

EnergyFlowVis: visualizing energy use flows on UBC campus

Claude Demers-Bélanger, and Sanyogita Manu

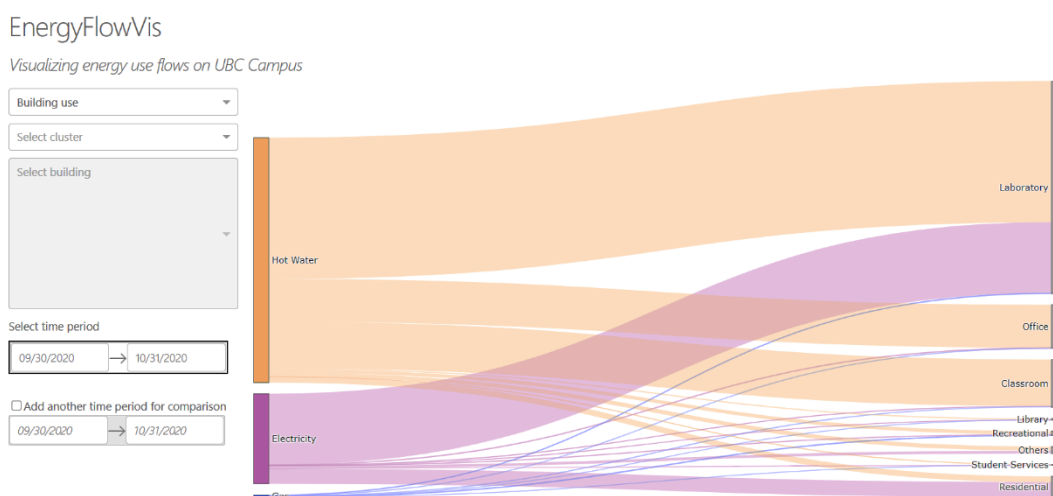


Fig. 1. The EnergyFlowVis interface and visualization

Abstract—This paper presents EnergyFlowVis, an interactive Sankey diagram generator that will help researchers and energy managers to visualise energy use flows by the type of energy source for the buildings on the University of British Columbia Vancouver campus. The application allows the user to compare energy use flows across different building types or clusters by enabling data filtration at multiple levels based on the primary function of the building, its age and size. It also allows the user to view the energy use flows over time by comparing two time periods. The ultimate goal of the application is to help inform decisions for better energy management on campus.

Index Terms—Building energy use, energy flow visualization, Sankey diagrams

1 INTRODUCTION

UBC campus is akin to a small town with over 350 institutional buildings, a hospital, a secondary school and five residential neighborhoods. In 1997, UBC was the first Canadian university to implement a campus-wide sustainability policy. Since then, UBC has developed several strategies to further the cause of reducing GHG emissions, one of them being energy efficiency measures for the built environment. This work falls under the scope of the UBC Energy and Water Services (EWS) Unit, which is responsible for managing and operating the buildings on campus. Since buildings are one of the most important utility assets, they need to be monitored closely to assess performance and implement efficiency programs. To achieve this, building management systems (BMS) have been installed in 152 buildings on campus. BMS monitors and controls the various mechanical heating, ventilation, and air conditioning (HVAC) systems in these buildings using a network of sensors and controllers. Commercial BMS systems come with their own data visualization tools, which have certain limitations. In addition to not being able to store historical data, these tools cannot be customized to add visualization capabilities specific to the tasks identified by the user, in

this case the EWS. So, over a period of time, EWS has developed its own custom applications by consolidating all the BMS data on the Skyspark platform.

Currently, EWS uses two different ways to visualize the energy use dataset [1]. One of them shows a bubble chart spread on the Campus map where the bubble position encodes the location of the building, the size encodes the amount of energy consumed by the building and the color encodes the type of building (Fig. 2). Users can view the total and daily (time series) energy consumption for each building by clicking on the corresponding bubble. The objective of this dashboard is to help the users identify the energy consumption for all buildings in the dataset, together and individually. The other dashboard is a set of bar charts that show historical annual data of the whole campus for GHG emissions, annual usage, annual cost, water usage and savings (Fig. 3). They both have their limitations and haven't been adopted by the research community. We, therefore, identified the need for an application that would offer a dynamic overview level understanding of energy use patterns by energy source for the campus and allow comparison between buildings. This gap, therefore, became our primary motivation for this project.

We both have several years of experience working with such data in the research (Sanyogita) and consultancy (Claude) realms, so this project engages us at a personal level. In addition to these motivations, we are a part of the Energy, Technology, and Architecture (ETA) Lab team. ETA Lab is an interdisciplinary research group that tackles research problems related to energy use in the built environment. This project, therefore, aligns perfectly with the Lab's research activities.

- *Claude Demers-Bélanger is with the University of British Columbia.
E-mail: claudio.d.belanger@gmail.com*
- *Sanyogita Manu is with the University of British Columbia.
E-mail: Sanyogita.manu@ubc.ca*

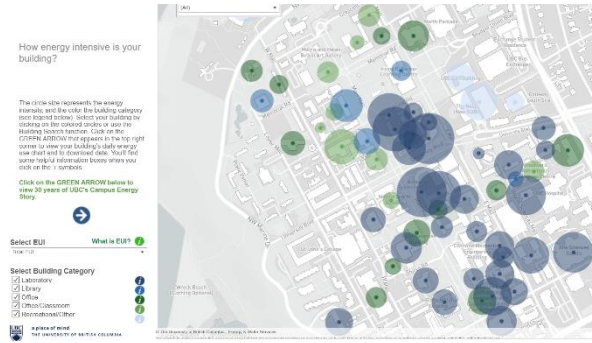


Fig. 2. Campus bubble chart of annual energy use

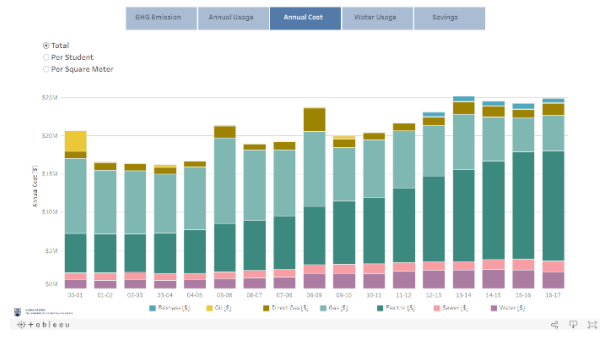


Fig. 3. Bar chart of annual energy cost for the whole campus

2 DATA DESCRIPTION

The data we have used for this project is collected, collated, and managed by EWS. This data is made available to students, researchers and UBC staff through the SkySpark platform. SkySpark’s proprietary data format does not allow for easy third-party application integration so we converted the data into InfluxDB [2]. Thereafter, we performed considerable pre-processing on the raw data to reshape it into a format that was pliant for the analysis we wanted to perform.

2.1 Pre-processing

In its most raw form, the data after conversion may be presented through three separate datasheets in a tabular format. We have explained the structure and the data available in each datasheet below.

The first sheet has the sensors’ metadata where each row is an installed sensor, and the columns are the attributes or “tags” assigned to the sensor. The examples of attributes are “bmsName” which is the name of the BMS system to which the sensor belongs; “siteRef” for the name of the building where the sensor is located; “navName”, which shows the specific variable being measured by the sensor, such as the “Discharge air pressure differential” or “Zone temp”. The raw data file has 66,904 sensors (and as many rows) and 360 columns for attributes associated with the sensors. We found that only 43,706 sensors were labeled, and this sub-set became our starting point. We have mentioned earlier that these sensors monitor a wide range of variables. Since our scope is this project is limited to energy data, we were interested in attributes called “gas”, “elec”, and “hotWater”, all representing the energy sources on campus. So, we filtered the sensors by these tags. Next, we performed additional filtering to keep sensors that were associated with the main meter, i.e., sensors that were monitoring building level energy use as opposed to those used for sub-metering zones within a building. This was done using “Main Meter” as the string search with the “equipRef” attribute. We found 219 main energy meters through this search.

The second table (or a set of tables) we used during the pre-processing stage was the actual metering data containing the time series monitored data for a meter type. For example, data from the “gas” meters was compiled in one table. So, there were three tables for the three meter types: electricity, hot water and gas. Each row in the table was a meter reading with a time stamp index and each column was a building. Several buildings have more than one meter types if they use energy from more than one sources. So, the number of columns (or buildings) varies between the three tables. The meters record data at 15-minute intervals. We aggregated the meter data to daily time series for our purpose.

The last table that we used at this stage was the building metadata sheet. The building level attributes were compiled into this sheet, where each row was a building and the columns had the following attributes: building code, building name, building area, year of construction, floor area allocated (in percentage) for these activities:

classroom, laboratory, library and office, and the building type (based on activity) assigned by EWS. The last column represents the categorization of the buildings by principal activity type and the categories were: classroom, laboratory, library, office, others, recreational and recreational_others. We used this table to identify building groups or clusters based on three criteria:

- Building use: We were not able to identify the basis for the categorization used by EWS. We also found several buildings without any category labels and a few that seemed to be labelled incorrectly. Furthermore, we felt the labels were not clear or distinct enough to use in the visualisation. So, we used the floor area allocation to determine the principal activity type – the activity with maximum area allocation was considered as the principal activity for that building. We assigned a new “building use” label to each building from a list of 8 labels. For buildings where the floor area allocation data was missing, we identified the primary activity type using information available about those specific buildings online.
- Year of construction: We used this data to identify clusters of buildings with similar age. We used the same cluster definitions from the CBECS dataset [3].
- Floor area (square meters): we used this data to identify clusters of buildings of similar size. We used the cluster definitions from the CBECS dataset [3] as a starting point and then performed several iterations on the cluster ranges to arrive at one that was well suited for our dataset.

2.2 Data abstraction

The dataset we are using is static even though energy monitoring is a continuous activity. The primary nature of the data type is a time series table containing energy use values corresponding to each time stamp. It is important to note here that the energy monitoring on UBC campus was implemented in stages. This means the buildings were included in the monitoring plan progressively. The earliest monitoring data is available starting 2010. We will now describe the data abstraction based on the idioms used in Visualization Analysis and Design.

2.2.1 Quantitative attribute: Total energy in kWh

The total energy use was monitored and recorded for each building through an energy meter. A building may have one energy meter monitoring energy use from a single energy source, or it may have up to three energy meters (for three energy courses), therefore three energy use values for a day, one from each meter.

2.2.2 Quantitative attribute: Energy Use Index (EUI) in kWh/m²

This attribute was derived from the total energy by normalizing it by the floor area. So, for each value of total energy, there is a corresponding EUI value.

2.2.3 Ordered key attribute: Date

Energy meters monitored and logged energy use at 15-minute intervals. Except for very specific tasks such as fault detection or identification of peak load occurrences that may require such precision, a time scale that is daily, or a higher-level aggregate (weekly, monthly, etc.) is usually more useful to analyse energy use data. So, we aggregated energy use into bins of one day. This also allows us to aggregate the data for any number of days within the monitoring period of a building. While the monitoring period varies between individual buildings, for the overall dataset, the date ranges from Dec 31, 2010 to November 20, 2020.

2.2.4 Categorical key attribute: Energy source

There are three types of energy sources from which energy is delivered to the buildings in the dataset. These are electricity, hot water and gas. As we have mentioned before, each building has at least one energy source and several have more than one.

2.2.5 Categorical key attribute: Building name

Each building was considered a separate category and there were 131 buildings in the dataset.

2.2.6 Categorical attribute: Building use

Building use refers to the primary activity associated with a building. This was identified for each building based on the floor space allocation for the four main activity types in the raw building metadata: classroom, laboratory, library, and office. We used 8 categories for building use: laboratory, classroom, office, residential, recreational, library, student services and others (Table 1).

2.2.7 Categorical attribute: Year of construction

Year of construction refers to the year in which a building was constructed. We used this attribute to derive the second cluster type indicative of the age of the building by binning the data into 9 categories (Table 1)

2.2.8 Categorical attribute: Floor area (m²)

Floor area refers to the sum of the area of all floors of a building expressed in square meters. We used this attribute to derive the EUI for a building. We also used it to derive a second cluster type indicative of building size by binning the data into 9 categories or clusters (Table 1).

2.2.9 Data type: Multi-dimensional table

Based on the discussion on our data's attributes, we conclude that our data forms a multidimensional table. The two quantitative attributes of total energy and EUI are indexed using the key attributes of date, energy source and building name. The remaining categorical attributes of building use, year of construction and floor area are used to aggregate the quantitative data to show cluster level data.

3 TASK DESCRIPTION

The domain-level tasks were articulated based on our initial discussions with multiple stakeholders. We met with the researchers from the ETA Lab as well as the UBC Community Energy Coordinator, to ask what information they would like to extract from the EWS sensor data to make better decisions. In addition to questioning the user-friendliness of the current vis solutions, their primary concern was to track the energy used by buildings on campus in terms of the energy source and to be able to do this for any time period. We articulated this requirement in terms of specific tasks, both

in the domain language as well as the abstraction framework used in Visualization Analysis and Design.

Table 1. Clusters within each cluster type and the distribution of buildings across each cluster type

“Building use”	No. of buildings	“Year of construction”	No. of buildings	“Floor area (m ²)”	No. of buildings
Laboratory	41	before 1945	7	<2000	12
Classroom	11	1946 to 1959	10	2000-3000	21
Office	24	1960 to 1969	29	3001-5000	15
Residential	24	1970 to 1979	9	5001-7000	21
Recreational	11	1980 to 1989	7	7001-9000	15
Library	8	1990 to 1999	15	9001-11,000	17
Student services	4	2000 to 2009	17	11,001-15,000	12
Others	7	2010 to 2014	23	15,001-22,000	8
		After 2014	14	>20,000	10

Domain specific tasks:

Task 1: Find out how much energy is used on campus across different energy sources for a specific time period. Compare this data with another time period.

Task 2: Find out how much energy is used across different energy sources for different building use clusters for a specific time period.

Task 3: Find out how much energy is used across different energy sources for different building size (floor area) clusters for a specific time period.

Task 4: Find out how much energy is used across different energy sources for different building age (year of construction) clusters for a specific time period.

Task 5: Find out how much energy is used across different energy sources by a specific building for a specific time period

Task 6: Compare cluster level energy use data for two different time periods.

Task 7: Compare energy use data across clusters for the same time period.

Task 8: Compare building level energy use data for two different time periods.

Task 9: Compare energy use data across different buildings for the same time period.

3.1 Task abstraction

In terms of the visualization framework, at the high level, our application will allow the user to analyse the data by way of discovery, where the user might find new things, such as finding that buildings built between 1960-1969 have lower EUI than those built after 2014. This tool could also be used to present energy use data to the UBC community or in an academic/research context, both in live mode as well as through static visuals. The tool will likely be used by the user to produce output that will be used as input. For instance, the user might initially use the tool to look at energy use data for building use cluster type and might discover that the office cluster uses more energy as compared to the other clusters. As a result, the user may decide to use the tool to look at the office cluster and the individual buildings within that cluster to look at more granular data. We have explained in the data abstraction how new data elements (such as EUI) were produced based on existing data elements (total energy and floor area), which speaks to the *derive* goal, although this transform was done by us. We also consider the three categorical attributes representing the cluster types as new data elements as well.

As a mid-level action, the tool will enable the user to *search* for elements of interest. The primary search type is *explore*, since the user will begin from an overview and then will find outliers or elements of interest as she proceeds. *Lookup* search will be used for cluster types and clusters since they are available at a specific location within the

interface and *locate* will be used to look for a specific building by navigating through the cluster type and cluster.

As a low-level action, user will be able to *query* the data to *identify* the amount of energy used (by source) by a building, for example (Tasks 1-5 are related to this action) or to *compare* energy use across all the buildings within a cluster (Tasks 6-9).

Table 2. What-Why-How analysis of EnergyFlowVis

What: Data	Multiple tables
What: Derived	Multidimensional table: Quantitative attribute: Total energy Quantitative attribute: Energy Use Index Ordered key attribute: Date Categorical key attribute: Energy source Categorical key attribute: Building name Categorical attribute: Building use Categorical attribute: Year of construction Categorical attribute: Floor area
Why: Tasks	Analyze: discover, present Produce: derive Search: explore, locate Query: identify, compare
How: Encode	Area marks for the quantity of energy use Energy sources by color Energy source and end uses (targets) by node position
How: Manipulate	Reorder, realign, animate transitions; navigate between aggregation levels: cluster type → cluster → building Highlighting: energy flows – increasing saturation Reduce items - aggregation Interactive item aggregation to change level of detail
How: Facet	Superimpose: two layers to compare time period; multiple layers to compare clusters Layers are constructed dynamically in response to user selection Superimpose static layers, distinguished by position of nodes

4 RELATED WORK

The range of tools and visualizations used to represent energy use in buildings is so broad and varied it warrants a standalone literature review paper. While an extensive review falls outside the scope of this paper, we present here a brief overview of our understanding of this domain.

Building energy visualization tools fall within two broad categories: tools that visualize energy use (and other metrics) for real/existing buildings using data from BMS, and tools that visualize projected energy use at the design stage based on virtual building energy and information models. We will focus on the first category. We have stated in the preceding section that each BMS has its own commercial toolkit for visualizing the monitored data. There are many BMS vendors in the market. The BMS system at UBC is comprised primarily of control systems from vendors such as Siemens, Delta/ESC and Johnson Controls Inc. (JCI). There are many other commercially available tools as well: Acuity Brands’ BuildingOS, Quality Attributes’ GreenTouchScreen, QA Graphics’ EEED, ProphetSuite by Stromquist & Company and EnergyFlow by Noveda, to name a few. Some of them are meant for non-expert users and present information related to resource use, carbon footprint and other sustainability measures, while others are meant for expert users with advanced knowledge of energy monitoring and management. These visualization tools are also referred to as “energy dashboards”. We were not able to study most of these tools due to lack of information about the visualization capabilities they offer. However, based on our domain knowledge of these tools, we might summarize the “displays” offered by these tools in terms of the vis idioms we see most often. Line/area graphs to show 24-hour load profiles or comparison with

prior load profiles, and bar graphs to show energy consumption by end uses are the most commonly used idioms. Another common way to display minute-by-minute or instantaneous consumption is by using “speedometers” or dials. Lastly, we have also observed an increasing use of heat maps.

We also looked at the related work in terms of solutions that were similar to those we were proposing; our primary objective was to understand how flow data was visualized across domains that were peripherally related to ours. We found that flow diagrams were used extensively to visualize flow of materials [4]–[7], flow of energy from primary source to end-uses or demand drivers from a global [8], [9] and country-level perspective [10], [11] to provincial scale [12], [13]. We also found several examples where flow diagrams were used to examine the interaction between two or more resources, such as the energy-water or energy-water-food nexus [14]–[16]. We came across fewer examples of flow diagrams being used to visualize money and associated goods flows between sectors [17] or energy and exergy flows for sub-systems in industrial processes [18].

The use of flow diagrams in the building energy efficiency domain seems to be limited. Belzer [19] developed separate flow diagrams for commercial and residential buildings sectors to depict flows of energy from source to end-use in these two sectors. We also came across energy flow charts that visualized the flow of energy use related annual metrics (total energy use, CO2 emissions, utility cost) between source and individual buildings on a campus [20]. The same authors proposed two other applications of flow diagrams to show (1) building-level energy flow, and the associated costs [21] and (2) energy and mass flows in HVAC sub-systems [22]. We see another application of the latter in [23]. It is interesting to note that the last two examples are reminiscent of the first ever (arguably) flow diagram made by Captain Henry R. Sankey and published in 1898, “the thermal efficiency of steam-engine” to represent the flow of heat energy in a steam engine.

From our limited study of related works, we found that the Sankey diagram was the preferred idiom for visualizing energy flows. They have been used to encode a similar use case with institutional buildings in [18] as shown in Fig. 4. In another example, the Sankey diagram is used to visualize country-level projections (for 2050) of energy flows for a specific combination of energy efficiency and behavior related scenarios (energy pathways) (Fig. 5). This visualization is a part of a tool that enables users to understand the impact of selecting a given energy pathway on the national carbon footprint [24].

The common thread between all the different versions of the Sankey diagram that we saw was the positioning of the nodes to encode the level of aggregation or the order in which that node appears in the path of the flow. The nodes on the left always represented the first source within the scope of that specific flow diagram and the nodes on the extreme right were the points at which the flow would stop. The width of the flows encoded the value of the quantitative feature represented by the flow, relative to the other flows/values in the diagram. In most examples the flow was one-directional, starting from the source and ending at the target. There were a few, however, where selected flows were cyclical in nature, such as in the case of mass flows in HVAC systems where the air (mass) is recirculated through the system. This cyclical behaviour was not relevant for our task.

The other important encoding we saw was related to the hues assigned to the flows, which may be grouped into three categories: (1) source nodes are encoded with different hues and the flows use the source node’s color; (2) target nodes are encoded with different hues and the flows use the target node’s color; and (3) nodes are encoded with different hues and all flows are the same color which is different from the colors used for the nodes, mostly a variation of grey. If there are too many target nodes in the diagram, assigning as many hues that are distinct could be a challenge if option (2) is used. The diagram in Fig. 5 uses a variation of option (1) where the flows use the color from the immediate source, not the primary source. This means that the colors are assigned to the flows after every node column. Since the

colors are repeated, it become difficult to find a clear rationale for the color encoding when one views the diagram as whole. Our observations from the study of related works have helped us greatly in the design of EnergyFlowVis.

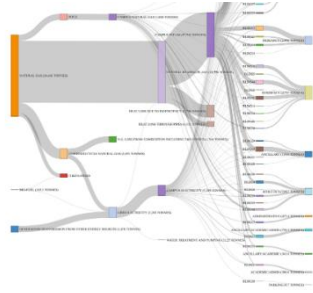


Fig. 4. Sankey diagram showing annual CO2 emissions

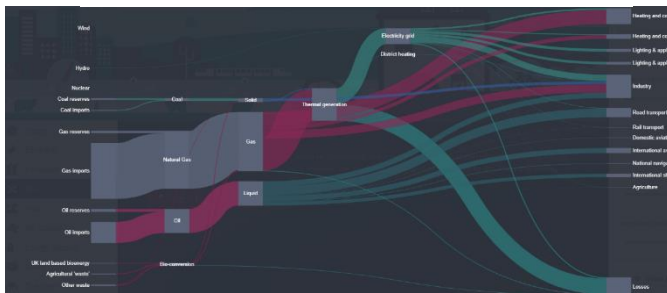


Fig. 5. Country level energy flow chart for 2050 based on a user-selected energy pathway

5 SOLUTION

In this section we describe our visualization solution and analyze it based on Table 2 which uses the What-Why-How framework provided by [25].

EnergyFlowVis is divided into two panels: the **view** panel (Fig. 1b) on the right shows the energy flow diagram, while the **input** panel (Fig. 1a, Fig. 6) on the left has a series of input fields through which the user can control and filter the data to generate the corresponding energy flow diagram.

Fig. 6. EnergyFlowVis user inputs

5.1 Inputs for data filtering

The user can select a cluster type from the first dropdown. The choices are: “Building use”, “Year of construction” and “Floor area (m²)”. Only one cluster type may be selected at a time. By default, the building use cluster type is selected. So, the user is first presented the energy flow diagram for the campus where the buildings are clustered

by building use. This diagram has two sets of nodes, the energy source nodes on the left and the building use cluster nodes on the right.

From the second dropdown list on input panel, the user can select the specific cluster (within a cluster type) on which she may want to focus. This dynamically updates the energy flow diagram which now shows the source nodes on the left, an intermediate node indicating the cluster selected and the end nodes showing the individual buildings within the cluster. At this point, if the total number of buildings within the selected cluster is greater than 20, the user is notified that only 20 buildings that have the highest energy use will be displayed on the diagram. We implemented this limit to ensure all the nodes were visible and there was no occlusion. The user can add more clusters to the diagram, which will increase the number of intermediate nodes. However, the total number of ends nodes will not be greater than 20.

The third dropdown list enables the user to select a specific building with the pre-selected cluster. This selection changes the energy flow diagram back to a two-node level structure, where the nodes on the left are the energy sources and end node shows the selected building. There are no intermediate nodes displayed. The user may add more buildings, which will increase the number of end nodes. Here again, the limit of 20 end nodes is enforced. It is important to mention here that typing the name in the same dropdown panel may also be used as a way to find a building in the list. Buildings already selected by the user do not appear on the dropdown list.

Lastly, in addition to the filtering provided by the aforementioned dropdown lists, the user can select a time period over which to aggregate the energy use data. This is done by selecting the start and end dates on the calendar that pops up when the user clicks on the input field. The default time period is the month of October 2020. A second time period may also be added in the same way. This will introduce additional intermediate nodes in the energy flow diagram corresponding to the two time periods selected by the user.

The highest aggregation level we have used for visualization is the cluster type and the lowest level of aggregation is individual building. This enables changing of the level of detail which is manifested visually in the level of nodes displayed in the energy flow diagram. The tool also allows aggregation at a temporal scale. These aggregation strategies enable filtering of items as well as attributes. The energy flow diagram is reordered, realigned and reshaped using animation in response to the changes made by the user on the input panel.

We initially split these aggregation levels into three separate tabs on the interface, each tab with its own set of input fields. Later, however, we decided that we could offer the same functions with a smaller number of input fields and found the current design to be more intuitive for new users to follow.

5.2 Energy flow diagrams

The view panel displays a single energy flow diagram based on the inputs provided by the user. The quantitative variable of total energy is shown by the flows (area marks) where the width of the flow encodes the quantity and the color encode the primary source of energy. We have used three different hues for the three primary energy sources: orange for hot water, magenta for electricity and indigo for gas. We used the Viz Palette [26] to help us choose three categorical colors for the energy sources. The flows from these energy sources are encoded using the same colors but with reduced transparency to ensure the overlaps between the flows are visible.

The horizontal position of the nodes encodes the level of aggregation, in other words, the categorical attributes. The nodes representing the primary energy sources always appear on the left end of the diagram because they are the highest level of aggregation used in the visualization even though this aggregation cannot be manipulated by the user. Other sets of nodes are added from left to right by the decreasing level of aggregation: cluster type, cluster and individual buildings. An exception to this structure occurs when two time periods are selected, in which case the time periods appear as a separate column of nodes. A maximum of four columns of nodes are

displayed in the diagram at a time; the cluster type nodes are removed once a cluster is selected. All nodes except the energy sources nodes have the same color (light grey) to indicate that they are the target nodes even though they may act as “sources” for the set of nodes to their right.

The energy flow diagram offers three kinds of interaction within the view panel. Hovering over the nodes and the energy flows offers additional information about them. For the nodes, total energy and EUI values are displayed along with the corresponding units. For the energy flows, total energy as well as the source and target node names are displayed. Additionally, the flow is highlighted by increase in opacity which enables that flow to pop out in comparison to the rest of the diagram. Hovering over the nodes highlights all the flows connected to that node. The second level of interaction allows the user to physically move the nodes. This is useful when the distance between two columns of nodes needs to be increased (this happens for some of the building nodes when the name of the building is too long) or when the nodes in the same column need to be rearranged. The third kind of interaction is offered by a set of tools available on the top right corner of the view panel (Fig. 7). These tools allow the user different ways to save the diagram, reset the diagram to its original state (if the nodes were moved by the user), and change the flow data displayed on hover; the two options are to display the data for the flow of interest (the flow that is highlighted) or to display the data for other flows that have the same target node but originate from different energy sources. When a building has more than one sources of energy, and if the contribution from one source is more prominent than from the other source, the energy flow from the second source is displayed as a thin line and may not be easily visible. The second hover option allows the user to compare these energy flows even if they are not visibly distinct in the diagram.



Fig. 7. Default graph controls from Plotly object

6 IMPLEMENTATION

EnergyFlowVis was implemented using the Dash framework from Plotly [27], with all backend programming in Python. The main packages required for data processing were Pandas (Dataframe format) and NumPy (Matrices and Vectors format). The framework builds the application's frame and enables the developer to create functions that will update other elements in real-time. The graphics were made with Plotly objects. The main graph was a Sankey graphical object [28]. It takes a source, a target, and a value per flow displayed. The objects will then automatically sum to display the values at the node level. Therefore, the backend uses three main blocks: the first one queries the database to extract the sum for the inputs (cluster, building, time range); the second block creates the inputs required by the Sankey diagram from this data; and finally, the third block generates and displays the Sankey diagram.

The data is stored in an Influx database. This type of database specializes in the storage of time series data; the resampling queries are optimized compared to using a typical relational database, which allowed us to increase our application's reactivity. The application is hosted in the Compute Canada cloud environment as it is a free resource for researchers to use. For the building metadata, we stored the data in comma-separated text files (CSVs) loaded to memory whenever the user loads the page. We filter for buildings using this data frame and filter the dates at the database level.

Table 3. Project milestones

	Hours				Timeline	
	Planned	Actual	Claude	Sanyogita	Planned	Actual
Pitch	10	10	5	5	Oct 01	Oct 01
Project Meetings	10	14	7	7		
Proposal	14	14	7	7	Oct 23	Oct 23
Design					Nov 25	Dec 05
<i>Related Work</i>	15	15	-	15		
<i>Research Task</i>	5	5	1	4		
<i>Abstraction Data</i>	5	7	1	6		
<i>Abstraction Design</i>	10	10	-	10		
<i>schematic and mock-ups</i>						
<i>Use cases and interaction scenarios</i>	10	10	-	10		
Implementation					Dec 07	Dec 09
Data	10	20	20	-		
Wrangling						
Database build	5	5	5	-		
Data creation	0	5	1	4		
Application	5	5	5	-		
frame programming						
Call back	15	15	15	-		
functions						
Testing	5	5	3	2		
Defect correction	5	5	5	-		
Project Update	10	10	5	5	Nov 17	Nov 17
Peer reviews	2	4	2	2	Nov 19	Nov 19
Final Presentation					Dec 10	Dec 10
Slides	5	6	3	3		
Practice	3	4	2	2		
Recording	2	2	1	1		
Final Paper	20	24	12	12	Dec 14	Dec 14
Total	166	195	100	95		

7 RESULTS

7.1 Use Case 1: Comparing two clusters for the same time period

The user wants to compare the energy use between the oldest and the newest buildings on campus for the month of October 2020. To do this, the user will first select the "Year of construction" from the cluster type dropdown list (Fig. 8a). This will activate the second dropdown list of all the clusters within this cluster type (Fig. 8b). From this list, the user will select clusters named “After 2015, for the most recently constructed buildings, and “before 1945” for the oldest buildings available in the dataset (Fig. 8c). Since the total number of buildings in these clusters is greater than 20, the user will see a pop-up window with a message saying that only the 20 most energy intensive buildings will be displayed on the diagram. Finally, the user will update the time period and select the start date as 10/01/2020 and the end date as 10/31/2020 (Fig. 8d). The resulting diagram shows that the cluster of buildings built constructed after 2015 consumed significantly more energy than the cluster built before 1945 in terms of total energy (Fig. 9). However, when the user hovers over the nodes for each cluster the EUI values are displayed, which indicate that the old buildings have twice the energy use as compared to the new buildings if it is normalized by the building floor area (Fig. 10).

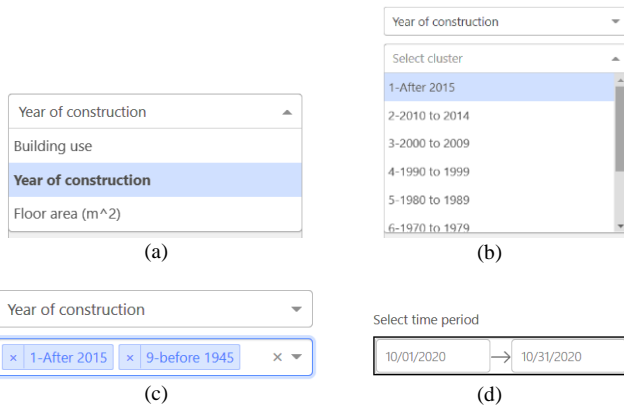


Fig. 8. (a) User selects the "Year of construction" as a cluster type; (b) the list of individual clusters is automatically updated; (c) from the list, the user selects clusters named "After 2015" and "Before 1945"; (d) user selects the time range for the data

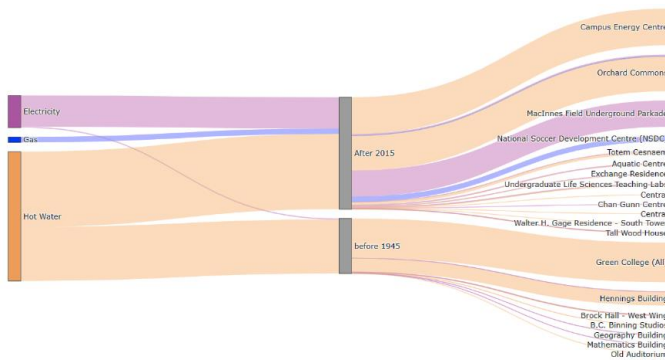


Fig. 9. The energy flow diagram is updated with the input, displaying the total energy consumption for the two clusters

7.2 Use Case 2: Comparing two individual buildings over time

The user wants to know how much energy was used by office buildings on campus in September this year as compared to the same month last year. To do this, the user will change the start and end dates to reflect the month of September 2020 and update the time period selection. The user will check the "Add another time period for comparison" checkbox and update the time period to select September in 2019. Since the

"Building use" cluster type is already selected by default; the user will select the "Office" cluster from the cluster dropdown list. The resulting energy flow diagram shows the office cluster consumed much higher energy in September 2020 as compare to September 2019 (Fig. 11). The most energy intensive office buildings were J.B. Macdonald, Anthropology and Sociology and the Leon and Thea Koerner University Center. In order to view these buildings in more detail, the user will select them from the third dropdown list on the input panel. The updated diagram will show that the difference in energy use between September 2020 and 2019 for those buildings is more pronounced than for all other office buildings combined (Fig. 12). The main difference between the two views stems from the source and target nodes. In Fig. 11, the flows go from energy source → time period → cluster → building. This means that hovering over the flow or an individual building node will not display the energy use for a building for a selected time period. To view that, the flow must look like this energy source → time period → building, as is the case in Fig. 12.

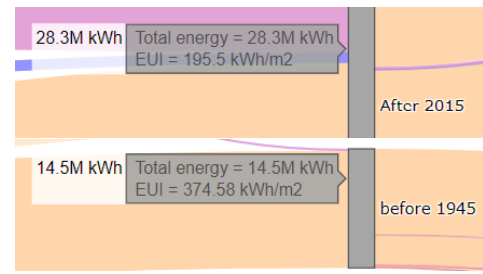


Fig. 10. Hover displays of total energy and EUI for the "After 2015" and "Before 1945" clusters

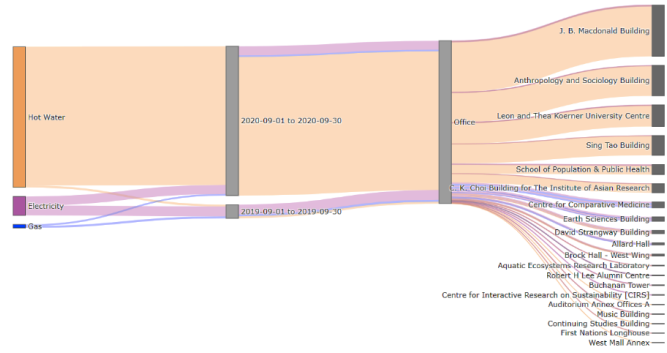


Fig. 11. Energy use for "Office" cluster for September 2020 and September 2019

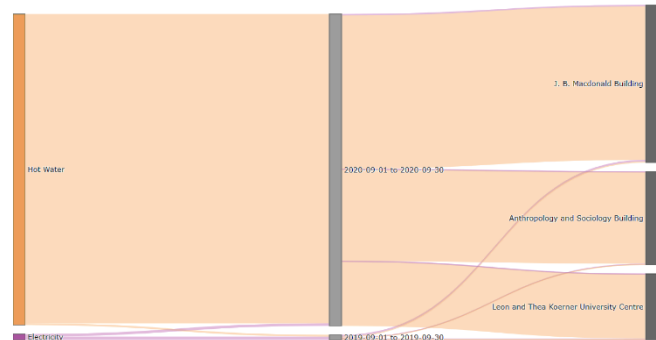


Fig. 12. Energy use for individual buildings within "Office" cluster for September 2020 and September 2019

8 LIMITATIONS

EnergyFlowVis offers an effective way to show the flow of energy on campus. Different levels of aggregation/clustering allow the user to zoom in, metaphorically speaking, on an as-needed basis without being overwhelmed with too much information which would have been the case if, for instance, all individual buildings in the dataset were displayed on the diagram, without the intermediate nodes represented by the clusters. However, both the application and the idiom of Sankey diagram have their limitations. For example, when comparing two time periods for a specific building cluster, the flow between the time period and the individual buildings is "obstructed" by the building cluster, as demonstrated in the second use case above. In general, whenever an intermediate node is added, the flow between the period layer and the end target nodes is interrupted.

Also, since the level of consumption on campus isn't evenly distributed, the visualization is biased towards buildings that use more energy. This means that at an aggregate level, the user will not be able to see buildings that have low energy use.

The application also suffers from some of the shortcomings inherent in the Dash framework. Dash is rigid in the components it uses, which means that while it may be used for prototypes, it is not suitable for implementing a widely used solution. An example of this is the date selection functionality we are currently using, which does not allow the user to change the year directly, without having to scroll through each month. Such weaknesses could become a barrier for broader adoption.

9 FUTURE WORK

This application was a first step toward helping the building science community at UBC to understand the energy use on campus better. It would require a broader user study to improve and validate the results. The team already has a plan to have researchers start using the tool and provide feedback. The idea would be to have a ready-to-use solution working by early 2021.

We have mentioned that we considered the energy dataset to be static for the current version of the tool. In future, the data pre-processing steps mentioned in data description will be adapted to allow a more dynamic visualization of the energy flows.

Since the application can only display high-level consumption at the moment, additional development must show information on both ends of the current energy flow diagram. To go left of the graph, we would need information on how the campus procures and produces its energy (i.e., BC Hydro, geothermal, etc.). While this information is currently available at campus level, there is no data for individual buildings which makes it difficult to integrate it within a solution like EnergyFlowVis. We could derive some of the values; we would need support from the research community to design this integration and show its value. Extending the right side of the energy flow diagram would mean displaying consumption at a lower level than individual buildings, such as building zone (e.g., a room within a building) → energy end use (e.g., lighting, ventilation, heating) → equipment (e.g., AHU, heat pumps). UBC BMS already has sensors that could help with this task. However, there is a lack of consistency across buildings in terms of the variables that are monitored, and the labels assigned to the sensors. Some buildings have an extensive network of power meters monitoring rooms or machines, but others need either additional sensors or calculations to derive energy consumption from the available sensors.

Finally, if the popularity of the tool increases, we may need to update the technology stack. At present, we do not have the means to monitor the performance of the tool and some components need more customization for a better user experience.

10 CONCLUSIONS

In this paper, we presented EnergyFlowVis as a Sankey Diagram generator for energy use flows on UBC campus. It serves as effective way to get an overview of how the energy use is distributed amongst buildings and the sources from which the energy originates. It also allows comparison between groups of buildings as well as user defined time periods. The application is a prototype which will be used as an input for a longer-term solution to help researchers and energy managers make better decisions to improve energy efficiency on campus. The results show great potential and are an obvious improvement over the current visualization options. The next step is to conduct structured user studies and improve the capabilities of the tool to extend the energy flow diagrams to include higher level sources and lower-level end-uses.

REFERENCES

[1] UBC EWS, "Campus Energy Visualization Map," UBC Energy & Water Services. [Online]. Available: <https://energy.ubc.ca/energy-and-water-data/ion-system/>. [Accessed: 22-Oct-2020].

[2] influxdata, "InfluxDB." [Online]. Available: <https://www.influxdata.com/products/influxdb/>. [Accessed: 12-Dec-2020].

[3] U.S. EIA, "Commercial Buildings Energy Consumption Survey (CBECS)," Washington, 2018.

[4] J. M. Cullen, J. M. Allwood, and M. D. Bambach, "Mapping the Global Flow of Steel: From Steelmaking to End-Use Goods," *Environ. Sci. Technol.*, vol. 46, no. 24, pp. 13048–13055, Dec. 2012, doi: 10.1021/es302433p.

[5] J. M. Cullen and J. M. Allwood, "Mapping the Global Flow of Aluminum: From Liquid Aluminum to End-Use Goods," *Environ. Sci. Technol.*, vol. 47, no. 7, pp. 3057–3064, Apr. 2013, doi: 10.1021/es304256s.

[6] P. G. Levi and J. M. Cullen, "Mapping Global Flows of Chemicals: From Fossil Fuel Feedstocks to Chemical Products," *Environ. Sci. Technol.*, vol. 52, no. 4, pp. 1725–1734, Feb. 2018, doi: 10.1021/acs.est.7b04573.

[7] H. Mu, H. Li, M. Zhang, and M. Li, "Analysis of China's carbon dioxide flow for 2008," *Energy Policy*, vol. 54, pp. 320–326, 2013, doi: <https://doi.org/10.1016/j.enpol.2012.11.043>.

[8] J. M. Cullen and J. M. Allwood, "The efficient use of energy: Tracing the global flow of energy from fuel to service," *Energy Policy*, vol. 38, no. 1, pp. 75–81, 2010, doi: <https://doi.org/10.1016/j.enpol.2009.08.054>.

[9] J. M. Allwood, J. M. Cullen, and R. L. Milford, "Options for Achieving a 50% Cut in Industrial Carbon Emissions by 2050," *Environ. Sci. Technol.*, vol. 44, no. 6, pp. 1888–1894, Mar. 2010, doi: 10.1021/es902909k.

[10] L. Ma, J. M. Allwood, J. M. Cullen, and Z. Li, "The use of energy in China: Tracing the flow of energy from primary source to demand drivers," *Energy*, vol. 40, no. 1, pp. 174–188, 2012, doi: <https://doi.org/10.1016/j.energy.2012.02.013>.

[11] C. Chong, W. Ni, L. Ma, P. Liu, and Z. Li, "The Use of Energy in Malaysia: Tracing Energy Flows from Primary Source to End Use," *Energies*, vol. 8, no. 4, 2015, doi: 10.3390/en8042828.

[12] V. Subramanyam, D. Paramshivan, A. Kumar, and M. A. H. Mondal, "Using Sankey diagrams to map energy flow from primary fuel to end use," *Energy Convers. Manag.*, vol. 91, pp. 342–352, 2015, doi: <https://doi.org/10.1016/j.enconman.2014.12.024>.

[13] M. Davis, M. Ahiduzzaman, and A. Kumar, "Mapping Canadian energy flow from primary fuel to end use," *Energy Convers. Manag.*, vol. 156, pp. 178–191, 2018, doi: <https://doi.org/10.1016/j.enconman.2017.11.012>.

[14] P.-C. Chen, V. Alvarado, and S.-C. Hsu, "Water energy nexus in city and hinterlands: Multi-regional physical input-output analysis for Hong Kong and South China," *Appl. Energy*, vol. 225, pp. 986–997, 2018, doi: <https://doi.org/10.1016/j.apenergy.2018.05.083>.

[15] G. Hu, X. Ou, Q. Zhang, and V. J. Karplus, "Analysis on energy–water nexus by Sankey diagram: the case of Beijing," *Desalin. Water Treat.*, vol. 51, no. 19–21, pp. 4183–4193, May 2013, doi: 10.1080/19443994.2013.768038.

[16] D. L. Bijl, P. W. Bogaart, S. C. Dekker, and D. P. van Vuuren, "Unpacking the nexus: Different spatial scales for water, food and energy," *Glob. Environ. Chang. POLICY Dimens.*, vol. 48, pp. 22–31, Jan. 2018, doi: 10.1016/j.gloenvcha.2017.11.005.

[17] W. Leontief, "The Choice of Technology," *Sci. Am.*, vol. 252, no. 6, pp. 37–45, Oct. 1985.

[18] K. Soundararajan, H. K. Ho, and B. Su, "Sankey diagram framework for energy and exergy flows," *Appl. Energy*, vol. 136, pp. 1035–1042, 2014, doi: <https://doi.org/10.1016/j.apenergy.2014.08.070>.

[19] D. B. Belzer, "Energy End-Use Flow Maps for the Buildings Sector," United States, 2006.

[20] A. Abdelalim, W. O'Brien, and Z. Shi, "Visualization of energy and water consumption and GHG emissions: A case study of a Canadian University Campus," *Energy Build.*, vol. 109, pp. 334–352, 2015, doi: <https://doi.org/10.1016/j.enbuild.2015.09.058>.

[21] A. Abdelalim, W. O'Brien, and Z. Shi, "Data visualization and analysis of energy flow on a multi-zone building scale," *Autom. Constr.*, vol. 84, pp. 258–273, 2017, doi: <https://doi.org/10.1016/j.autcon.2017.09.012>.

[22] A. Abdelalim, W. O'Brien, and Z. Shi, "Development of Sankey diagrams to visualize real HVAC performance," *Energy Build.*, vol. 149, pp. 282–297, 2017, doi: <https://doi.org/10.1016/j.enbuild.2017.05.040>.

- [23] L. Perez-Lombard, J. Ortiz, and I. R. Maestre, "The map of energy flow in HVAC systems," *Appl. Energy*, vol. 88, no. 12, pp. 5020–5031, 2011, doi: <https://doi.org/10.1016/j.apenergy.2011.07.003>.
- [24] UK-DECC, "2050 Energy Calculator," UK Department of Energy and Climate Change. [Online]. Available: <http://2050-calculator-tool.decc.gov.uk/#/calculator>. [Accessed: 21-Oct-2020].
- [25] T. Munzner, *Visualization Analysis and Design*, A K Peters. CRC Press, 2014.
- [26] S. Lu and E. Meeks, "Viz Palette." [Online]. Available: <https://projects.susielu.com/viz-palette>.
- [27] Plotly, "Dash." [Online]. Available: <https://plotly.com/dash/>. [Accessed: 12-Dec-2020].
- [28] Plotly, "Sankey Diagram in Python." [Online]. Available: <https://plotly.com/python/sankey-diagram/>. [Accessed: 12-Dec-2020].