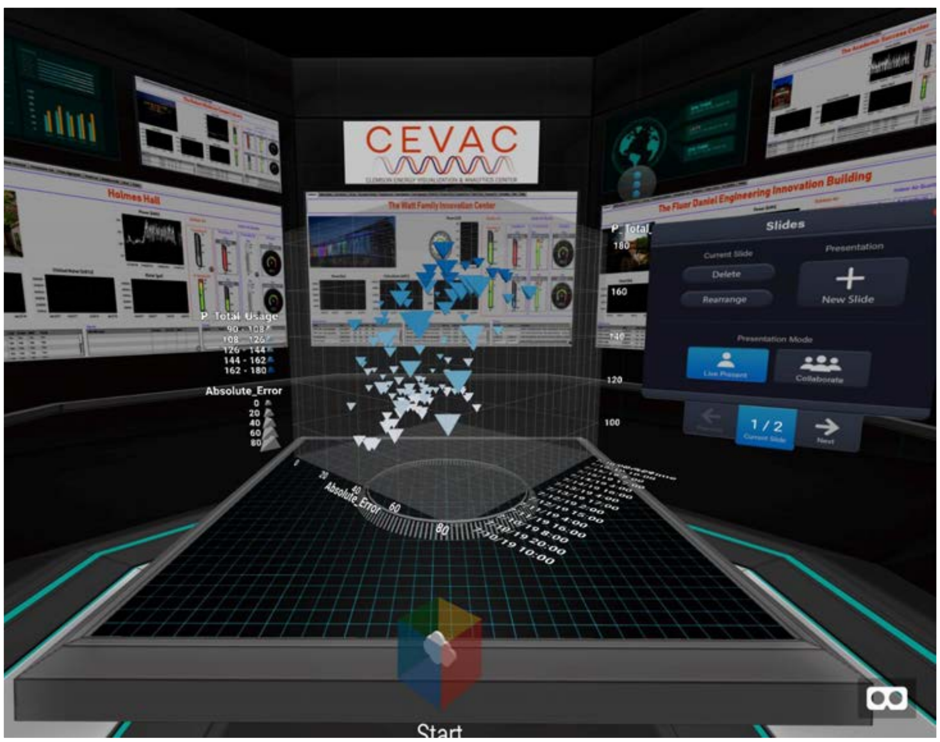

Figure 5. CEVAC Virtual Reality Operations Center Displaying Watt Center HVAC Power Forecast Model and SAS® VA Slides in the Background

Additional use of VA is focused on the reporting of carbon footprint analyses for each building (Figure 6). Carbon footprint calculations provide a snapshot of building performance for CO_2 equivalent emissions from energy use (Airaksinen and Matilainen, 2011). CEVAC calculations do not take into consideration the CO_2 equivalent emissions due to construction and the embodied emissions of building materials. Over the lifetime of a building, operational CO_2 emissions are likely to account for a higher percentage, but embodied material emissions are critical for maximizing sustainability efforts in the construction industry (Alwan and Jones, 2014). The calculation of a carbon footprint is not a straightforward process. There are several direct and indirect measures of energy use and generation that result in carbon emissions for a building on the Clemson campus. At this time, the CEVAC Carbon Footprint analysis takes into consideration several emission sources for building operations. The report is an annual measure for each building on campus that is currently under study by CEVAC. Modeling dates back to 2017 when complete records are available. The inputs into the CEVAC carbon footprint model includes 1) chilled water use, 2) total metered electricity use at each building, and 3) steam produced by natural gas at the Clemson power plant. CO_2 equivalent emissions from these processes are based on reported equivalent carbon emissions from similar systems to those deployed at Clemson. It is also assumed that there is no net loss of energy in the generation and distribution of chilled water and steam at the campus level. The preliminary analyses are promising, but further research is required to improve accuracy and reliability. However, overall trends indicate that steam generation is the largest source of carbon emissions on the Clemson campus. Campus buildings are connected to facilities boilers through an extensive underground pipe network. Behind steam generation, is building electricity and chilled water, respectively. A notable exception to this trend is the Lee III building, which has a geothermal system to heat and cool the interior space. A geothermal system uses the earth's energy to heat and cool a refrigerant that is piped into deep wells drilled near the building (Omer, 2008). Lee III is the most energy-efficient building on campus per square foot of conditioned space.

POWER FORECAST

Building operations and construction account for approximately 40% of annual global carbon emissions (Airaksinen and Matilainen, 2011). Of that percentage, building operations account for about 28%, which includes all utilities such as heating, cooling, receptacles, etc. A university campus can be thought of as a compact town or city, thus building operations likely account for an even higher percentage of total emissions for a campus. Reduction of just a few percentage points at each Clemson building will make a measurable impact on the overall campus carbon footprint. The CEVAC team is working towards the development of a power forecast model for every building that enters into the CEVAC portfolio. The process of building a power forecast model in and of itself will not immediately realize improved efficiencies. It is an exploratory process that can assist in the identification of 1)

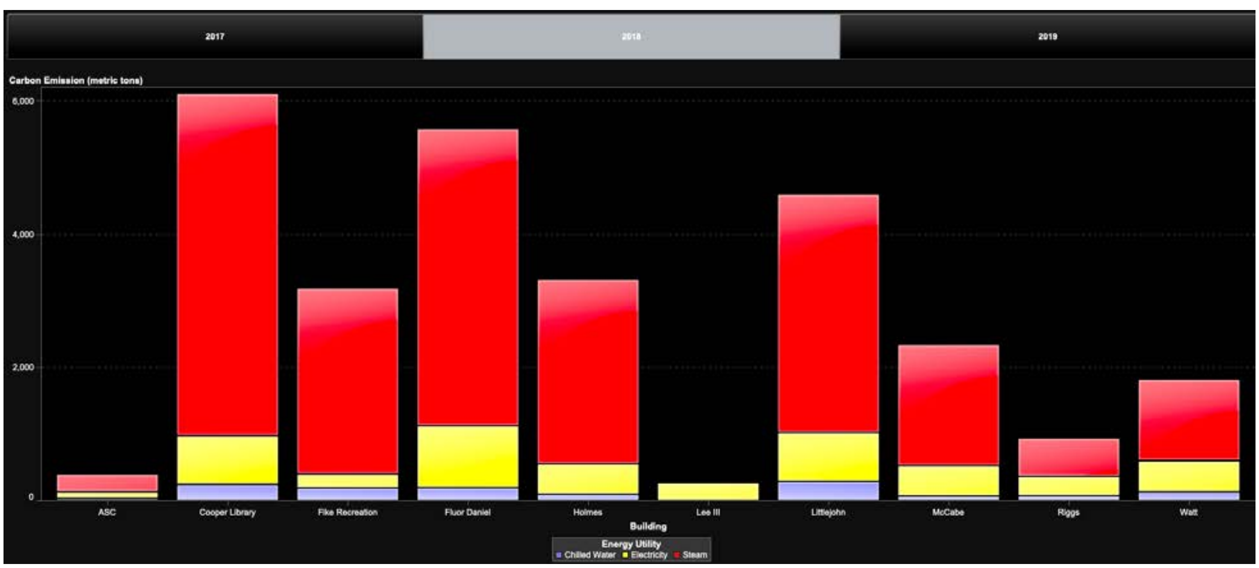

Figure 6. Carbon Footprint for Ten Buildings on the Clemson University Campus. Units are CO_2 Emissions in Metric Tons; Legend: Red = Steam, Yellow = Electricity, Blue = Chilled Water

enhanced building efficiencies, 2) lead to an increased understanding of how different buildings and HVAC systems perform when compared, and 3) detect and diagnose building system faults (Wang and Srinivasan 2017, Zhong et al. 2019). Additionally, a long-range forecast of six months could allow Clemson Facilities to better plan for energy needs that are purchased through contracts with local utilities.

A forecast model of building power consumption requires the identification of likely parameters that will explain the most variability of day to day operations; building occupancy and weather were identified as those two primary factors. Initially, building occupancy presented a unique problem because it is difficult to monitor the occupancy of the building with a high degree of accuracy without dedicated monitoring technologies. However, the CEVAC team has made some advancements in this area and will begin integrating occupancy into machine learning models. One obvious factor that became apparent was the impact of weather conditions on total building power consumption. Several meteorological and temporal features support a deep learning neural network. Currently, the model uses the month of the year, percent through the month, the hour of the day, day of the week, cloud coverage percent, temperature, and humidity. Cloud coverage, temperature, humidity are forecast products made available through the Dark Sky API. These products are also available through the National Weather Service (NWS) APIs.

A python Keras Deep Learning Model forecasts a 24-hour power consumption model for the Watt Family Innovation Center. The Hybrid Scaler model has an approximate 80% accuracy and is the best performing of three models (blue/green trend line) as compared to observed data (red trend line)(Figure 7). The scaler model is data normalized, where input variables are transformed between a range of 0 and 1. The data set consisted of 58,731 recorded power observations at thirty-minute intervals for training and validation. Twenty percent of the records were set aside for training. While this is a rich data set, this only represented approximately 40 months' worth of data at thirty-minute intervals.

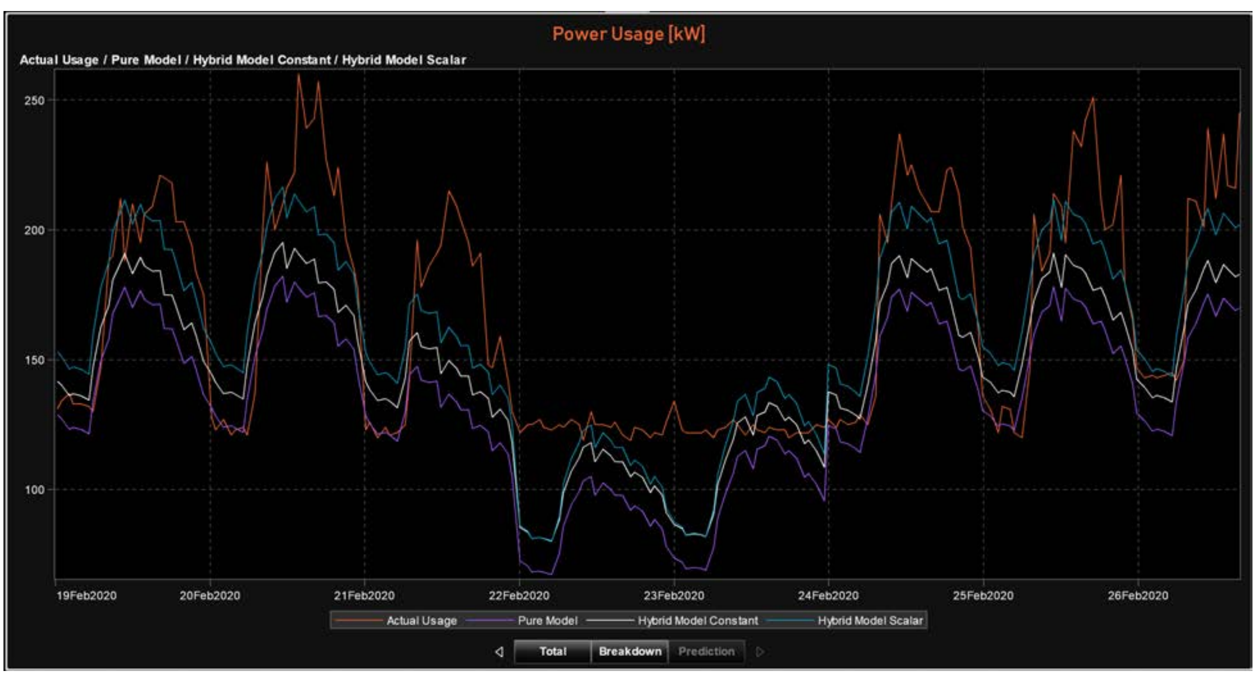

Figure 7. Watt Center Power Forecast Model for February 19, 2020 to February 26, 2020; Legend: Red = Observed, Purple = NN, White = Hybrid Constant, Blue/Green = Hybrid Scaler

A model run of eight days dating from February 19, 2020, to February 26, 2020, is presented (Figure 7). Observed data show a daily pattern with peak power consumption occurring in the early afternoon regularly except for the weekend as a result of the building being in unoccupied mode (February 22–February 24) (red trend line). The Watt Center has a large atrium that rises to the third floor of the building. The afternoon drops and spikes in power usage are hypothesized to be the effects of clouds and daily sun angle. Overall, the building is in occupied mode from 6 am to midnight during the weekdays. That trend is visible in the data and captured by all models. However, all three models tend to underestimate power consumption peaks, and none predicted the unoccupied mode correctly.

CONCLUSION

CEVAC made impressive strides in under a year. The student interns continue to advance the project and are gaining valuable work experience. The team is excited by the possibilities and collaborations that have evolved. SAS® Visual Analytics™ is a critical component to this work, and the upcoming upgrade to SAS® Viya™ at Clemson University will provide a more developed suite of tools to continue the CEVAC work, including visual analytics and machine learning.

Power forecast modeling will play a more significant role in CEVAC operations going forward. The past several months focused on the development of a data platform that is reliable, and data are documented and well-understood. Version one of the Watt Center Power Forecast Model is promising but needs work. The current model underestimates peaks and does not capture weekends well when building systems stay in unoccupied mode. Another line of modeling is to include building occupancy into future efforts to adjust occupied and unoccupied modes depending on current and future use. For example, if the Watt Center remains in occupied mode until midnight, but there are just a few occupants should the building move into unoccupied mode earlier. Very few buildings on the Clemson campus are of a similar design and use different construction materials. A CEVAC hypothesis is that each building on campus will have a unique power fingerprint in the sense of how it responds to varying environmental conditions and occupant loads. The CEVAC team will be pursuing these topics and others in the second year of CEVAC operations and research.

REFERENCES

Airaksinen, M. and Matilainen, P., 2011. A carbon footprint of an office building. Energies, 4(8), pp.1197-1210.

Alwan, Z. and Jones, P., 2014. The importance of embodied energy in carbon footprint assessment. Structural survey.

Global Status Report, Towards a zero-emission. (2018). Efficient and resilient buildings and construction sector. Global Alliance for Buildings and Construction (GlobalABC).

International Standards Organization (ISO) 16484-6:2014, Building automation and control systems (BACS) — Part 6: Data communication conformance testing, https://www.iso.org/standard/63754.html.

Omer, A.M., 2008. Ground-source heat pumps systems and applications. Renewable and sustainable energy reviews, 12(2), pp.344-371.

Wang, Z. and Srinivasan, R.S., 2017. A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. Renewable and Sustainable Energy Reviews, 75, pp.796-808.

Zhong, H., Wang, J., Jia, H., Mu, Y. and Lv, S., 2019. Vector field-based support vector regression for building energy consumption prediction. Applied Energy, 242, pp.403-414.

ACKNOWLEDGMENTS

Mark Jenson, Director, Watt Center IT Technologies at Clemson University
Ty Standridge, Watt Center IT Systems Engineer at Clemson University
University Professional Internship / Co-op Program at Clemson University
CEVAC Students: Solomon Brown, Andrew Emery, Harrison Hall, Karr Hersh, Zach Klein, Lingxiao Li, Bennett Mears, Tyler Piel, Zach Smith

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

David L. White
Research Assistant Professor
whitedl@clemson.edu