Monitoring the Built Environment: Analysis and Visualization of Sensor Data at Winterthur Museum

Elizabeth Keim

27 April 2020

Submitted in partial fulfillment
of the requirements for the degree of
Bachelor of Science in Engineering
Department of Civil and Environmental Engineering
Princeton University

I hereby declare that I am the sole author of this thesis.

I authorize Princeton University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Elizabeth Keim

I further authorize Princeton University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Elizabeth Keim

Abstract

Much research has been done in the separate fields of microclimate environmental monitoring and data visualization for large datasets, but the two have not been combined. This thesis examines how large, heterogeneous monitoring datasets can be synthesized into hierarchical and actionable information using principals of human-machine interface design and machine learning. Winterthur Museum, Garden, and Library, located in Delaware, collects large amounts of microclimate monitoring data to best preserve its collections, but to date has been limited in how it can analyze and visualize that data. This thesis performs data analysis and visualization for the data collected by Winterthur, enabling better conservation at the museum by improving understanding of microclimate control and promoting prompt action to remedy problems. Bounds and Swing Analysis, Cross-Correlation Analysis, and Factor Analysis are used to analyze the data collected by Winterthur's sensors, and visualization on floorplans is explored. The outcome of this thesis is a novel workflow for analyzing and visualizing data from Winterthur, implemented with a Python-based set of software that makes the data interpretable by a diverse audience. This thesis also provides a framework for other monitoring situations, such as structural health monitoring, to utilize for better understanding of the many factors involved in any monitoring situation.

Acknowledgements

I cannot thank my advisors on this paper, Rebecca Napolitano and Professor Branko Glišić, enough. Thank you for your endless guidance, encouragement, and support as I have worked on this thesis. Thank you to Wes Reinhart for your important contributions to this work. Thank you to Joelle Wickens and Maddie Cooper at Winterthur for your insights. Thank you to my friends and family for your constant support and willingness to lend an ear.

This thesis was generously funded by the McIntosh Independent Work/Senior Thesis fund, managed by the Princeton University School of Engineering and Applied Science.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

While we can collect large amounts of data with ease, synthesizing this data into meaningful results is not a trivial task. Furthermore, visualizing these results so that a user can understand them and act promptly off of them is an additional challenge. Winterthur Museum, Garden, and Library, located in Delaware, provides a perfect case study for addressing these challenges. Winterthur collects large amounts of microclimate monitoring data in order to best preserve its collections, but to date has been limited in how it can analyze and visualize that data. This thesis examines how large, heterogeneous monitoring datasets can be synthesized into hierarchical and actionable information using principals of human-machine interface design and machine learning. Much research has been done in the separate fields of microclimate environmental monitoring and data visualization for large datasets, but the two have not been combined, as the literature review in Chapter 2 explains in further detail. This thesis performs data analysis and visualization for the large dataset that the Winterthur microclimate monitoring sensors return, and thus combines the two often separate ideas. The outcome of this thesis is a novel workflow for analyzing and visualizing the data from Winterthur, implemented with a Python-based set of software that makes the data interpretable by a diverse audience. This, in turn, betters conservation work at Winterthur through a better understanding of microclimate control, which enables prompt action to remedy problems and thus best preserve museum artifacts. This thesis also provides a framework for other monitoring situations, such as structural health monitoring, to utilize for better understanding of the many factors involved in any monitoring situation.

1.1 Motivation

Winterthur Museum, Garden, and Library is currently a museum open to the public, and also offers graduate-level programs in conservation. The museum collects temperature and relative humidity data with microclimate monitoring sensors every 15 minutes in 175 rooms. This generates an enormous amount of data to be processed and understood. As many artifacts on display in the museum are vulnerable to any changes in relative humidity or temperature, the sensors are currently used to monitor such changes. Environmental monitoring for climate control in museums is of great importance, especially when those museums are located in historic structures. Winterthur, being both a museum and a historic structure, has a great need for monitoring to understand its microclimate.

There are many factors to consider when controlling a museum microclimate. Dario Camuffo et al. found that both temperature and relative humidity increase in a room with human occupants [1]. Human beings release water vapor when breathing and warm up areas with body heat, thus visitors to Winterthur can affect microclimate. Some of Winterthur's rooms are located on a regular tour route through the museum, but not all of the 175 rooms of the museum are open to the public. Some are used to store Winterthur's many collections, or function as offices or preservation rooms for the many conservators who work there. In addition to considering visitors, a museum must also account for lighting so that visitors can see artifacts. A museum must balance the desire for visitors to be able to clearly see its collections with the fact that strong lighting can damage those items. Sunlight too brings heat and harmful radiation. Finally, the connections between rooms are also important. Some doors between adjacent rooms are always propped open, whereas others are always closed. Some doors are continually opened and closed throughout the day. Each of these states affects the circulation of people, moisture, and heat throughout the structure. Human occupancy, lighting, and connections between spaces are important factors in microclimate control, but understanding how they interact is often difficult.

Elena Lucchi writes that the main points of action for preventative conservation for museums are: “damage preservation and environmental management; architecture and exhibit design; environmental and energy simulations; monitoring, recording and controlling of the environmental agents; management and training” [2]. Fluctuations in temperature and relative humidity, especially rapid jumps, are the leading cause of deterioration of museum items, but even periods of controlled temperature and relative humidity can be damaging if they are not the appropriate value for temperature and relative humidity [3]. Corgnati et al. note that even with HVAC systems to control environmental variables, it is very difficult to completely ensure the exact desired parameter value [3]. G. Pavlogeorgatos also offers a thorough examination of microclimate parameters and their relation to museums, emphasizing that control of the parameters is of utmost importance for preventing degradation of museum objects [4].

For a museum seeking to control environmental factors, being located in a historic structure brings its own set of challenges. Winterthur is a historic building that has been modified several times over its lifetime, so it is more difficult to control microclimate in some areas due to unknown connections between spaces. A further discussion of the many modifications to the building can be found in Chapter 3. Luísa Pereira et al. found during the study of a museum located in an 18th-century structure that long-term monitoring of indoor museum spaces located in historic structures is of the utmost necessity for the conservation of museum objects [5].

Many factors are best understood spatially, as they focus on relationships between rooms. These spatial relationships are difficult to translate to the two-dimensional world that most data visualization tends to occur in, thus a method for visualizing the results of large analyses of data that are best understood spatially is much needed.

A better data analysis and visualization system allows for better conservation of the many important and valuable artifacts on display at Winterthur; the museum's current system makes it difficult to monitor when significant, meaningful changes in microclimate have occurred. With the methods provided in this thesis, problems can be identified in advance, and relationships between various factors can be recognized and adjusted for. For instance, it can be determined why a room fluctuates more than surrounding rooms, such as receiving more sunlight or being more exposed to the wind; these issues can then be remedied with curtains or planting a tree outside the room. Identifying the factors that are at play gives conservators a clear understanding of what actions need to be taken to remedy problems. By improving visualization methods, this thesis permits a better understanding of Winterthur's data, and thus, improves conservation efforts. There is a clear need for the interfaces developed by this thesis for analysis and visualization of microclimate monitoring data at Winterthur.

1.2 Objectives

This thesis develops a novel methodology and software package for synthesizing and visualizing large, heterogeneous datasets for monitoring the built environment. Although it is directed towards Winterthur, this methodology can be adapted on a case-by-case basis to other museums and historic structures. Thus, this methodology can improve conservation abilities of any museum that monitors temperature and relative humidity data, especially those located in old buildings whose microclimates are harder to control. With this ability, important elements of history will not be lost to relative humidity and temperature fluctuations. While this work focuses on a museum environment in a historic structure, the methodology and visualization tools are developed to be flexible so that they are easily transferable to other fields that collect monitoring data, such as structural health monitoring.

This thesis produces four highly intuitive user interfaces that present the results of analyses on large collections of data from Winterthur's microclimate monitoring sensors with excellent visualization. Additionally, this thesis provides methods to perform analyses not common to the field of microclimate monitoring to better understand trends at Winterthur and help conservators account for these. Finally, this methodology has been formalized so that it can be applied to other structures that use microclimate or structural monitoring systems.

The core research objectives are listed here:

- Design an intuitive human-machine interface to translate the results of analyses to a diverse audience.
- Implement bounds and swing analyses to show basic relationships within and between datasets.
- Implement cross-correlation analysis to understand what sensors and rooms are highly correlated with each other.

- Implement factor analysis to highlight trends between monitored variables and other parameters such as external temperature, external relative humidity, solar gains, human presence, and precipitation.
- Formalize the methodology so that it can be applied to other monitoring situations.
- Create methods for visualizing analysis results, focusing especially on ways to highlight the spatial relationships between variables.

1.3 Thesis Organization

Chapter 2 provides a literature review of prior work that has been done in this field. It covers prior work that has been done regarding the analysis of microclimate monitoring data, human-machine interface design, and the visualization of analysis results. Chapter 3 details the history and context of Winterthur and further explains the motivation for this thesis. Chapter 4 highlights the data that is used in this thesis. Chapter 5 presents the methods used in this thesis. Chapter 6 presents and discusses the results of this thesis. Chapter 7 summarizes all that has been presented, offers concluding remarks, and proposes future work.

Chapter 2: Literature Review

While much work has been done in the separate areas of microclimate monitoring sensor data analysis and visualization of big data, the two areas have not been combined. This literature review first examines work that has been done in the analysis of environmental monitoring data and discusses its limitations. Then it turns to prior work in human-machine interface design. Next, prior work in the visualization of microclimate monitoring data is examined. Finally, a summary of prior work is provided that emphasizes why this thesis is novel and needed.

2.1 Analysis of Environmental Data

Monitoring of environmental variables through sensors is done in a variety of contexts, ranging from structural health monitoring to energy usage monitoring to microclimate monitoring. In structural health monitoring settings, sensor data is used to determine loads, stresses, and displacements of structures. For microclimate monitoring, variables such as temperature and relative humidity are recorded by sensors.

A 2015 paper by Fabio Sciurpi et al. examines microclimate monitoring at a museum in Florence, Italy [6]. Sciurpi et al. [6] collected temperature and relative humidity data with sensors every fifteen minutes, just as Winterthur does. They analyzed the data with a bounds analysis, by looking at whether the variable being examined went outside of a recommended range. Sciurpi et al. compare the interior environmental values to exterior weather data to look for relationships between the two [6]. Corgnati et al. also only analyze environmental data from a museum by looking at how well the data stays within the desired range [3]. Other works mirror these studies use of only bounds analyses with environmental data.

As well, past work has focused on analyzing environmental factors in a single room. Sciurpi et al. [6], Corgnati et al. [3], Camuffo et al. [7], and Gysels et al. [8], among others, all focus on analyzing single rooms inside of museums and only perform a bounds analysis on their data. They do not examine data for the museum as a whole. Camuffo et al. [9], focusing on four European museums, does compare multiple rooms within a museum at once, but only at one specific time instance. In general, microclimate monitoring papers focus on bounds analysis (seeing if a parameter stays within a certain range) and not on any other types of analyses.

2.1.1 Factor Analysis in Environmental Monitoring

However, some papers do perform more sophisticated analyses. When dealing with large datasets, exploratory factor analysis (EFA) is often used to determine unobserved underlying variables that relate to observed variables and thus explain variability. There are two common types of EFA: Principal Component Analysis (PCA) and Common Factor Analysis (referred to as "factor analysis" in this work). A UCLA seminar notes that "both methods try to reduce the dimensionality of the dataset down to fewer unobserved variables," but PCA assumes there is no unique variance while factor analysis assumes unique variance [10]. A further technical discussion of the differences between PCA and factor analysis can be found in Section 5.2.3.

PCA is the most frequently used method in prior work in the field of sensor monitoring. It has been especially popular in the social sciences and the field of chemical engineering, such as in Gallagher et al. [11]. In the field microclimate monitoring, Garcia-Diego and Zarzo (2009) used PCA methods based on batch chemical processing methods to analyze microclimate monitoring data from a Spanish cathedral [12]. Garcia-Diego and

Zarzo found that the differences between sensors were explained by two factors determined through PCA [12]. Determining these two factors essentially reduced the number of variables to be worked with. Further work by Zarzo et al. in 2011 continued to use PCA to verify that the relationships between the monitoring sensors had remained the same over a period of three years [13]. Thus, analysis of sensor monitoring data through exploratory factor analysis has focused on using PCA to reduce the number of variables. No research has been done in using (common) factor analysis to determine relationships among the sensor data.

Costello and Osborne suggest that factor analysis ought to be used over PCA [14]. They argue that PCA was mostly used as a method due to limited computer capacity in the past and is not as robust as factor analysis. For these reasons, this thesis will use factor analysis to determine underlying variables affecting measured parameters.

2.2 Human-Machine Interface Design

The question of how to design an intuitive human-machine interface has been addressed by many in various fields. D.A. Norman writes in *The Design of Everyday Things* that discoverability and understanding go hand in hand for creating an effective product [15, p. 3]. The user must be able to discover what is possible for a certain system and understand how to use it. A simple item should not require a manual in Norman's opinion; the user should simply be able to intuit how to go about using it [15].

Interfaces exist for innumerable situations. However, there has been no specific research in interface design for microclimate monitoring. For the works described in Section 2.1 [3], [6-14], the authors ran the analyses and then returned the results to the museum conservators. No interface was created to allow museum conservators to run the analyses themselves, as new data was collected. Even if those studies' methodologies are replicable by other researchers, they do not enable conservators to best utilize data collected. An intuitive user interface brings greater ease and efficiency to the conservators who make use of the results of analyses, while also allowing for analyses to occur over time. This leaves a great opening in the field of microclimate monitoring to be filled.

Data from Winterthur's sensors is currently stored in eClimateNotebook, an online cloud data storage system created by the Image Permanence Institute that also does basic visualization of parameters [16]. This interface is difficult to navigate, especially as it does not indicate when data or charts are loading. It is difficult for the user to know that their click has done anything as no feedback is provided upon user action.

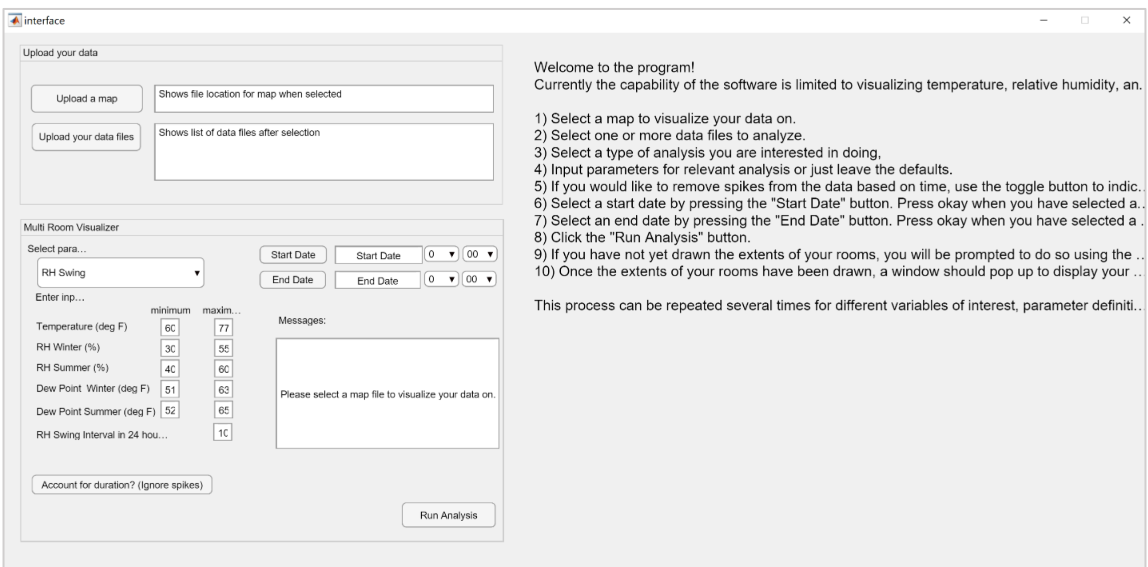

Figure 1: Initial user interface made in MATLAB. King et al. [17].

To remedy this lack of an effective interface for data analysis and visualization, King et al. created an initial user-machine interface for Winterthur in MATLAB [17]. Figure 1 shows the MATLAB interface that was used to analyze and visualize data from Winterthur. The user uploads the floor plan, or map, that corresponds to the data uploaded to the software. If the map has not been previously uploaded, the user outlines and marks the various rooms on the floor plan. The user selects the type of analysis desired, as well as the set of dates to be analyzed. The interface includes a “Messages” box to provide feedback to the user.

Analysis options are limited to bounds and swing analyses, which are also tied to a parameter and season. This limits the ability of a user to customize the analysis to their needs, by forcing them to select from the following analysis options: “RH Swing,” “Temperature Bounds,” “RH Bounds, winter,” “RH Bounds, summer,” “Dew Point Bounds, winter,” and “Dew Point Bounds, summer.” The user is unable to select the parameter, the type of analysis, and the months to analyze (effectively the season) separately, as they are all joined in this interface. As well, the descriptions are confusing, and so not only would customizability permit more control over the analyses run, but it would ensure the user knows exactly how a term like “winter” is being defined.

As well, the MATLAB interface workflow is not clear and could leave untrained users confused, evident by the list of steps that have to be provided at right in Figure 1. King et al. do attempt to enclose sets of steps within grey bounding boxes, but these are faint and difficult to discern against the grey background. Rather than listing all of the steps, a better layout could be used so that any user intuitively understands what must be entered before the final submit button is clicked. There is a need to improve the visual appearance of the interface and provide a more evident workflow, eliminating the need for steps to be listed.

2.3 Visualization of Data

Edward Tufte’s *The Visual Display of Quantitative Information* is an excellent primer in presenting data clearly and without unnecessary design additions [18]. Tufte writes that “of all methods for analyzing and communicating statistical information, well-designed data graphics are usually the simplest and at the same time the most powerful” [18, Introduction]. He continues, “Modern data graphics can do much more than simply substitute for small statistical tables. At their best, graphics are instruments for reasoning about quantitative information” [18, Introduction]. A strong visual representation of results of analyses can be much more telling than any numerical, textual output of the same results. For instance, a time series is much more informative than reading a chart of the maximum and minimum values in that same dataset. Thus, it is important to visualize data.

Many studies in microclimate monitoring do not focus on data visualization at all. For example, Silva et al. (2020) use crude time series and bar charts, and do not visualize their data on a spatial map like a floorplan, even though their work with sensor data would be best understood in that way [19]. Some papers acknowledge that proper data visualization requires extra effort. For instance, a paper by J. Valach et al. examines how to best collect and analyze museum sensor data, but then leaves visualization as future work [20]. The impact of that work would be much greater with proper visualization.

Many papers perform data analysis but do little beyond publishing confusing time series graphs to visualize the data. Boncaina et al. present the results of 20 years of microclimate monitoring and only use time series graphs to do so. They even publish a floorplan within their paper but do not use that to present results; it is there only to help the reader understand the structure being studied [21]. Sciurpi et al. rely on time series graphs and bar charts to demonstrate their results, such as in Figure 2 [6]. Corgnati et al. make use of bar charts to visualize distributions over time, but it is difficult to make quick interpretations from the graphs, one of which is seen in Figure 3 [3].

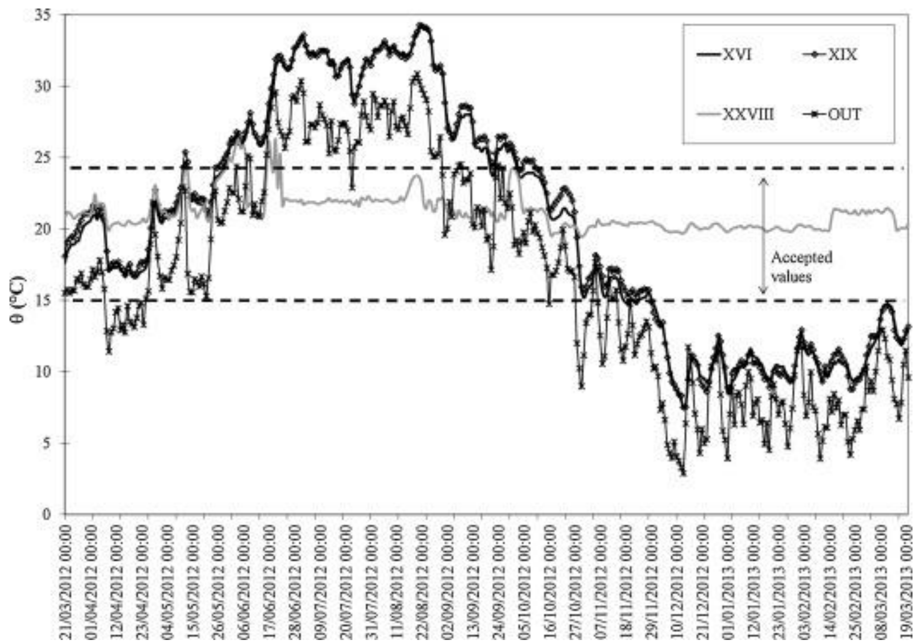

Figure 2: Times-series plot for temperature. Sciurpi et al. [6].

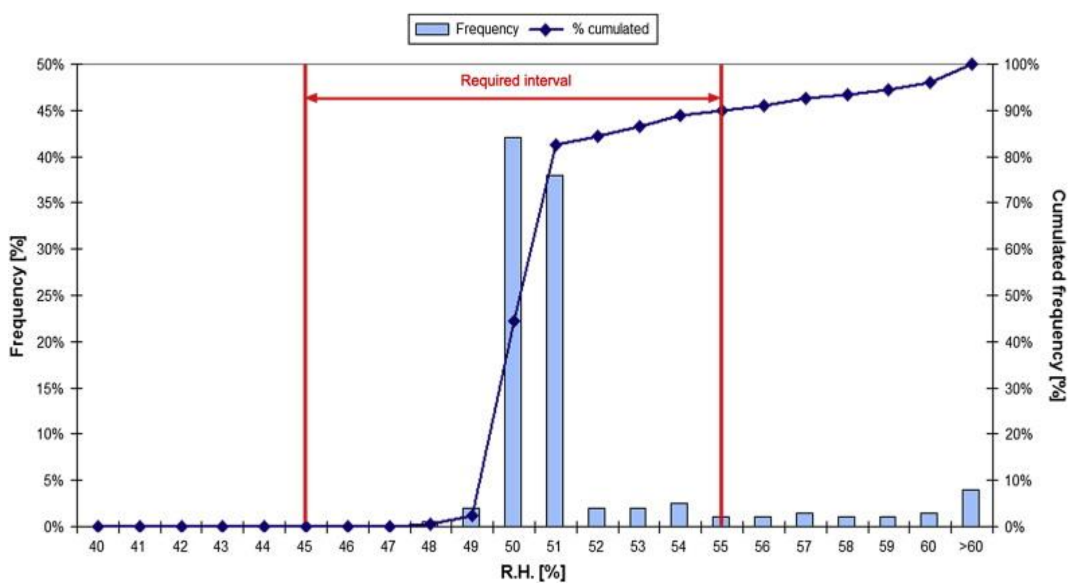

Figure 3: Relative humidity frequency distribution. Corgnati et al. [3].

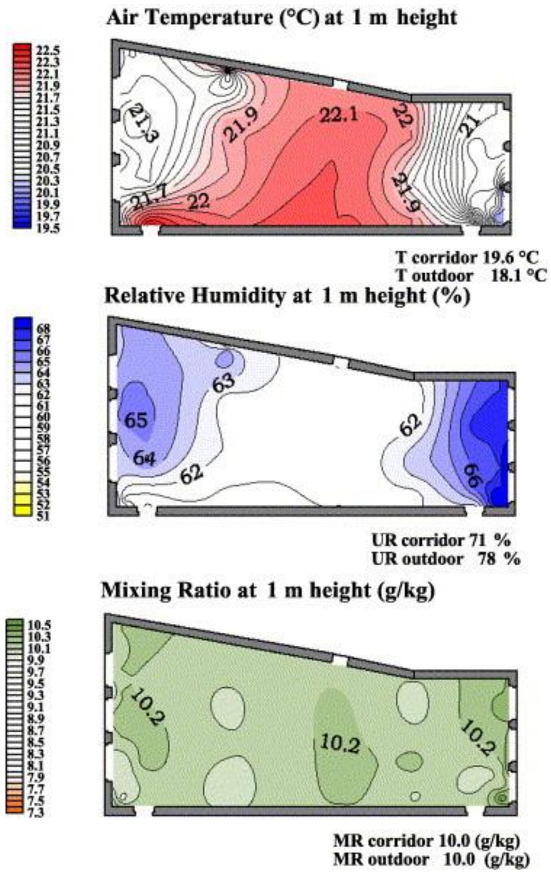

Figure 4: Air temperature, relative humidity and mixing ratio in the hall after the end of a concert. Camuffo et al. [1].

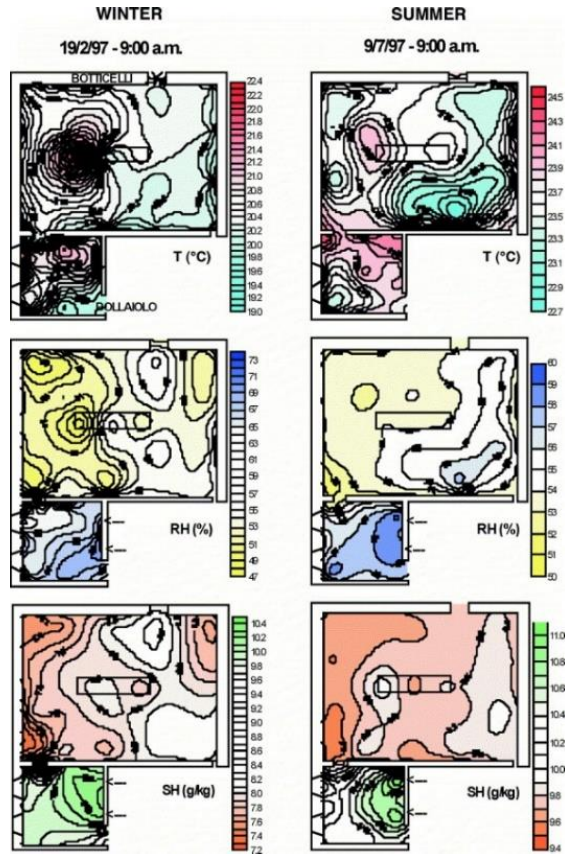

Figure 5: Temperature, Relative Humidity, and Specific Humidity. Camuffo et al. [7].

Others have gone one step further, to visualize the results of analysis on a floorplan. Figure 4 shows an image used in Camuffo et al. to show the environmental climate of a concert hall after a concert [1]. This presents detailed information about one single room, but only for one single instant in time. Camuffo et al. [7] use similar mapping in a study of specific rooms in the Uffizi Gallery in Florence, Italy (Figure 5). Gysels et al. [8] also visualize single rooms in a similar manner, shown in Figure 6. Again, this only maps microclimate variables for one instant of time, and thus is not effective for looking at trends over time. It is important to note that these studies are looking at the distribution of microclimate variables within each room; they are not examining the microclimate variables throughout an entire structure. This method of visualization may be helpful for studying specific rooms, but it is less useful when attempting to visualize the results of many rooms at once. The images are not presenting the results of analyses – merely the results of data collection.

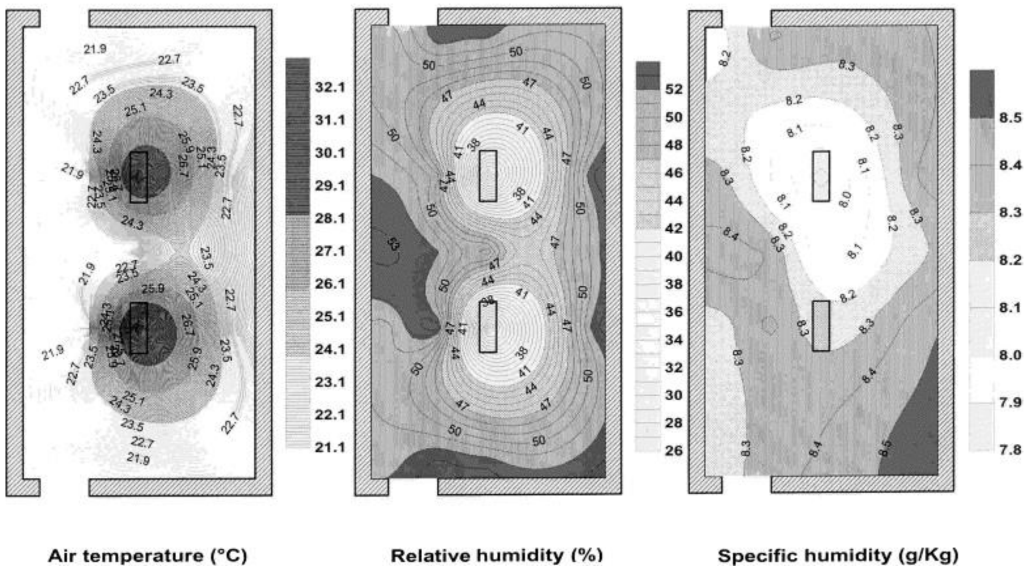

Figure 6: Distributions within one room. Gysels et al. [8].

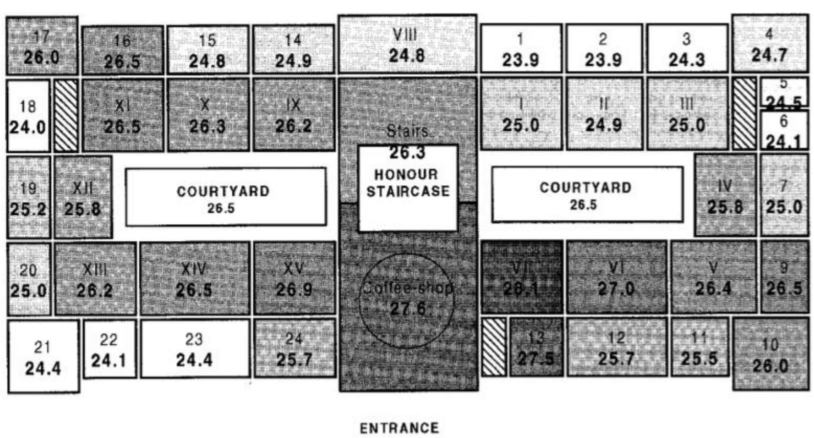

Figure 7: Temperature distribution across a floorplan, August 26, 1997, 4 p.m. Camuffo et al. [9].

Some works do seek to visualize data across an entire floorplan. Camuffo et al. do so in Figure 7 with a temperature distribution of the Kunsthistorisches Museum [9].

However, once again, the data being visualized is from one moment in time (August 26, 1997 at 4 p.m.) and merely shows the data collected – no analysis has been performed.

eClimateNotebook is Winterthur's current data management program [16]. It does not visualize results of analysis, only the parameters over time using time series graphs. Information is often shown in charts, which are not as easy to read as a visualization. For the times series, data can be viewed for up to eight datasets at a time. This method can introduce diagnostic bias because the comparison between two time series is qualitative; the user must look at the time series plots and estimate for themselves how much the two plots differ. This thesis aims to eliminate qualitative biases from data analysis by using quantitative methods for analysis.

King et al. sought to remedy this issue for Winterthur by plotting not just one time-instance of data on a floorplan, but rather the results of analysis [17]. In this way, parameters over time can be plotted on one floorplan. Figure 8 shows the preliminary data visualization work by King et al. [17]. This visualization examines how well Dew Point stays within desired bounds on the 4th floor of the Winterthur Research Building for the time period of October 13, 2018 to March 31, 2019. The opacity of the shading indicates how over or under the desired bounds each room is for dew point. Yellow shading indicates the dew point went both over and below the desired range. Red shading indicates the dew point only went over the bounds, and blue shading indicates the dew point only went below the desired range. Green shading indicates the dew point stayed within the desired range for the entire time range selected. The minimum and maximum dew point for each room are printed on each polygon. Note how the figure uses color, opacity, and text to communicate multiple pieces of information to the viewer. This work is the first step towards examining the results of analyses on floorplans, bringing them into a time-dimension.

Figure 8: Preliminary data visualization on floorplan for Winterthur. King et al. [17].

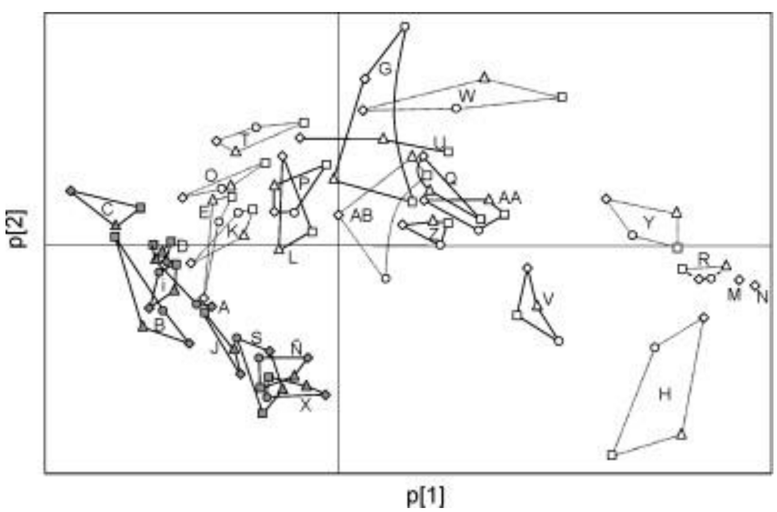

Figure 9: PCA results. Garcia-Diego and Zarzo [12].

Visualizing bounds and swing analyses on a floorplan opens the door to plotting the results of other analyses onto floorplans. No study has sought to graph the results of factor analysis on a floorplan. Garcia-Diego and Zarzo plot PCA results on a confusing graph seen in Figure 9, in which points joined together represent data from one sensor and each axis represents one underlying factor [12]. The viewer does not get a good sense of the spatial connection between sensors, nor of what the underlying factors might in fact be, due to poor labeling of the points on the graph.

2.4 Summary of Literature Review

An examination of prior literature reveals many gaps in the field of analyzing and visualizing microclimate monitoring data. Prior literature has performed simple bounds analyses of microclimate monitoring data and visualized that data with time series analyses. PCA has been used to analyze sensor data, but factor analysis has not been used. The literature focuses on analyzing single rooms, not entire buildings, and only visualizes single time-instances on floorplans. Current user interfaces are not as intuitive as they could be and do not facilitate the best use of the data. This leaves an opening for implementing analyses other than bounds and swing analyses to analyze microclimate monitoring sensor data. As well, visualization can be improved through showing results of various analyses on floorplans, adding a time element. Finally, an interface available to conservators to analyze their data in real-time can be made that will allow for conservators to visualize their data as it is collected, improving conservation efforts.

Chapter 3: Winterthur Museum, Garden, and Library

Located in northern Delaware, Winterthur Museum, Garden, and Library began as a residence for the du Pont family [22]. The first Winterthur house to stand at the site was built in 1837 and had 12 rooms, as seen in Figure 10 [23]. By 1951, when Winterthur was permanently opened to the public as a museum by H. F. du Pont, the house had 150 rooms [22]. The transition from a family home to a museum was driven by H. F. du Pont's love of American antiques and history. Winterthur historians note, "Between 1928 and 1931, H. F. du Pont greatly expanded the house, installing important interiors from early American houses and filling the rooms with his burgeoning collection of American antiques" [22]. Today, Winterthur is a popular site to visit for tourists, with its large collections and pleasant gardens, as seen in Figure 11 [24].

Figure 10: Front elevation of Winterthur, pre-1884 [23].

Figure 11: Current view of Winterthur [24].

3.1 Structural History

As noted before, Winterthur began as a 12-room house and grew over the years to the 175 rooms it has today. Many additions and alterations have been made by its various residents to bring it into its current state. Thus, there are many locations where changes were made that could be impacting the ability to control the microclimate in surrounding areas. For example, an outer wall that becomes an interior wall might have had its windows filled in; on the surface, the wall is just a regular wall, but it is likely to exchange heat with surrounding rooms in an unusual way due to this discrepancy. The museum has published several articles that explain some of the changes that the house has undergone [23], [25]. Jonathan Fairbanks wrote a paper in 1967 detailing the development of the house [26]. The current floor plans for all nine floors of the museum can be viewed on the museum's website [27], and more detailed copies were obtained from Winterthur's conservators [28]. As well, HVAC floorplans have been obtained from Winterthur's conservators. Understanding Winterthur's structure is a complex task which can be facilitated through the analyses of this thesis. The results of this thesis assist in identifying when these hidden structural elements of the museum are affecting microclimate control.

3.2 Location

Local site conditions can affect the ability to control microclimate in a structure. In Figure 12, one can see that Winterthur is located near a small creek and is nestled in a shallow valley [29]. Figure 13 provides another topographic look at Winterthur's location [30]. The effect of tree cover and terrain on sunlight and rain are also important to consider. Understanding site conditions provides

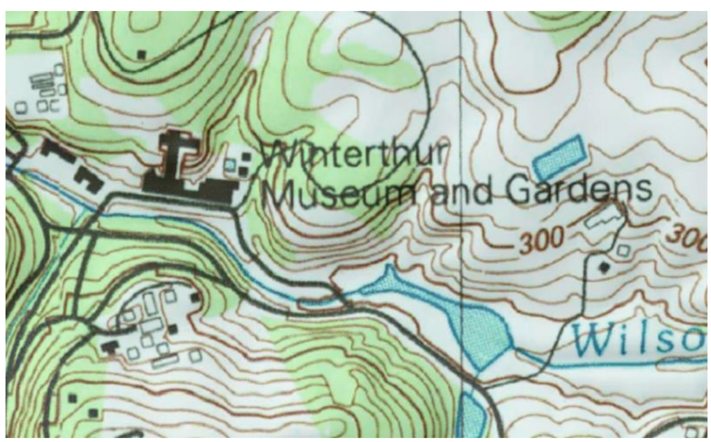

Figure 12: Local topography of Winterthur [29].

important context to conservation efforts at the museum and provides data that can be used in analyses.

Figure 13: Local topography of Winterthur [30]. Museum is located at starred location.

3.3 Climate and Extreme Weather Events

A coastal area, Delaware has a moderate year-round climate, with about 10 degrees less yearly temperature fluctuation than non-coastal areas [31]. Average monthly temperatures range from 75.8 to 32.0 degrees Fahrenheit in the area [31]. This may seem a wide range to control for, but it is due to seasonal fluctuations, and thus occurs over a long period of time that is easier for conservators to control. However, predicting the seasonal changes will become more difficult as climate change continues. Understanding how the local and regional climate affects microclimate control at Winterthur will greatly assist conservation efforts.

Chapter 4: Data

There are several sources of data for this thesis. The interfaces are designed to work with sensor data collected from the Winterthur sensors. Weather data is used for factor analysis. Site visit data is used to determine which rooms should be analyzed.

4.1 Sensor Data

Winterthur's PEM2 (Preservation Environment Monitor) Datalogger microclimate monitoring sensors, made by the Image Permanence Institute [32], have been collecting temperature and relative humidity data every 15 minutes since 2012 in 175 rooms at the museum. The sensors provide accurate data, as they are calibrated and checked with other measurement devices by conservators. This data is currently stored online using a cloud software called eClimateNotebook, as discussed in Chapter 2 [16].

4.2 Weather Data

Weather data was sourced from the National Oceanic and Atmospheric Administration's (NOAA) National Center for Environmental Information (NCEI) Climate Data Online Search [33]. Local Climatological Data was obtained, which records hourly data points for many variables, of which the following are used in this thesis: dew point temperature, dry bulb temperature, precipitation, relative humidity, station pressure, visibility, wet bulb temperature, and wind speed. Wilmington Airport, Delaware was selected as the location to collect data from as it was the closest available source of data to Winterthur, at 11.2 miles away. The Station ID for this location is WBAN:13781. Data was obtained for the date range of 13 January 2003 to 3 March 2020 for Wilmington, DE. Weather data arrived in .csv format. Explanations on how to filter a .csv file can be found in Section 5.5.

4.3 Site Visit Data

During a site visit to Winterthur on January 27, 2020, a better understanding of the needs of Winterthur was gathered [34]. Winterthur conservators were consulted with to understand the key issues present at Winterthur regarding monitoring needs. Specific rooms that are subject to frequent microclimate fluctuations were identified. In particular, conservators noted four rooms surrounding a portion of the roof on the third floor, what they call the "swimming pool" due to its tendency to collect water, are particularly difficult to control [34]. These rooms are the Textile Study room, the Textile Room – Needlework Study, the Rug Storage room, and the Pad Room, also known as the Costume Storage room. In Figure 14, the four rooms are highlighted in orange and the "swimming pool" portion of the roof is shaded in blue. The empty boxes to the left of the roof are also rooms that surround the "swimming pool," but they are not labeled on the third-floor plan. The Textile Room – Needlework Study has a heavily cycling temperature and relative humidity, and conservators wish to better understand why. Also on the third floor, the China Trade Room leaks. Conservators noted that the first, second, and third floor of the museum are all partially subterranean, in that a portion of rooms are underground whereas those at the other end of the building are above ground [34]. These different areas are identified on the third floor by green circles in Figure 14.

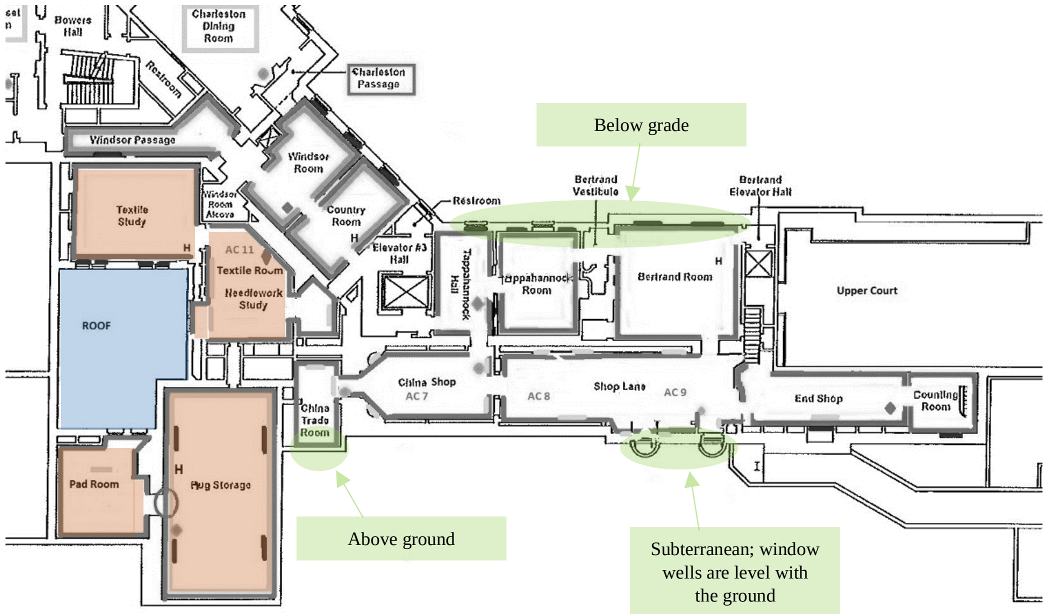

Figure 14: Winterthur Third Floor plan [28]. Areas of interest highlighted.

Another problem location that conservators pointed out is the seventh floor. In winter, the temperature and relative humidity on the seventh floor are not controlled. The conservators believe this is due to the exterior exposure on the upper floors, as the footprint is narrow and so there is lots of exposure on either side. Figures 15 and 16 show the narrowness of the structure on both the seventh and eighth floors. Conservators would like to know at what exterior temperature it becomes difficult to control temperature and relative humidity on the seventh floor.

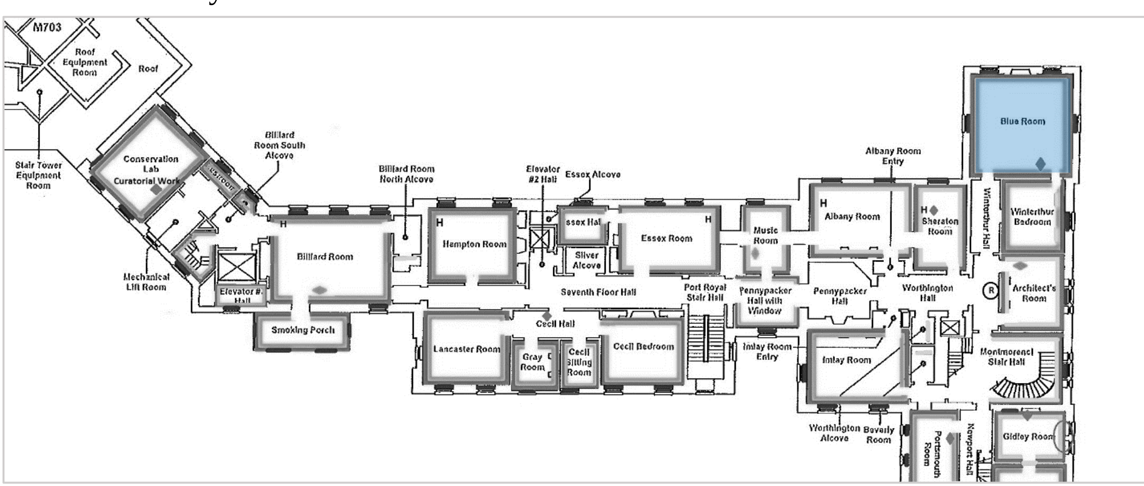

Figure 15: Winterthur Seventh Floor [28]. Blue Room highlighted in blue.

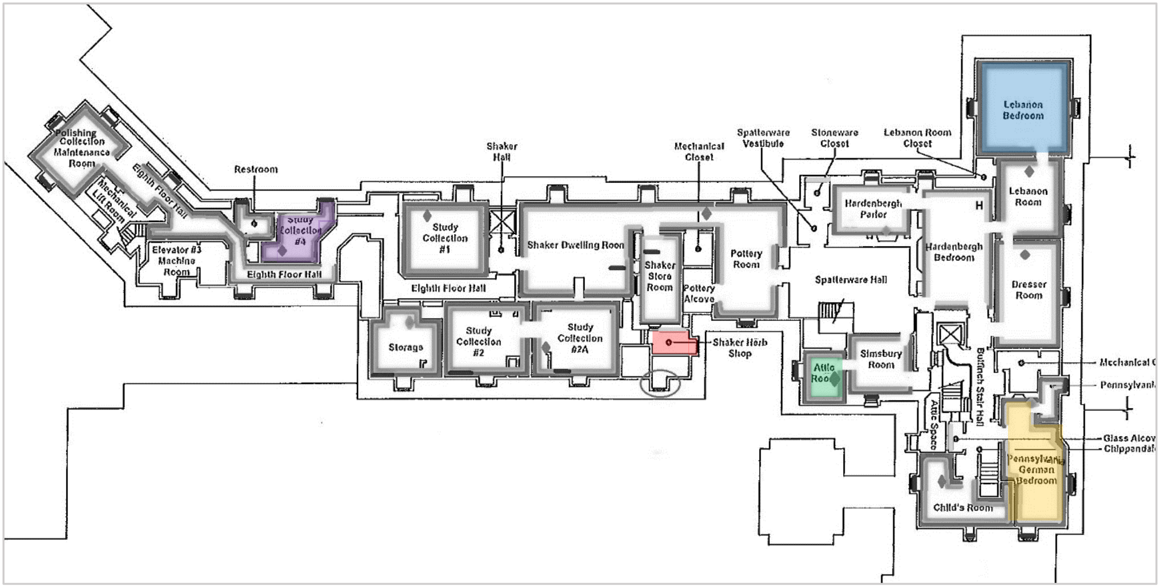

Figure 16: Winterthur Eighth Floor [28]. Rooms of interest highlighted.

On January 27, 2020, there was condensation on all of the windows on the seventh floor. As well, the seventh floor was much warmer than the eighth floor. On the eighth floor, there have been leaks and mushrooms in the Pennsylvania German Room highlighted in orange in Figure 16. The 8th floor Attic Room, highlighted in green in Figure 16, is difficult to control in the winter. In the Attic Room, there have been mercury amalgam drips; scientists differ on what causes mercury amalgam to drip, but it is likely due to high relative humidity. The Lebanon bedroom has leaked along the back wall, which went down into the Blue Room on the seventh floor (both rooms are highlighted in blue in Figures 15 and 16). There is spalling in the Shaker Herb Shop, highlighted in red in Figure 16. Study Collection IV, highlighted in purple in Figure 16, has mold by the window; conservators have opened the window as of winter 2019-2020 above the hallway door to see if that causes any improvement.

There are many fireplaces within the museum, but they were all capped in 2013-2014. Some fireplaces do not even have chimneys, having been purely decorative from the start. Fireplaces with chimneys, despite being capped, could still influence relationships between floors. Doors are generally left open in all rooms unless they are being used as storage rooms. This is important to note because it indicates that rooms likely influence each other in some way due to being constantly connected by doorways.

There is data as to what tour routes are followed at what time of year and through what portions of the house as well as how many visitors are present each time. As well, there are plans of the landscaping that surrounds the house that could be used to estimate shading and wind blockage. Although this data has not been incorporated into this thesis, it will provide ample data for future work.

Chapter 5: Methods

The various analyses and principles that are used to achieve the results of this thesis are outlined in the following sections.

5.1 General Programming Languages and Packages

Python was the main programming language used for this project. It was selected due to its versatility and ability to be integrated with a variety of other programming languages. Project development occurred within the PyCharm development environment from JetBrains [35]. The open-source framework for web applications Dash by Plotly was used for interface development [36]. Visualization on floorplans was enabled by the dash-floorplan component developed by Wes Reinhart, discussed further in Section 5.4.1 [37]. The Pandas library was used for managing large datasets [38]. Ploty.graph_objects was used for visuals such as time series and bar charts. NumPy was used for creating arrays and matrices [39]. Base64 was used to decode and encode image data after upload to an interface [40]. Google Drawings enabled uploading images for floorplan visualization [41]. HTML and CSS were used to style the interfaces, using CSS stylesheets from a Dash sample app [42]. All code can be found in the Appendix, Section 8.1.

5.2 Analyses and Optimization

5.2.1 Bounds and Swing Analysis Method

A bounds analysis determines how often the data goes outside of bounds set by the user, i.e. how often it is over the maximum and under the minimum desired value. Figure 17 shows a visualization of a bounds analysis. The green box represents the desired range for the variable shown in blue. This is determined by comparing every data point to the desired minimum and maximum. A percentage is returned for how out many points are out of bounds. This total percentage out of bounds can be broken down into the percentage over the bounds and the percentage under the bounds.

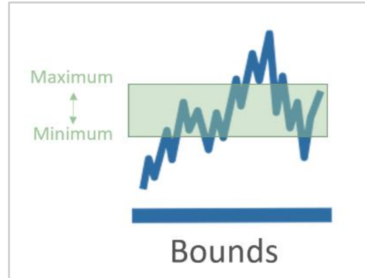

Figure 17: Visualization of bounds analysis.

A swing analysis determines how often the difference between the maximum and minimum value of a parameter within a 24-hour time period exceeds a desired value. Figure 18 visualizes a swing analysis. The green box represents the time period to be examined for swing. The difference between the minimum and maximum value in the time period, shown by the orange arrow, is the swing for that 24-hour period. The analysis compares every set of points spanning 24 hours in a submitted dataset and returns the percentage of times swing was over the desired value.

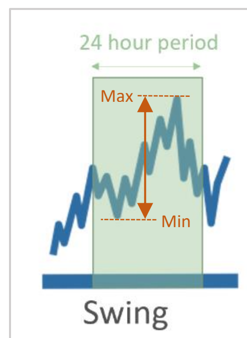

Figure 18: Visualization of swing analysis.

Bounds and swing analyses provide a helpful basic method for understanding what is going on within a room and a building. By examining data from specific seasons, trends can be identified regarding seasonal changes, which allow conservators to adjust how they control temperature and relative humidity depending on the season. As well, comparing various rooms using bounds and swing analyses can identify outlier rooms. If all but one room tend to follow a similar pattern regarding being within a desired range for bounds or swing, then that one room can be examined more closely in other ways for explanations as to its differing behavior. For example, Room B, Room C, and Room D might all tend to be within bounds during the summer, but then in the winter Room D might suddenly diverge from the rest of the group and go out of bounds, indicating that somehow control has become a problem in this room, in a way that is more worrisome than if it had been out of bounds all along. This helps conservators focus their action and understand what changes are affecting every room and what changes might be evidence of a room-specific problem. Bounds and swing analyses also indicate to conservators exactly what is going on inside each room and give them concrete values with which to work. While there are some insights to be gained about the relationship between rooms, primarily bounds and swing analyses serve to highlight current problem areas in the structure. A conservator is able to identify that Room A might have an issue with its temperature control, as it has been swinging over the desired range for the past month, or that Room B was once out bounds constantly but after an effort to fix relative humidity control, has stabilized and stays within bounds. Bounds and swing analyses are an important part of the toolbox that conservators can utilize to understand the microclimate control of the building.

5.2.2 Cross-Correlation Methods

Cross-correlation analysis seeks to understand the relationship between two variables relative to each other. For instance, one could examine the time of year, number of people in a room, and temperature; cross-correlation determines how related each one is to the other variables [43]. Figure 19 shows a hypothetical example of a cross-correlation matrix.

Every set of data, in this case data from Room A, Room B, Room C, Room D, and Room E, is plotted against one another. With a matrix such as this, the user can see relationships between all of the variables at once. Note that any variable plotted against itself is linear, because each variable has 100% correlation with itself. Some rooms, such as Room A and Room C, show no visible correlation. Others, such as Room E and Room C appear to be highly correlated due to the linear nature of their correlation plot. Cross-correlation returns a correlation number for each of the scatter plots in the matrix in Figure 19, informing the user of precisely how correlated two variables are. Examples of numerical matrices can be seen in Section 5.2.3.

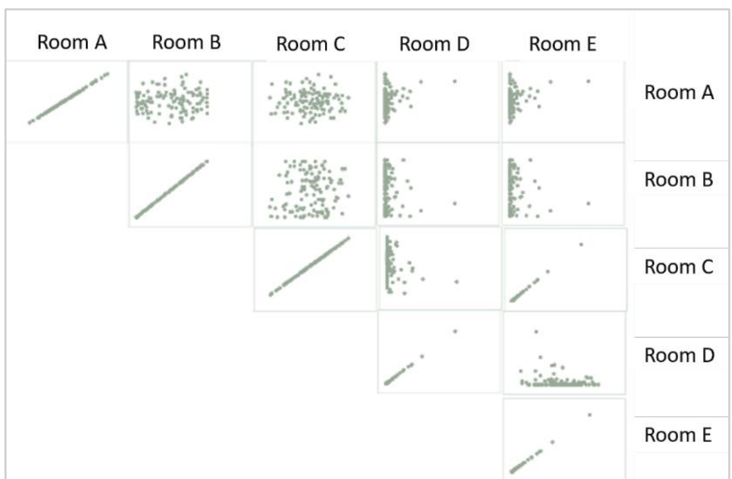

Figure 19: Hypothetical example of a cross-correlation matrix scatterplot.

Cross-correlation is implemented in Python by storing the data to be used in a Pandas dataframe and then using the Pandas function Dataframe.corr() [44]. Any empty (NaN) values are automatically excluded.

Cross-correlation allows for a better understanding of relationships between rooms in a different way than bounds and swing analyses do. Cross-correlation puts numbers to the trends that might be seen in visualization of bounds and swing analyses and allows the user to know just how much two rooms are related to each other. It is important to note that correlation does not denote causation, but it is all the same a valuable tool for conservators. For instance, conservators could examine whether all the rooms on the same floor are correlated or whether rooms on different floors on one side of a structure are correlated. Cross-correlation helps conservators begin to develop a sense of what factors might be at play within a structure. For instance, Room A and Room B might be on different floors and not directly above one another, but upon noting that they are highly correlated, the conservator can begin to look for underlying factors uniting them. Identifying precisely the factors that are at play behind correlations is examined further in the next section on factor analysis.

5.2.3 Factor Analysis Methods

Factor analysis is a machine learning technique used to analyze big datasets. It accounts for variables that are not measured but that can affect the measured variables [45]. Factor analysis helps one determine what variables are most important to look at. For example, if one measures sunlight strength, the number of people in a room, and temperature, factor analysis would determine underlying factors, and weight each of the measured variables according to how much they correlate to the underlying factors. The variable with the highest weight is most correlated with the factor. At Winterthur this allows conservators to see what unmeasured variables are affecting the measured temperature and relative humidity data.

Exploratory Factor Analysis seeks to determine the underlying, unobserved variables that can explain relationships between observed variables. There are two types of Exploratory Factor Analysis: Principal Component Analysis (PCA) and common factor analysis (referred to as “factor analysis” in this work). Both seek to explain variance among variables through fewer unobserved factors than the number of parameters being measured. The literature review in Section 2.1.1 noted that almost all research done with sensor monitoring has used PCA and not common factor analysis.

Exploratory Factor Analysis looks at variance amongst factors. Common variance is that which is shared among a set of observed variables, while unique variance is all other variance, that can only be attributed to the one observed factor [10]. Figure 20 shows how common variance and unique variance sum together to form total variance for one observed variable. Unique variance is also split into specific variance and error variance; specific variance is related to the observed variable itself, while error variance accounts for randomness in data collection.

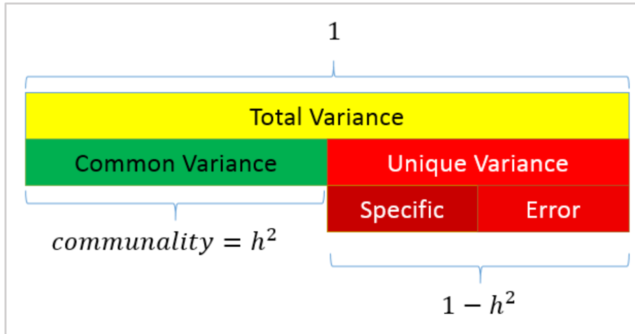

Figure 20: Factor analysis variance explanation [10].

In contrast to PCA, factor analysis assumes there can be unique variance. PCA assumes that the total variance is equal to the common variance, in other words, that there is no unique variance [10]. UCLA writes that PCA is more appropriate when one simply wants to reduce the number of variables down, to simplify the problem; factor analysis is more useful when one wants to understand what factors are causing what things [10]. The primary purpose of PCA is to reduce the number of variables. Factor analysis assumes that variance for a given observed variable is made up of both common and unique variance. Factor analysis accounts for errors in data collection and does not merely reduce the number of variables to be worked with as PCA does, but rather can help identify relationships between observed variables [10]. For these reasons, factor analysis, and not PCA, is used in this thesis.

Before beginning factor analysis, it must be confirmed that the dataset is suitable for factor analysis. This is done using Bartlett's Test of Sphericity and the Kaiser-Meyer-Olkin (KMO) Test. The correlation matrix of the variables being examined is used throughout the factor analysis process.

Bartlett's Test of Sphericity determines how close to the identity matrix the correlation matrix of the data is [46]. If the correlation matrix for observed variables was a perfect identity matrix, such as in Figure 21, then it would not be suitable for factor analysis, as the observed variables are not related in any way. However, a correlation matrix like that in Figure 22 is more likely to be suitable for factor analysis as each room is highly correlated with all the other rooms. Bartlett's Test returns a chi-square value and p-value that indicate whether or not the data is suitable for factor analysis. A p-value of less than or equal to 0.05 indicates that the data is suitable for factor analysis.

	Room A	Room B	Room C
Room A	1	0	0
Room B	0	1	0
Room C	0	0	1

Figure 21: Sample correlation matrix in which the correlation matrix is the identity matrix.

	Room A	Room B	Room C
Room A	1	0.8	0.9
Room B	0.8	1	0.7
Room C	0.9	0.7	1

Figure 22: Sample correlation matrix in which all rooms are highly correlated with one another.

Bartlett's Test is implemented in Python using the factor_analyzer.factor_analyzer package [47]. The formula is:

$$-1 * (n - 1 - \left(\frac{2p + 5}{6}\right) * \ln(\det(R)) - 1 * (n - 1 - \left(\frac{2p + 5}{6}\right) * \ln(\det(R))$$

where det(R) is the determinant of the correlation matrix and p is the number of variables. The function tests the hypothesis that the correlation matrix of the data is equal to the identity matrix, and outputs the chi-square value and p-value.

The KMO Test estimates how much of the proportion of the variance of each observed variable is due to common variance [46]. It measures how many factors would be associated with very few variables. This test returns a KMO value, which indicates whether the data is suited to factor analysis. KMO values range from 0 to 1, and the closer to 1 the KMO value, the better suited to factor analysis the dataset is. Generally, factor analysis should not be run for a dataset if the KMO value is less than 0.60. A KMO value near 0 indicates that the sum of partial correlations was large compared to the sum of correlations, which means that correlations are occurring between many variables and not just a few, which is a problem for factor analysis [48]. High partial correlations mean that factors determined by factor analysis would not be distinct, nor reliable [49]. The KMO test formula is

$$MSA = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} u}$$

where MSA is the Measure of Sampling Adequacy which ranges from 0 to 1, $R = [r_{ij}]$ is the correlation matrix, and $U = [u_{ij}]$ is the partial correlation matrix [50].

The KMO Test is also implemented in Python using the factor_analyzer.factor_analyzer package [47]. It returns the KMO score per item, as well as the KMO total value. The KMO total value is compared to 0.60, and if it is greater, the dataset is suitable for factor analysis.

In sum, Bartlett's Test checks that there is some correlation amongst variables, and the KMO Test ensures that that correlation between variables will allow for factor analysis to be fruitful.

After determining that the dataset is suitable for factor analysis, the number of meaningful factors with which to run factor analysis must be determined [14]. This is done by examining the eigenvalues of the correlation matrix of the dataset, which represent the total amount of variance that can be explained by a factor [10]. The number of eigenvalues over one indicate the maximum number of factors that should be analyzed for factor analysis, but the meaningful number of factors could be less than the number of eigenvalues greater than one [51].

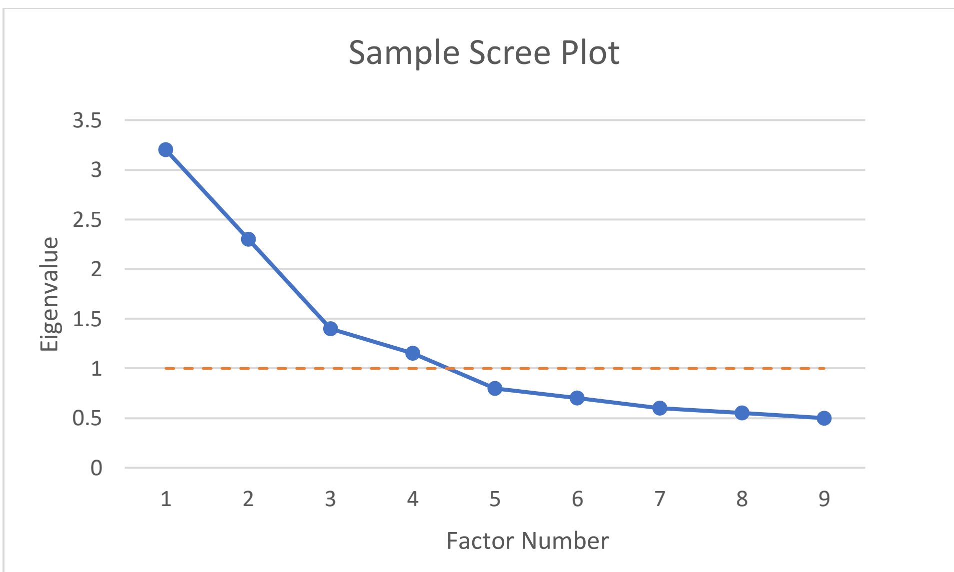

Figure 23: Sample scree plot.

To understand what number of factors should be selected, a scree plot is generated of each factor versus its eigenvalues, a method developed by Raymond Cattell in 1966 [52]. The factor number on the x-axis simply refers to how many total eigenvalues there are, and thus potential factors. Figure 23 shows a sample scree plot. In this sample, there are four eigenvalues that are greater than one; thus, the maximum number of meaningful factors is four. To determine the number of factors that should be used in factor analysis from this plot, the first change in slope is determined, what Cattell calls the “elbow” of the graph [52]. In Figure 23, there is a break in slope at the 3rd factor. This would indicate that 3 factors would be a good number of factors with which to start factor analysis.

However, there is no set method for determining the number of factors to analyze with factor analysis from a scree plot. Another method used to determine the best number of factors is to run factor analysis for various numbers of factors and look at how much cumulative variance is explained by each set of factors. Some would set a threshold, such as once 60% of the variance is explained by the factors used, running factor analysis with more factors is unnecessary [53]. Ledesma et al. suggest using Parallel Analysis to determine the number of factors in order to make reading a scree plot less subjective [54]. However, as determining if a factor is meaningful is a subjective task in and of itself, this thesis does not incorporate that method, giving the user freedom to interpret the scree plot.

Once the dataset has been confirmed as a good candidate for factor analysis and the number of factors to be used has been determined, factor analysis can be run. The varimax rotation, an orthogonal rotation, is chosen for factor analysis. The varimax rotation makes results easier to understand because the axes are rotated such that each factor ends up with a few parameters with large weightings and many parameters with small weightings associated with it [55]. This simplifies understanding because the parameters with large weightings associated with a factor are the most significant for that factor. The choice of orthogonal over oblique axes is made to maximize the difference between factors to ensure better interpretation.

Running factor analysis determines the loadings, or weightings, of the parameters associated with each factor. The parameters are the observed variables that the analysis was run on. The weightings indicate how associated the parameter is with the factor. For instance, Factor 1 might be associated with temperature in Room A, temperature in Room C, and relative humidity in Room B. The weightings are 0.90 for Room A temperature, 0.82 for Room B relative humidity, and 0.50 for Room C temperature. Thus, Room A has the highest association with Factor 1 and Room C the lowest. This thesis ignores parameters associated with a factor with a weighting less than or equal to 0.25. This number was selected because if a parameter is less than 25% associated with a factor, it is not very significant and then only serves to confuse those trying to interpret the results.

Finally, the total cumulative variance is determined. As discussed before, looking at the total cumulative variance indicates how much variance is explained by the factors determined through factor analysis. As well, the amount of proportional variance explained by each factor is presented. This enables the user to understand how much the factors resulting from analysis account for variance in the dataset.

In sum, factor analysis enables Winterthur to look at relationships between rooms in an even more sophisticated way than cross-correlation by identifying underlying factors relating them.

5.3 Human-Machine Interface Design Principles

Napolitano and Glišić sum up D.A. Norman's three principles that should be used for an optimal user interface: "Simplify the structure of the interface so that the tasks are intuitive," "visualize the results of actions and feedback," and "correctly direct a user to the actions they must take for the intended result" [56]. Napolitano and Glišić outline the importance of a user knowing that their click has done an action through state buttons [56]. As well, tasks should be divided up in a way that indicates to users the order in which tasks must be done. Napolitano and Glišić use tabs to divide processes in their work, effectively guiding the user through a series of steps [56]. In the case of Winterthur, the goal is that the user knows what environmental changes have occurred, how significant these changes are, and where in the house the changes have occurred. As well, visuals should be intuitive, easy to manipulate, and legible.

D.A. Norman writes that for interaction design, "the goal is to enhance people's understanding of what can be done, what is happening, and what has just occurred" [15, p. 14]. Norman continues by discussing affordances and signifiers. For the user of the interfaces created in this thesis, the interfaces' affordances are that they can be clicked on with a computer mouse and typed into with the keyboard, and the signifiers for the user are input text boxes and upload buttons that resemble buttons seen elsewhere. The signifiers tell the user where to click or enter text.

Mapping helps the user understand what is associated with what and how a system might function. For instance, turning a steering wheel clockwise turns a car to the right; this association feels intuitive to most [15, p. 21]. In this thesis, mapping is used in several ways. Just as a page being read is begun at the top of the page, the process of analysis for the interfaces begins at the top of the page. As well, steps of analysis are broken down into boxes so that subconsciously the user is inclined to group those actions together.

However, the designer should not expect the user to perfectly intuit the proper usage of the text boxes and buttons and so helpful hints that guide the user are necessary. Instructions are given as needed, as well as outputs during and after a process has completed. Outputs inform the user as to what has occurred. Error messages describe what the error is and offer potential solutions. These elements are what Norman refers to as feedback [15, p. 25].

This interface aims to have the order of steps of analysis be very clear. Steps are divided in visual bounding boxes. Steps that are completed together are grouped within the bounding box with a submit button at the end so that the user knows those steps function in conjunction. Steps are also broken up in bounding boxes on the interface to mirror how the steps are broken up in the code. Thus, users can understand what steps have to be re-run to generate new analyses; for example, in the Bounds and Swing Analysis for One File interface, laid out in Section 6.1, the user could change the date selection and analysis performed but not have to change the file uploaded. This is in contrast to having one final submit button, which would take much longer as any adjustment to parameters requires rerunning every step. As well, breaking the analysis down into multiple steps gives the user intermediate feedback that is difficult to provide with only one final submission button. In this sense, there is great importance in the buttons of this thesis's interfaces. They not only notify the user that a process has been done, but in requiring the user to click to begin a process, they also ensure that the user understands exactly what processes have been run because the user themself initiates every process. A further discussion of design principles learned through designing the interface takes place throughout Chapter 6 as each interface is explained in depth.

5.4 Visualization Methods

Upon completing analysis, the results must be translated to the user. For Winterthur, that end user is a diverse audience and so the easier the human-machine interface is to use and the simpler the data visualization is to understand, the more effective the final product will be. By improving data visualization, it will be easier to take action on important insights.

5.4.1 Visualization with Dash Floorplan

Visualization on a floorplan builds off of the floorplan visualizations by King et al. detailed in Section 2.3 [17]. This thesis takes the concept of visualizing the results of analysis of multiple rooms in a structure at once on a floorplan and applies it to more areas than just bounds and swing analyses. This enables conservators to easily see spatial relationships between associated variables.

Visualization on floorplans was enabled by the dash-floorplan component developed by Wes Reinhart [37]. The component is written in React, a JavaScript library. The user sends data to the visualizer in dictionary form, in which each variable serves as a key with its entry being the desired polygon color. The color is based on a spectral scale. The names of the data in the variables show up in the visualizer and then the user can draw polygons and associate the polygons with each variable using the toolbar at the top of the component. The polygon will then change from grey to the color associated with that variable. Full documentation can be found at https://github.com/wfreinhart/dash-floorplan [37]. All output floorplans can be saved by using the Print-Screen function on the user's computer or another similar screen snipping tool. Zooming in and out of the browser window ensures that all desired parts of the floorplan are captured in a screen capture.

Specifics to visualization on a floorplan for each analysis are discussed in the following sections.

5.4.2 Bounds and Swing Analysis Visualization Methods

For understanding bounds and swing analyses results, times series, contour plots, and visualization on a floorplan are used to visualize the data and results. Time series plots are used to visualize the data collected. For bounds analysis, a time series graph allows the user to see when the data goes outside of the desired bounds, which are also plotted on the graph. For swing analysis, the user is able to examine instances of swing reported through the analysis. Contour plots are used to help visualize seasonal changes. The contour plot graphs year on the x-axis and month on the y-axis. The color of the plot changes based on the percentage of points outside of the desired bounds or swing range according to a defined scale. This allows the user to quickly spot seasonal tendencies. Lastly, visualization on a floorplan is used to determine what rooms are experiencing similar fluctuations. When analyzing multiple files at once for bounds and swing, this helps the user see potential correlations. Also, after running a correlation analysis to determine relationships between rooms, the user can go back and run a bounds or swing analysis to determine if the correlated rooms share a tendency to go out of bounds or to swing.

5.4.3 Cross-Correlation Visualization Methods

Cross-correlation results are visualized using a scatter plot matrix, a heatmap, and visualization on a floorplan. The scatter plot matrix shows how every variable analyzed is correlated to the other variables through a scatter plot. By presenting the correlation scatter plots of all of the variables in one matrix, the user is better able to understand the relationships amongst all variables at once. A heatmap plots variables against one another in the same way that a scatter plot does, but instead of a scatter plot, it shows a certain color depending on the correlation value. The scatter plot provides a visual sense of the amount of correlation while the heatmap gives the actual value of correlation so that the user can check their intuition and know the precise correlation value.

For cross-correlation, visualization on a floorplan highlights the rooms analyzed based on their correlation to one selected "main" room. This is, in essence, plotting the values from one column of the scatter plot matrix. The rooms are highlighted in a color defined by a scale that indicates the extent of their correlation to the "main" room.

5.4.4 Factor Analysis Visualization Methods

To visualize factor analysis, bar charts and visualization on floorplans are used. Bar charts show the weightings of the parameters associated with a factor, allowing the user to quickly note which parameters have the greatest weighting, and which are positive or negative. Visualization of weightings on the floorplan is done through the usage of a color scale. One factor is visualized on a floorplan at a time; each parameter associated with that factor is highlighted in the color corresponding to its weighting to that factor. Exterior weather can be visualized on the floorplan as well in the same way. Rather than highlighting a room, a box is drawn outside of the rooms to indicate it is exterior weather data.

5.5 Combining Data Types for Analysis

Data from Winterthur comes in .pm2 format, while weather data used in this thesis arrives in .csv format. In order to use this data in the Factor Analysis interface (Section 6.4), the data must be combined into one file. There are several steps to combine the two datasets into one.

First, the user launches the Factor Analysis interface, and uploads all .pm2 files to be used. These files will be made into a dataset that can be saved to the user's computer as a .csv file. Second, the user adjusts and runs the code in prepareCSV.py (see Section 8.1 for the code) to interpolate, select the proper date range, and save the exterior weather data to a new .csv. The user can adjust the code in prepareCSV.py to save the file where they would like and to change the date range selection. Next, the user then opens both new .csv files, one made from the .pm2 files and one from the exterior weather data, using Microsoft Excel [57] or another .csv visualizer. The user must confirm the datetimes match up when they copy and paste the columns from the exterior weather data file into the other new .csv file. The user now has a combined .csv file. The last step is to label the datetime column, which should be column A, as "Date and Time in GMT" and then save the file. The user can then upload this combined .csv to the Factor Analysis interface and perform factor analysis. An example of this process is found in Section 6.6, and a sample of a .csv produced using this method can be found at the link in Section 8.1.

5.6 Dew Point Calculation

Dew point is based on temperature and relative humidity; at Winterthur, dew point is calculated using the temperature and relative humidity values collected by the sensors. Dew point is calculated using the following formula:

$$Dew\ Point\ (Degrees\ Celsius) = \left(\frac{RH}{100}\right)^{\frac{1}{8}} * (112 + 0.9 * T) + 0.1 * T - 112$$

where RH is Relative Humidity in Degrees Celsius and T is Temperature in Degrees Celsius.

Chapter 6: Interfaces: Results and Discussion

The following sections show each interface produced by this thesis in great detail. Images of each interface in use are provided along with a walkthrough of how each one works. Most of the images of the interfaces in the following sections were generated using rooms surrounding the "swimming pool" identified in Section 4.3, providing a useful example. The link to the final code for all interfaces can be found in Section 8.1. All output graphs can be downloaded as .png files to the user's computer using the tools that appear when the user's mouse hovers over the figure.

6.1 Bounds and Swing Analysis for One File Interface

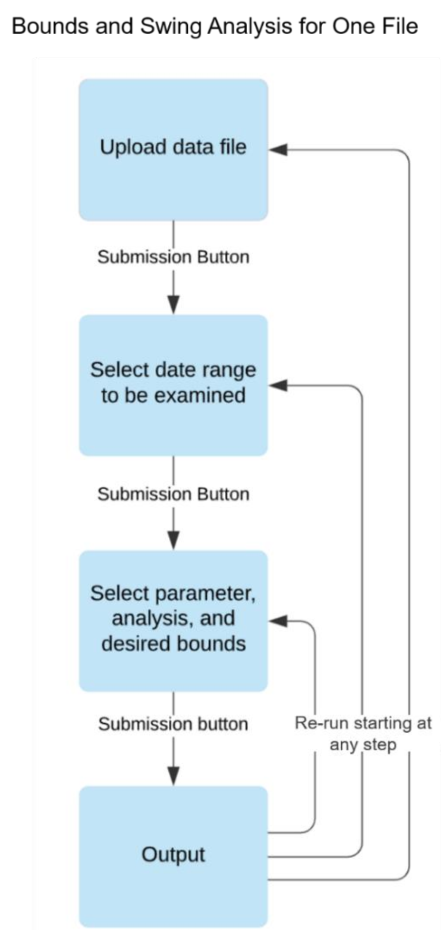

Figure 24: Flow map of Bounds and Swing Analysis for One File interface.

Figure 24 outlines the Bounds and Swing Analysis for One File interface. To begin, the user uploads the .pm2 data file that they would like to visualize. A file dialog opens upon clicking to upload the file, which allows the user to easily see all files available on their computer. Upon selection of a file, the name of the uploaded file is printed, as seen in Figure 25, allowing the user to at once confirm that the correct file was uploaded.

Figure 25: Upload data. Bounds and Swing Analysis for One File interface.

Next, the date range to be examined is selected by the user depending on what they wish to analyze (Figure 26). The user can select what specific months they want to analyze as well; this can be used to examine one season in specific, or to examine one month through the years. Date selection is submitted upon clicking the submit button. The dates selected are confirmed upon clicking submit, as are the dates that the data file was constrained to. If the data file's date range is smaller than the desired range, the output notifies the user as to the date range that the file was constrained to. In this way, no user will be under false assumptions as to what date range is being used.

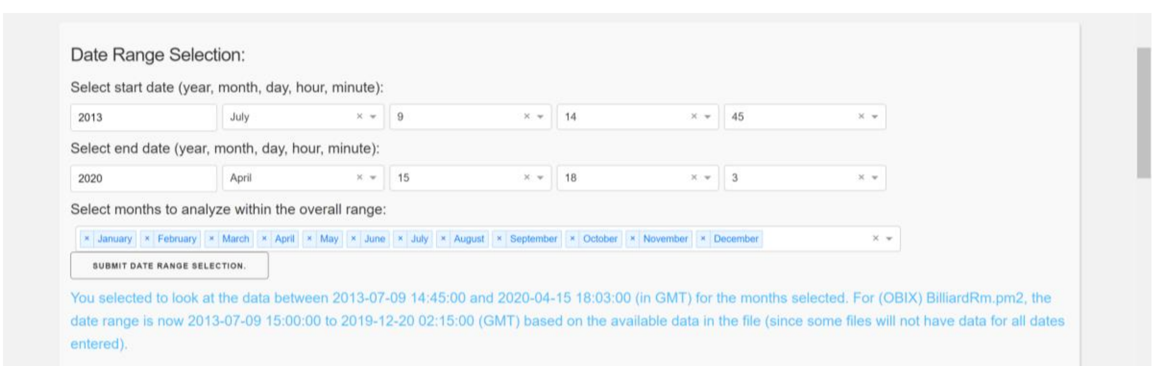

Figure 26: Select dates. Bounds and Swing Analysis for One File interface.

Next, the type of analysis to be done is selected. Users choose the parameter they wish to analyze from a dropdown menu; the options are temperature, relative humidity, and dew point. Then the type of analysis to be performed is selected – either a bounds analysis or a swing analysis. Finally, the user enters the desired range for the parameter if performing a bounds analysis, or the desired maximum swing for a swing analysis. To minimize user confusion of where to enter values, two input boxes show up for entering the bounds analysis range and only one for the swing analysis range. The input also features text that shows up when a user completely deletes an entry, telling them which box is for minimums and which is for maximums, shown in Figure 27. Based on which parameter is being analyzed, the interface suggests various entries. For a bounds analysis, the suggested entries for each parameter can be seen in Table 1. For a swing analysis, the suggested entry is a maximum swing of 10. Upon selecting and entering parameter, analysis, and bounds/swing range, the user clicks submit. This is seen in Figure 28.

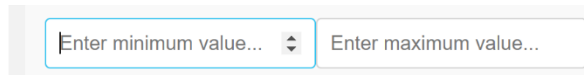

Figure 27: Input boxes offer suggestions when empty. Bounds and Swing Analysis for One File interface.

Parameter	Suggested Minimum	Suggested Maximum
Temperature (°F)	63	74
Relative Humidity (%)	35	57
Dew Point (°F)	37	56

Table 1: Values the interfaces suggests to the user for bounds analyses.

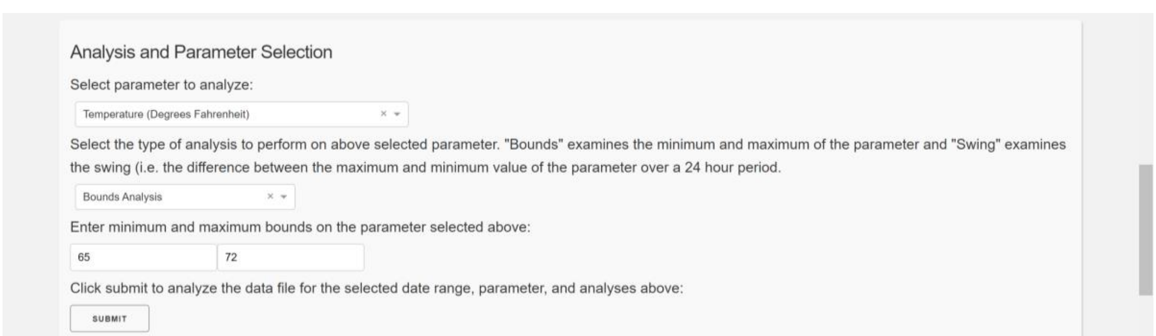

Figure 28: Analysis and Parameter Selection. Bounds and Swing Analysis for One File interface.

This generates several final outputs. For a bounds analysis, a text output provides the overall maximum and minimum value of the parameter over the time range, seen in Figure 29, as well as the percentage of data points over the maximum and under the minimum. The total percent of data points out of bounds is also given. For swing analysis, the total percentage of points that had a swing greater than desired is given as well as the maximum swing.

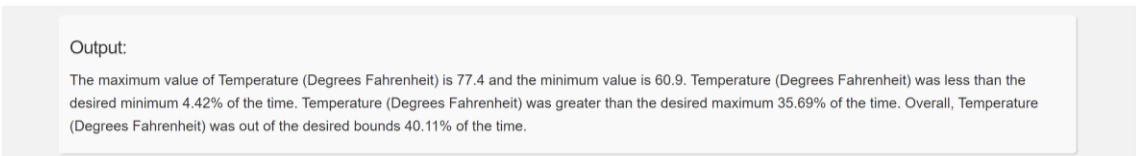

Figure 29: Output text for a bounds analysis. Bounds and Swing Analysis for One File interface.

A time series is also output, seen in Figure 30. For a bounds analysis, the time series shows the parameter value as well as horizontal lines indicating the desired bound range. For a swing analysis, the parameter is plotted over time and no bound lines are shown.

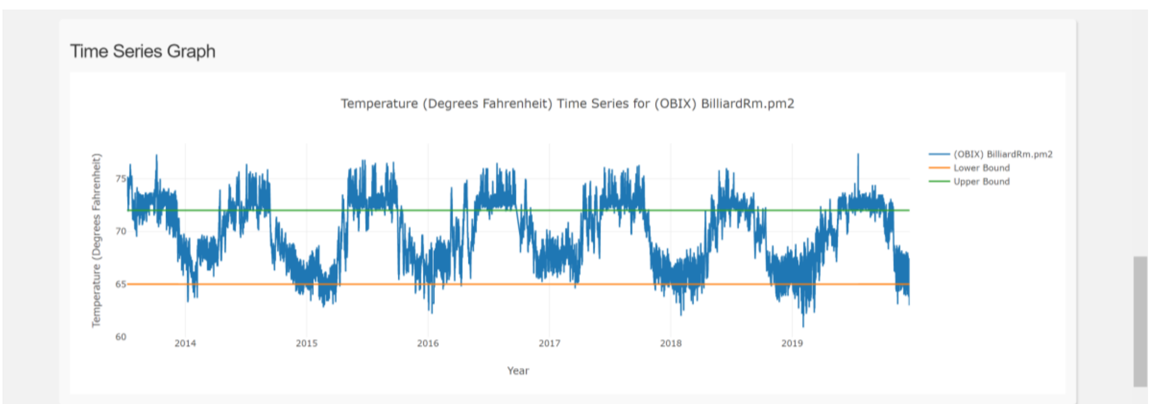

Figure 30: Time series for a bounds analysis. Bounds and Swing Analysis for One File interface.

Lastly, contour plots are output, seen in Figure 31. For a bounds analysis, two contour plots are output. One shows the percentage of points below the minimum bound. The other shows the percentage of points over the maximum bound. For a swing analysis, one contour plot is output. It shows the percentage of points over the desired swing.

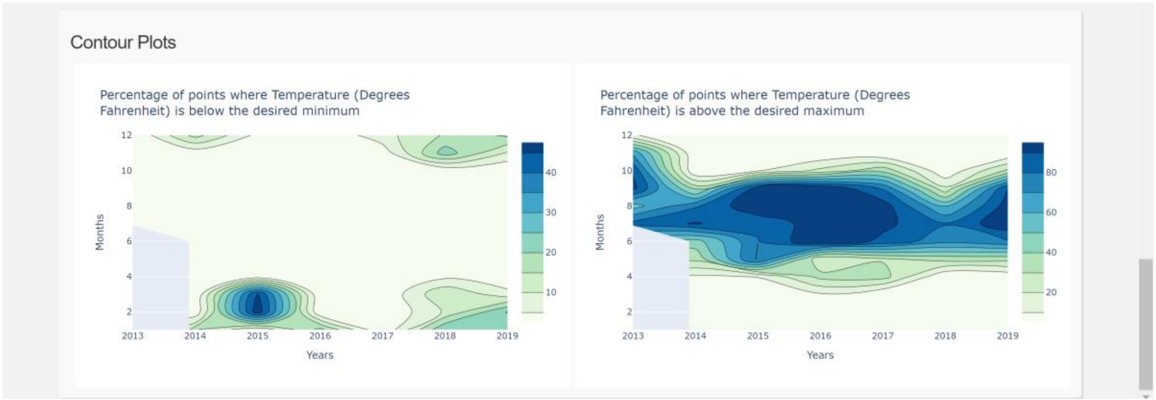

Figure 31: Contour plots. Bounds and Swing Analysis for One File interface.

Tasks are separated into small operations so that the user can easily go back and adjust the specific part they wish to adjust. For instance, after selecting a date range, the user is able to change the analysis selection without adjusting the date selection. This makes the interface more efficient from a coding perspective and streamlines user interaction.

6.2 Bounds and Swing Analysis for Multiple Files Interface

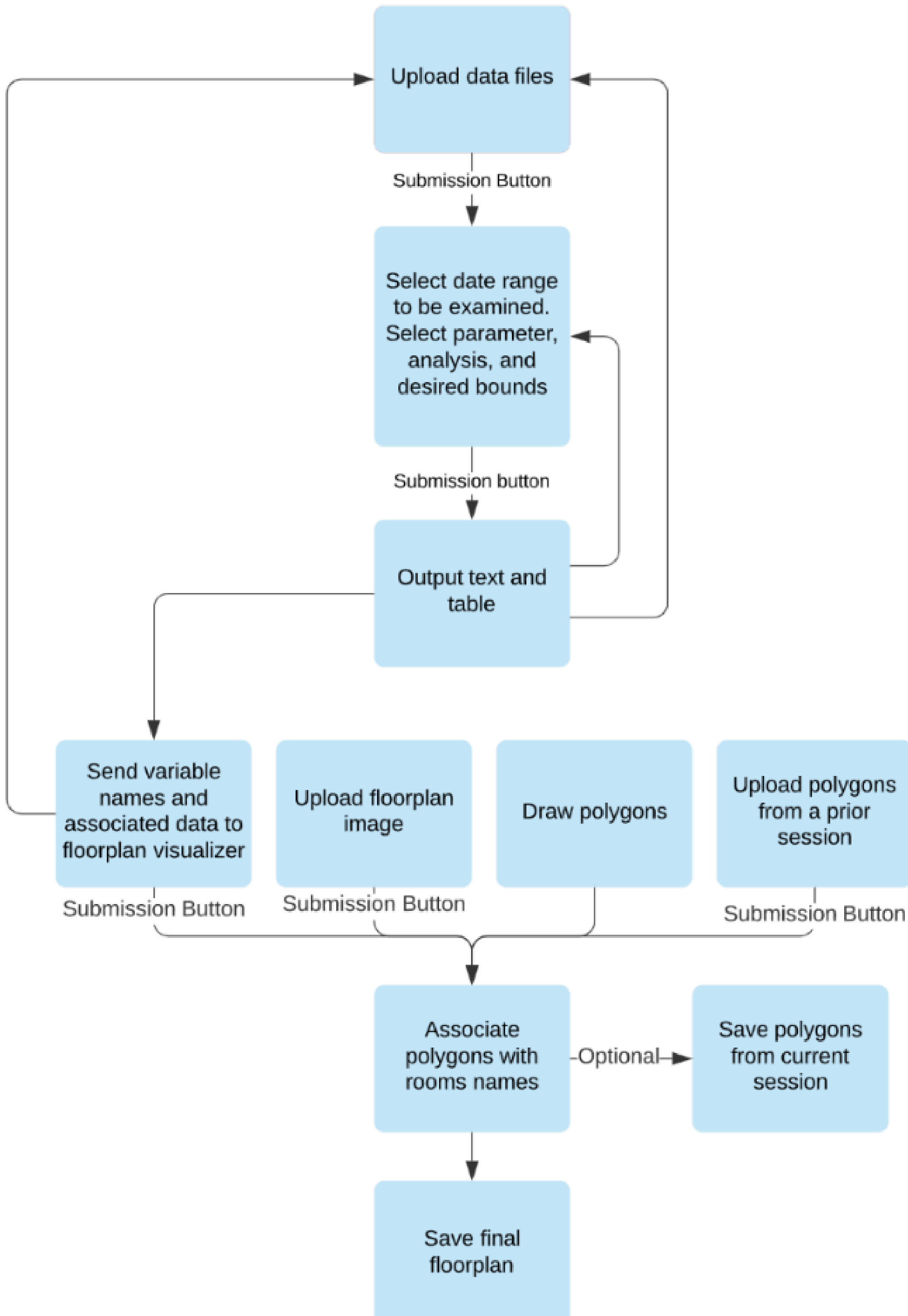

Figure 32: Flow map of Bounds and Swing Analysis for Multiple Files interface.

Figure 32 outlines the Bounds and Swing Analysis for Multiple Files interface. To begin, the user uploads the .pm2 data files that they would like to visualize, seen in Figure 33. The names of the uploaded files are printed, allowing the user to at once confirm that the files were uploaded and that they are the desired files.

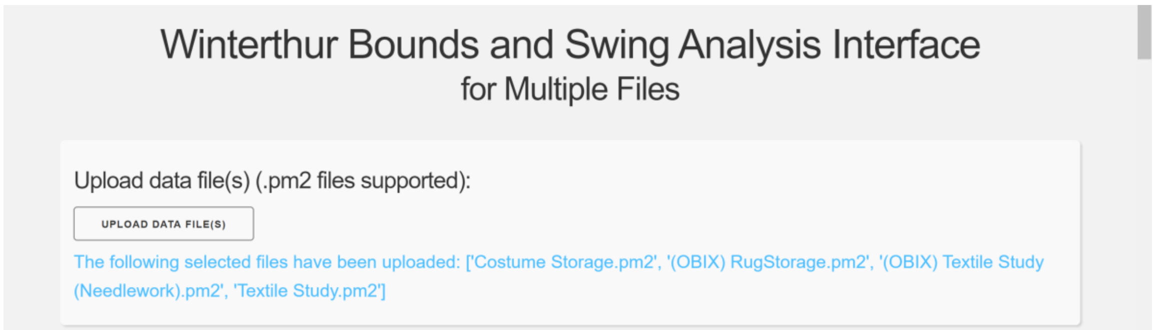

Figure 33: Upload data. Bounds and Swing Analysis for Multiple Files interface.

The date range selection and parameter, analysis, and bounds selection are the same as in Bounds and Swing Analysis for One File, described in Section 6.1, except that rather than having a submission button after data range selection, there is only one after the second set of selections, seen in Figure 34. This is due to the nature of how the code handles multiple data files; by visualizing what processes must occur together in bounding boxes, the user can subconsciously comprehend how the code operates from the interface layout.

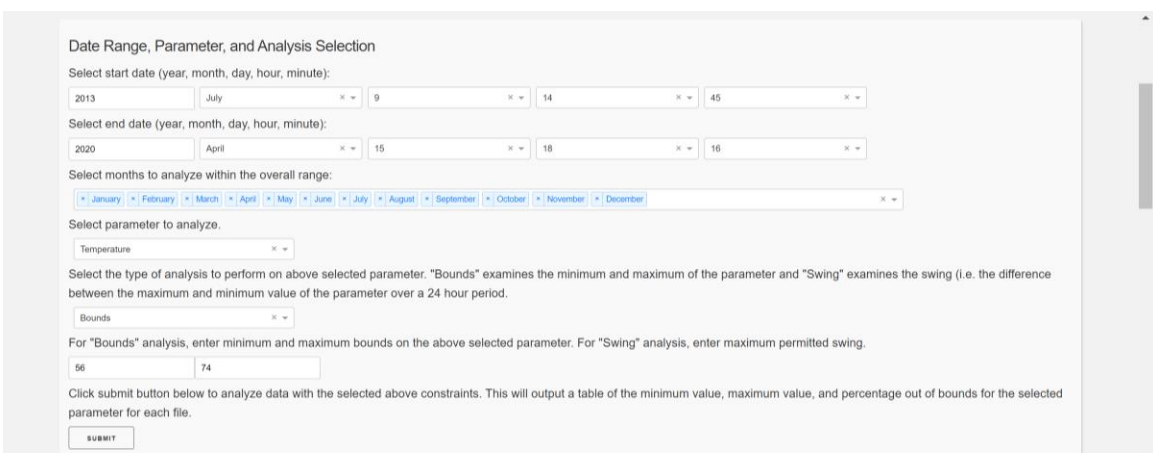

Figure 34: Date, Parameter, and Analysis Selection. Bounds and Swing Analysis for Multiple Files interface.

Submission of desired parameters outputs a text file and a table, shown in Figure 35. The text details the date range selected and the bounds selected so that the user can confirm the analysis was run for the parameters they wanted. For a bounds analysis, the table displays the minimum value, the maximum value, and the percentage out of bounds for the selected parameter for each file. For a swing analysis, the table outputs the maximum swing, and the percentage out of bounds for the selected parameter for each file. Upon seeing the table, the user can then go back and adjust the analysis bounds if they see that their bounds were not as desired.


Output:

You selected to look at the data between 2013-07-09 14:45:00 and 2020-04-15 18:16:00 (in GMT) for the months selected. You have selected to perform a Bounds Analysis for Temperature (Degrees Fahrenheit) where the minimum bound was 56 and the maximum bound was 74.

Room Name	Minimum Value	Maximum Value	Percent Out of Bounds Total (%)	Percent Over Upper Bound (%)	Percent Under Lower Bound (%)
Textile Study.pm2	57.7	73	0	0	0
(OBIX) Textile Study (Needlework).pm2	54.4	90.7	2	0	2
Costume Storage.pm2	52.8	77	1	0	1
(OBIX) RugStorage.pm2	54.6	75.5	0	0	0

Figure 35: Output text and table for a bounds analysis. Bounds and Swing Analysis for Multiple Files interface.

Next, the user can turn to visualization on a floorplan. Steps are separated both for computational efficiency, and so that the user can understand what each action outputs. One button sends the room names to the floorplan visualizer. Another sends the input floorplan image link to the floorplan, effectively “uploading” the image. Both can be seen in Figure 36.

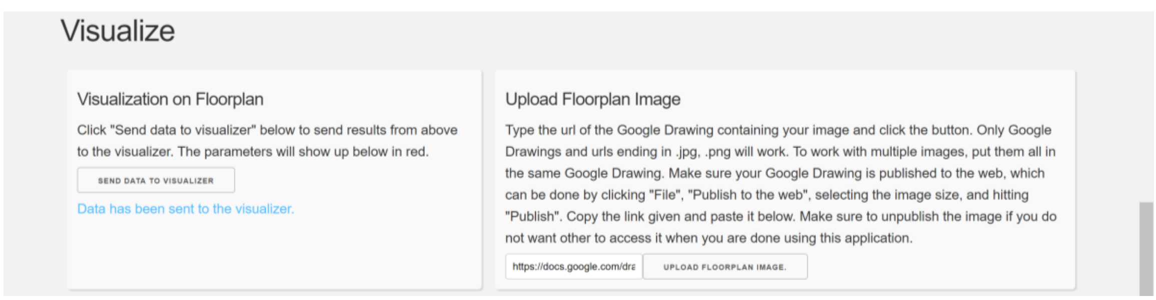

Figure 36: Visualization on floorplan instructions. Bounds and Swing Analysis for Multiple Files interface.

A section on polygons explains to the user how to create polygons, offers an input box to upload polygons saved from a previous session, and also has a button to save the polygons drawn during the current session (Figure 37).

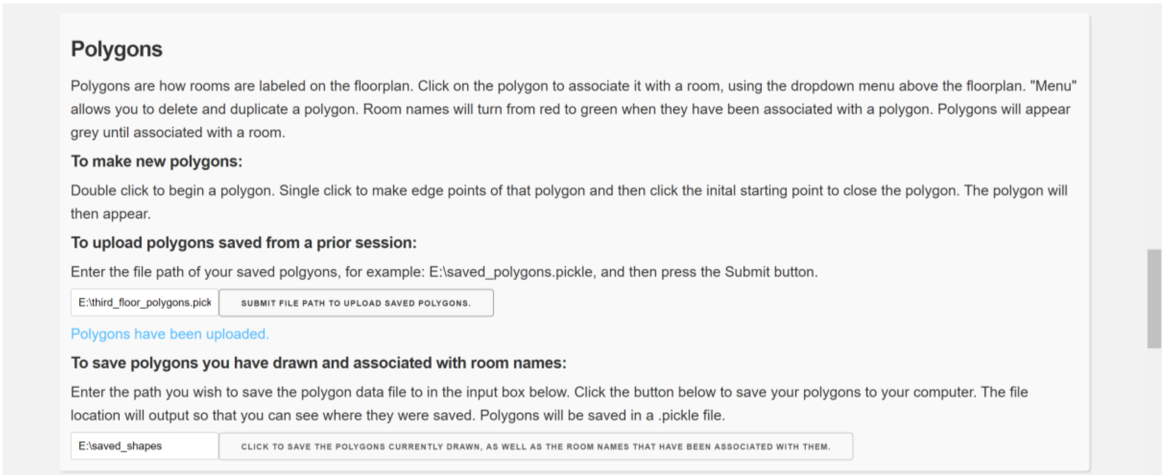

Figure 37: Polygons. Bounds and Swing Analysis for Multiple Files interface.

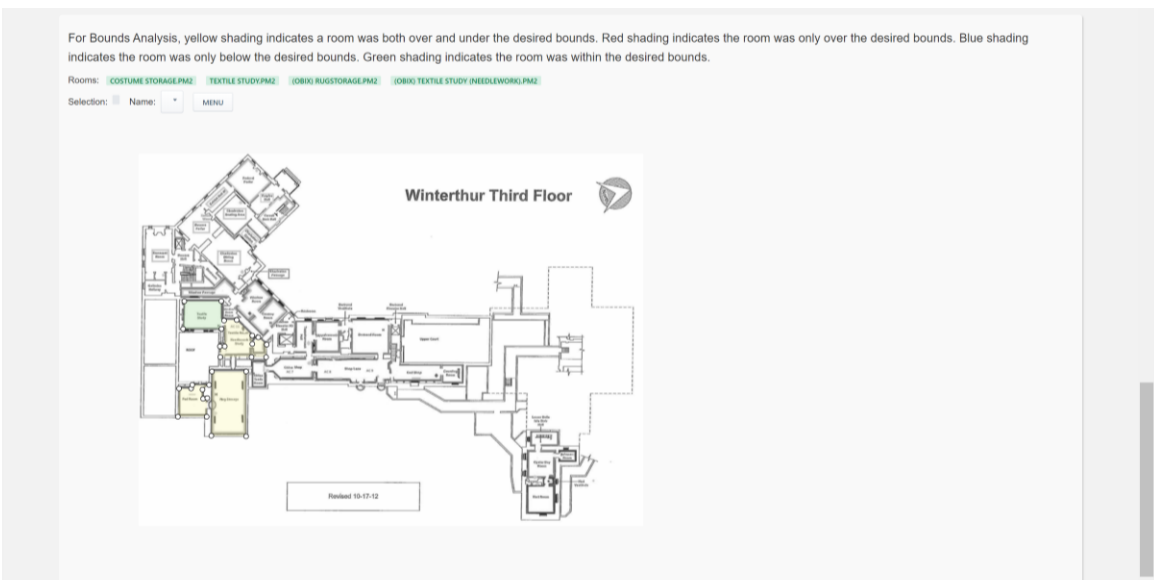

Figure 38: Floorplan visualization. Bounds and Swing Analysis for Multiple Files interface.

Finally, the user turns to the floorplan visualizer and is able to draw polygons, and associate polygons with a room name which will then color the polygon the appropriate color (Figure 38). The floorplan is visualized using the dash-floorplan component explained in Section 5.4.1. For a Bounds Analysis, yellow shading indicates a room was both over and under the desired bounds. Red shading indicates the room was only over the desired bounds. Blue shading indicates the room was only below the desired bounds. Green shading indicates the room was within the desired bounds. For Swing Analysis, red shading indicates a swing over the maximum desired swing in that room. Green indicates a swing within the desired swing range.

6.3 Cross-Correlation Interface

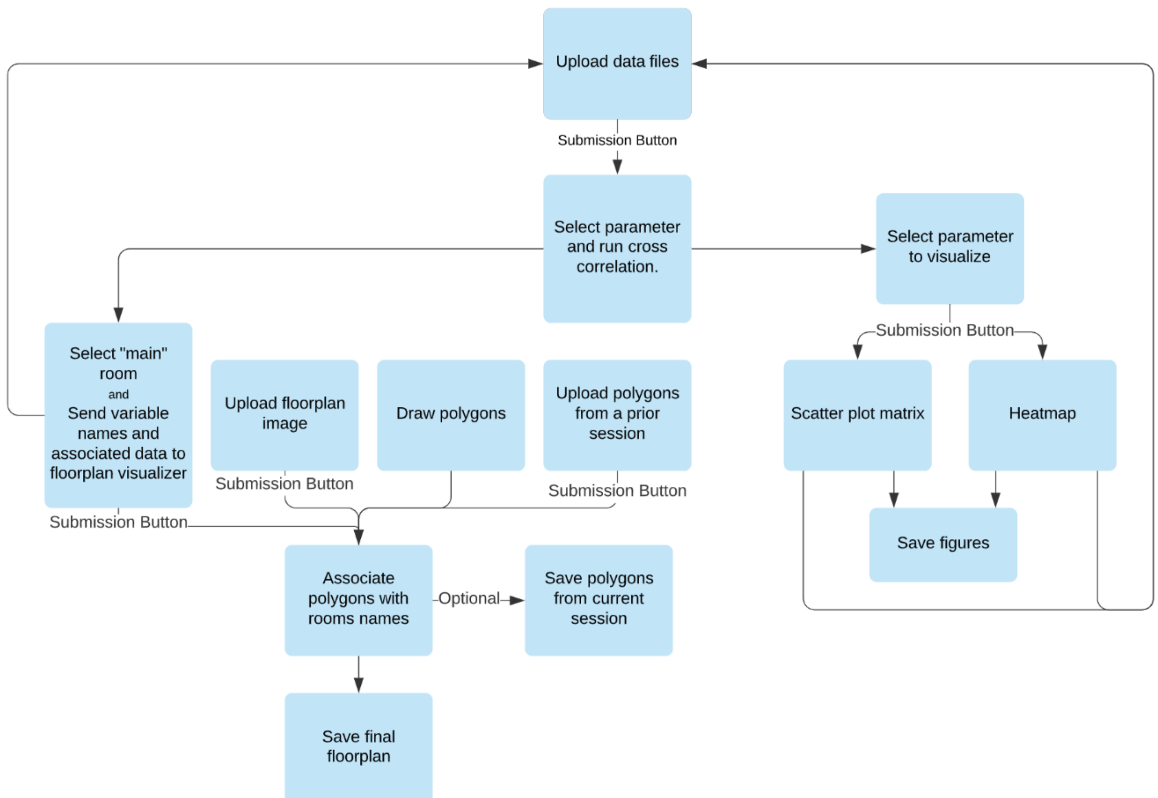

Figure 39: Flow map of Cross-Correlation interface.

Figure 39 outlines the Cross-Correlation interface. To begin, the user uploads the .pm2 data files that they would like to visualize, seen in Figure 40. The user then selects whether they want to examine temperature correlation or relative humidity correlation, and then clicks the “Run Correlation” button to process the data, seen in Figure 41.

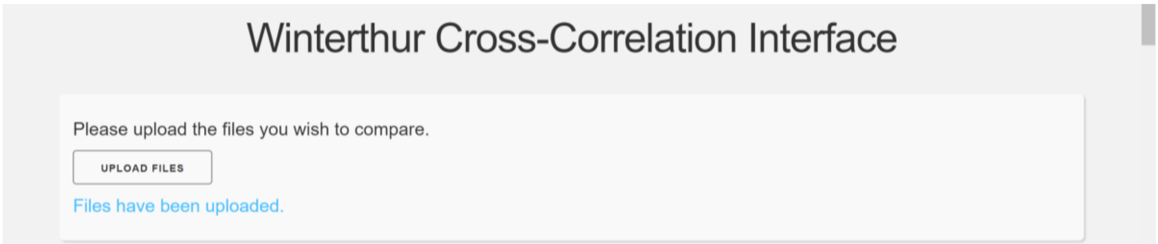

Figure 40: Upload data. Cross-Correlation Analysis interface.

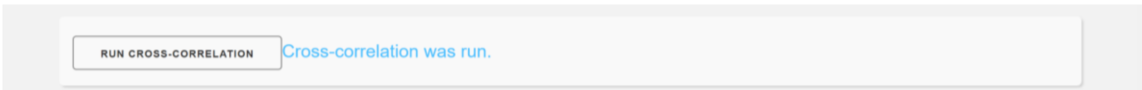

Figure 41: Run cross-correlation analysis. Cross-Correlation Analysis interface.

The user can select what graphs to visualize (Figure 42). The user has three options for ways to visualize correlation data: scatter plot matrix, heatmap, and floorplan visualization. The scatter plot matrix and heatmap are discussed in section 5.4.3. The heatmap shows values when it is hovered over, as shown in Figure 43. An example scatter plot matrix is shown in Figure 44.

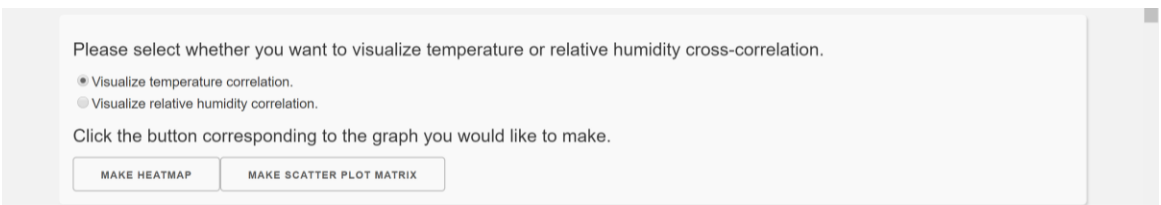

Figure 42: Graph visualization submission. Cross-Correlation Analysis interface.

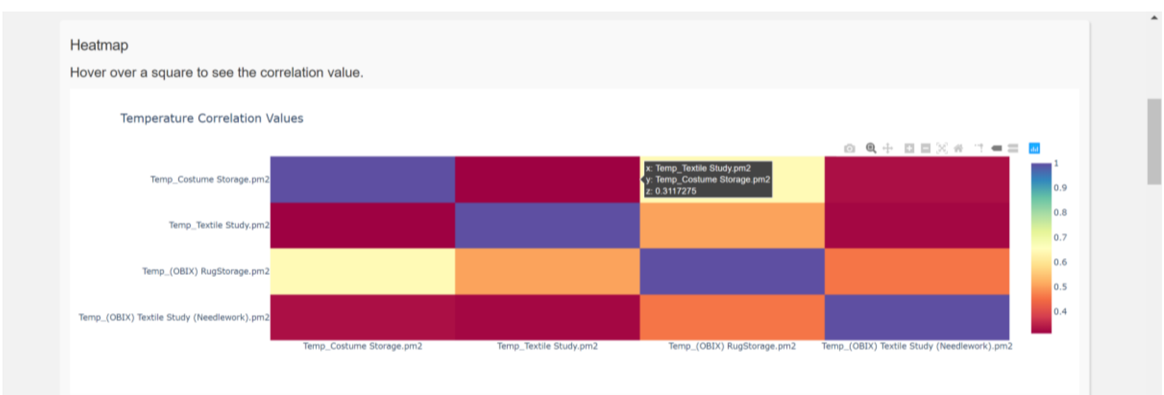

Figure 43: Heatmap. Note the box that appears when the heatmap is hovered over with the mouse. Cross-Correlation Analysis interface.

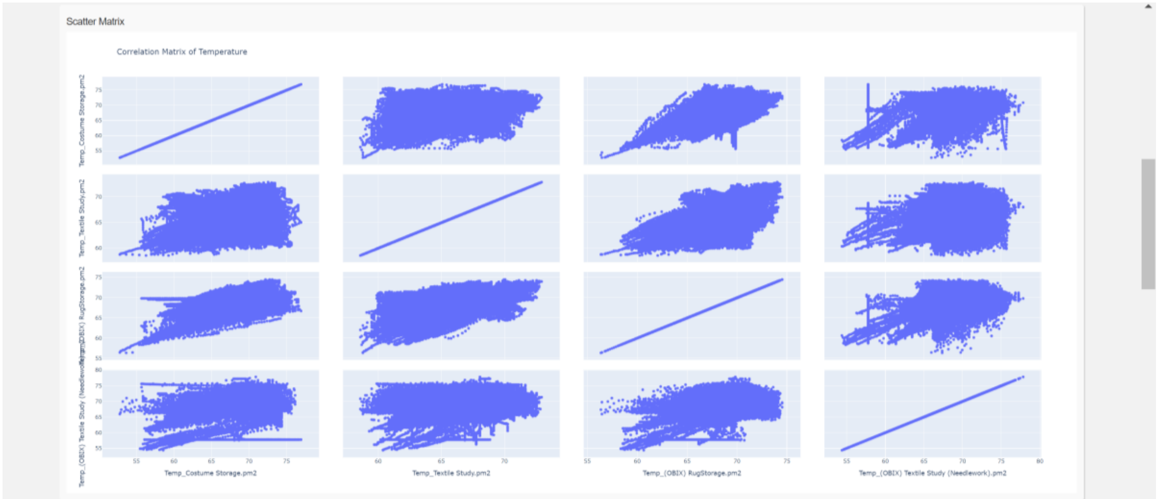

Figure 44: Scatter plot matrix. Cross-Correlation Analysis interface.

Next, the user can turn to visualization on a floorplan, which functions in the same way as described in Section 6.2. However, the user must also select which room is the "main" room to which all other rooms are compared before sending data to the visualizer, shown in Figure 45.

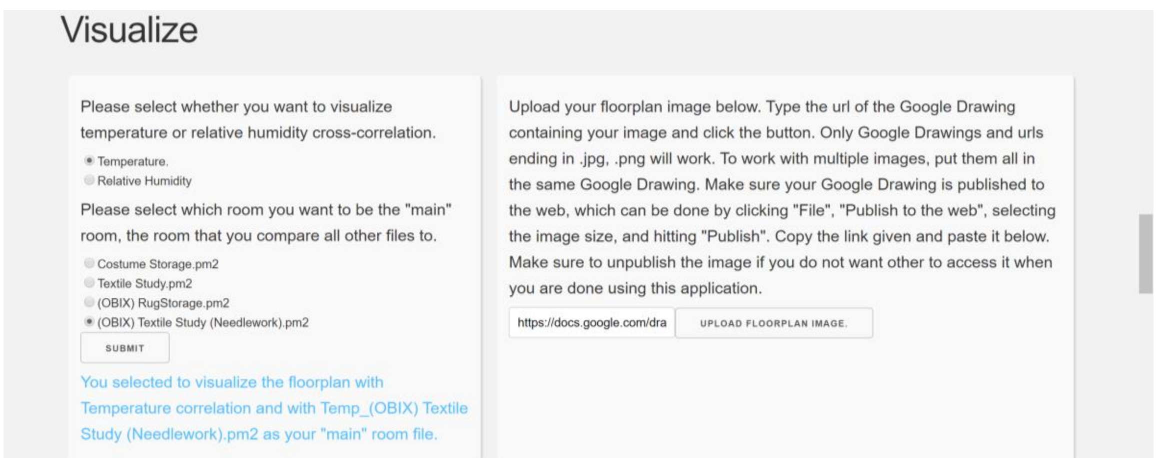

Figure 45: Send data and image to visualizer. Select "main" room and parameter. Cross-Correlation Analysis interface.

For cross-correlation, the color of the polygons is determined by the correlation of the room associated with the polygon to the room selected as the "main" room. A helpful color scale is located at the bottom of the floorplan (Figure 46).

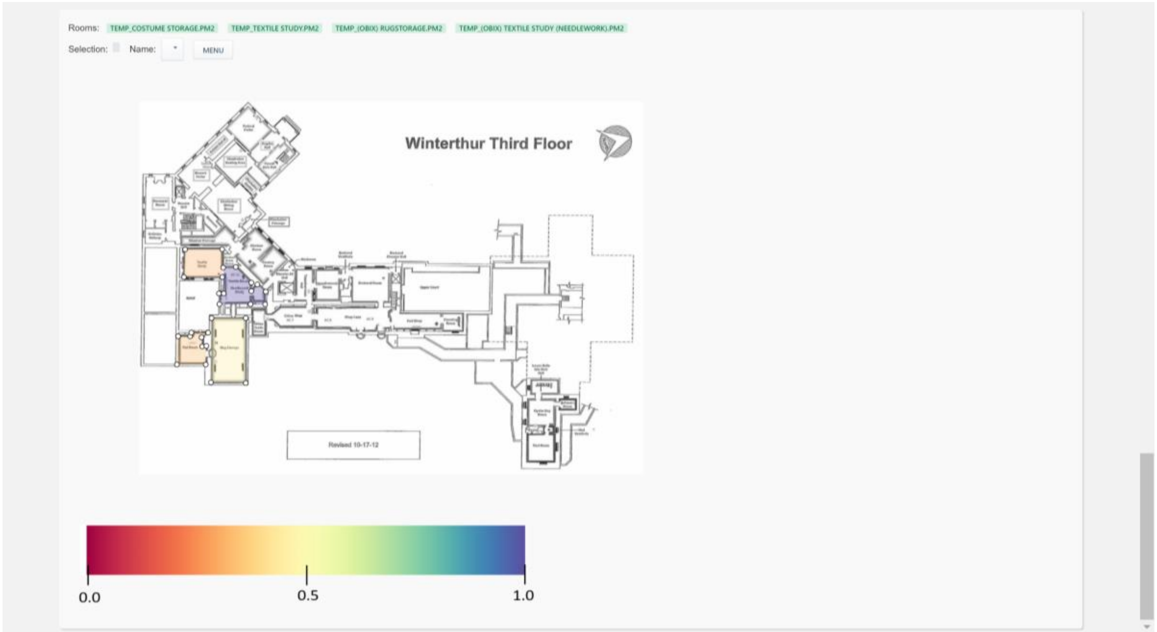

Figure 46: Floorplan visualization. Cross-Correlation Analysis interface.

6.4 Factor Analysis Interface

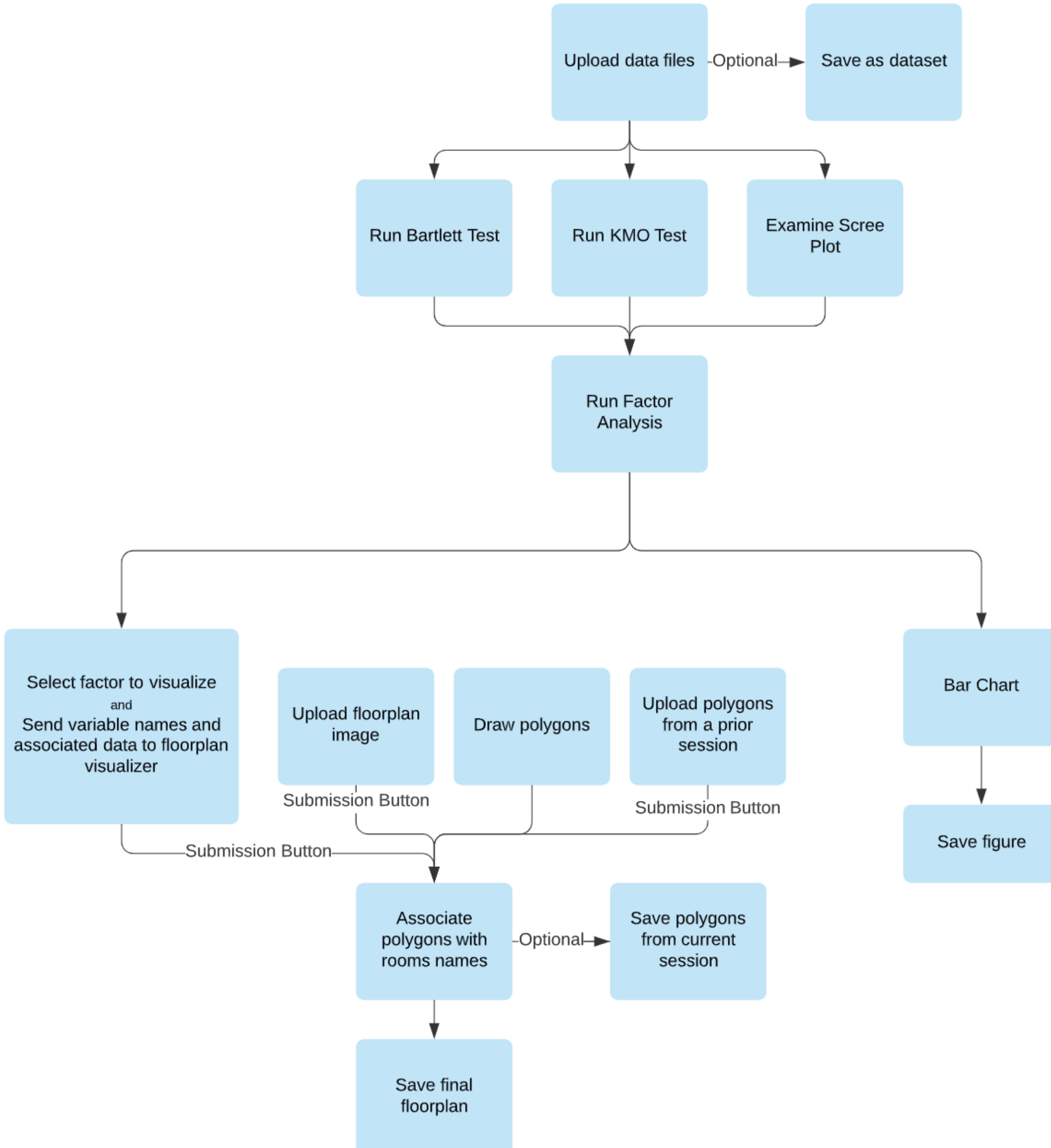

Figure 47: Flow map through Factor Analysis interface.

Figure 47 outlines the Factor Analysis interface. To begin, the user selects whether they need to upload .pm2 files or whether they already have a saved file in .pickle or .csv format from a prior session to upload. If the former, the user uploads the .pm2 files to be used. If the latter, the user uploads the .pickle or .csv file. The names of the uploaded files are printed, allowing the user to at once confirm the desired files have been uploaded. If uploading .pm2 files, the user will be presented with the option to save those files as a .pickle or .csv for use in a later session (Figure 48). If uploading a .pickle or .csv file, the user is presented with the option of saving a .pickle file as a .csv and vice versa (Figure 49). Saving of the dataset can be completed at any time, giving the user the flexibility to first determine the dataset is suitable for factor analysis before saving it.

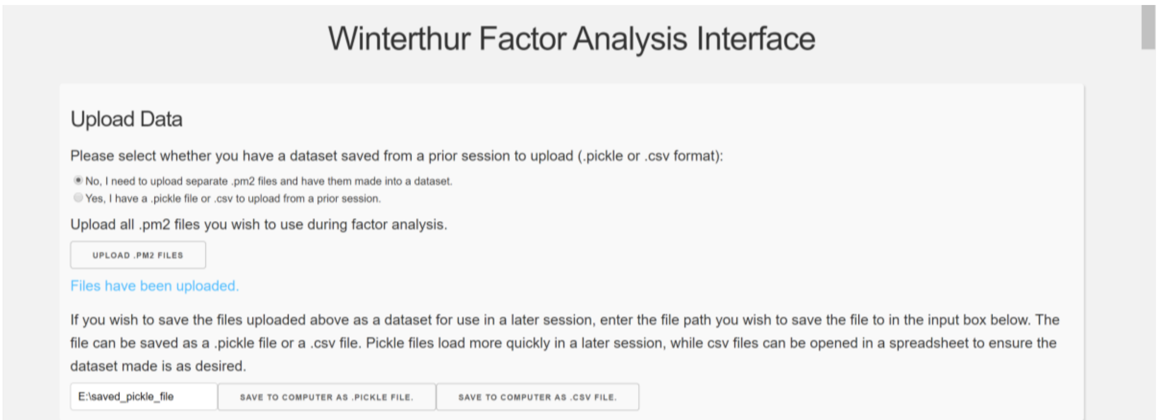

Figure 48: Upload data and save dataset for .pm2 files. Factor Analysis interface.

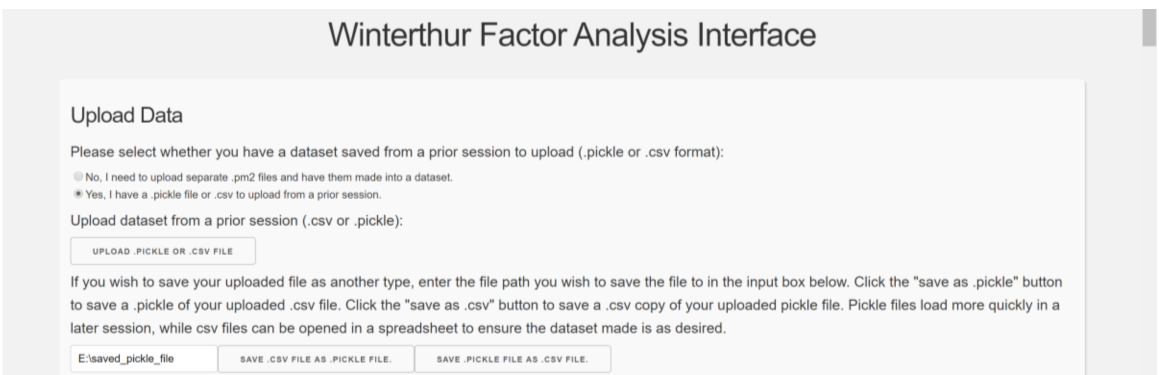

Figure 49: Upload data and save dataset for .csv and .pickle files. Factor Analysis interface.

Next the user checks that the dataset is suitable for factor analysis through the Bartlett Test, KMO Test, and a scree plot of eigenvalues. Usage of these tests is described in Section 5.2.3. The user is directed within the Bartlett and KMO Tests as to whether they should proceed with factor analysis depending on the results of the tests (Figure 50). Figure 51 shows the eigenvalue scree plot, in addition to stating the number of eigenvalues over 1, i.e. the maximum number of potentially meaningful factors.

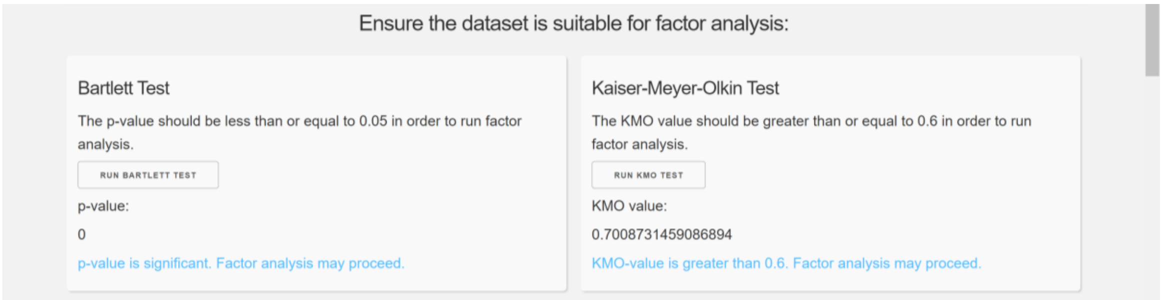

Figure 50: Bartlett Test and KMO Test. Factor Analysis interface.

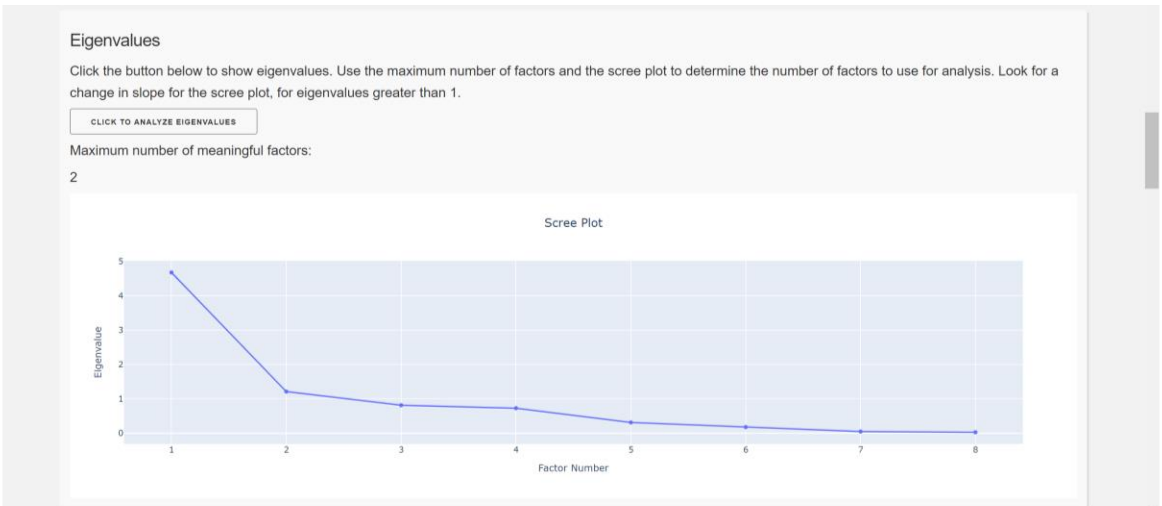

Figure 51: Eigenvalues and scree plot. Factor Analysis interface.

A radio button selection will appear once the user runs the eigenvalue test, from which the user selects the number of factors they want to use for factor analysis based on the results of the suitability tests (Figure 52). For this example, two factors were selected based on the "elbow" in the scree plot discussed in Section 5.2.3. Upon running factor analysis, the user is given several text outputs: the weighting of the parameters associated with each factor, the variance explained by each factor, and the total variance explained by all of the factors. As explained in Section 5.2.3, the total variance can be used to determine whether the number of factors used is suitable for the dataset. The user can visualize the results on a bar chart or on the floorplan. The factor analysis bar chart is explained in Section 5.4.4, and bar charts can be seen for each factor of this example in Figures 53 and 54.

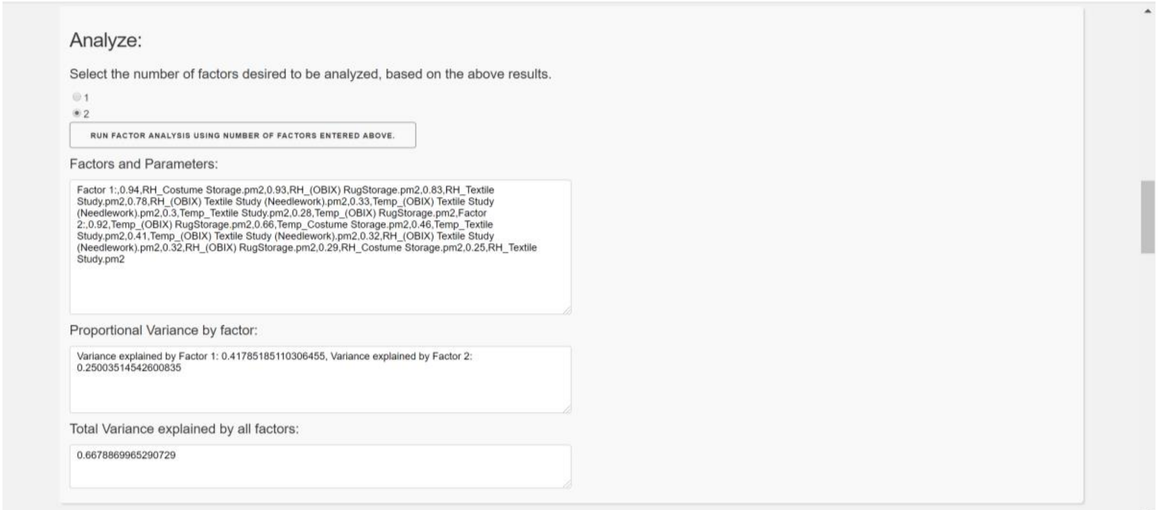

Figure 52: Run factor analysis. Factor Analysis interface.

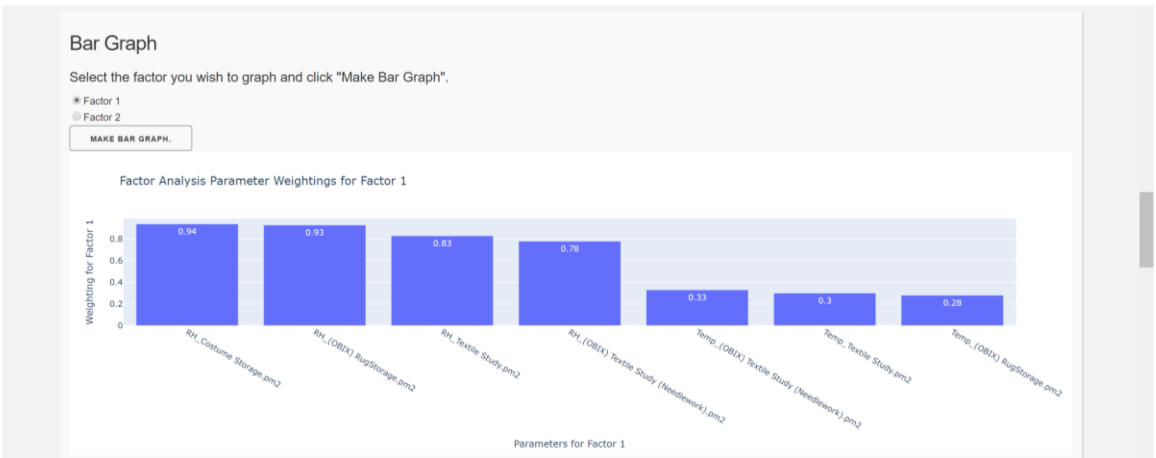

Figure 53: Bar graph for example Factor 1. Factor Analysis interface.

Figure 54: Bar graph for example Factor 2. Factor Analysis interface.

Next the user can turn to visualization on a floorplan, which functions in the same way as described in Section 6.2. The color of the polygons is determined by the weighting of the room associated with the polygon to the factor selected to be visualized. This allows the user to see spatial relationships amongst variables, which is very helpful in enabling the user to give each factor a name. Figure 55 shows the factor analysis floorplan visualization of Factor 1 in this example. Note that some room names are still red, as each room has two parameters, temperature and relative humidity, associated with it. Rather than double up polygons within one room, the parameter with a higher weighting is drawn for a cleaner graph.

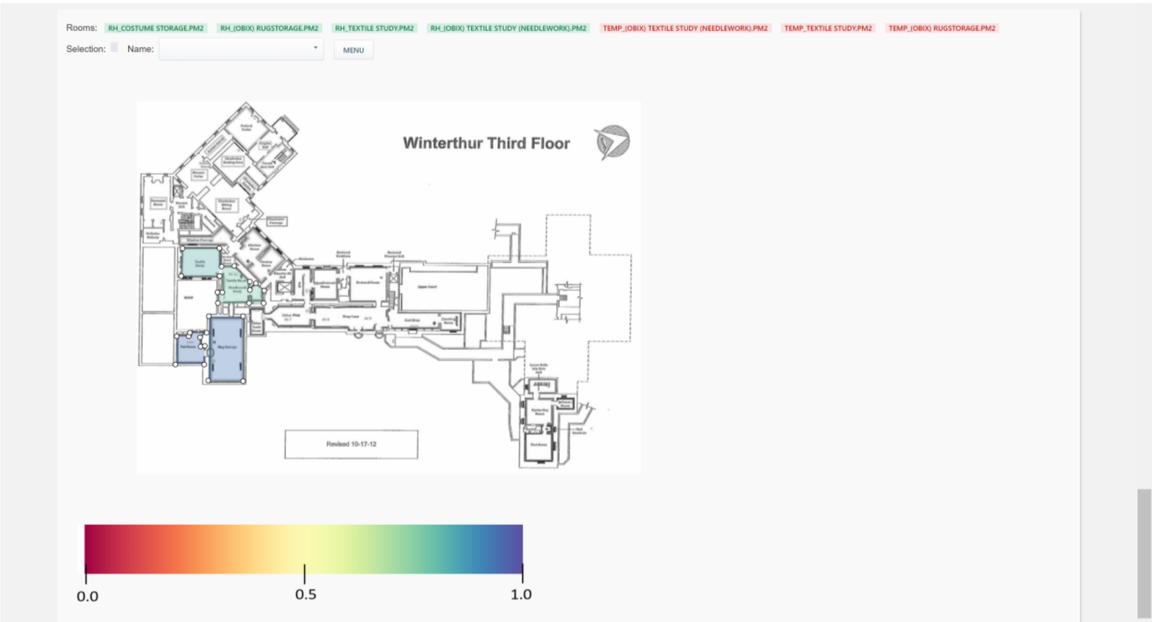

Figure 55: Visualization on floorplan of Factor 1. Factor Analysis interface.

6.5 Human-Machine Interface Design Results

While each interface has a unique layout as seen in the prior sections, it is important to note general principles that were implemented. While a process is running, the browser tab that Dash runs in indicates that the interface is “Updating…” This is seen in Figure 56. After a process has finished, the user is given further indication of what occurred via text outputs for buttons that do not output anything else. These outputs show up in blue text, to further assist the user in identifying what about the interface has changed and distinguish output from what was already there. An example of blue output text can be seen in Figure 50. Regarding polygons, keyboard commands work to copy and delete the polygons in addition to the provided toolbar of options. This makes adjusting polygons even easier for the user, as the toolbar of options does not have to be clicked to duplicate or delete a polygon. A polygon will stay highlighted unless the user clicks on another polygon. To have no polygon shaded to save the image, it is best to make one extra polygon, click on it, and then delete it. That leaves no polygons shaded and provides an optimal final visualization.

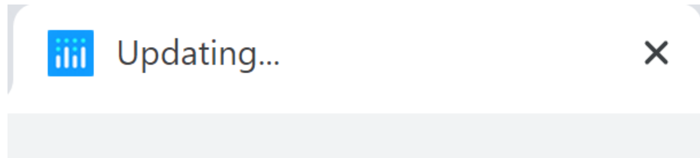

Figure 56: Indication that the interface is running a process.

6.6 Factor Analysis with Exterior Weather Data Results

The prior sections demonstrate how each interface functions. This section shows the results of using the methodology described in Section 5.5, the Factor Analysis interface shown in Section 6.5, and weather data from Section 4.2.

The .pm2 files for the four rooms surrounding the “swimming pool,” described in Section 4.3, and the .pm2 files for the 3 HVAC systems associated with those rooms were uploaded into the Factor Analysis interface. The dataset was saved as a .csv and combined with exterior weather data from Section 4.2 as described in Section 5.5. Factor Analysis was then run for this combined dataset. The Bartlett Test p-value was 0, and the KMO value was 0.85, thus the dataset was suitable for factor analysis.

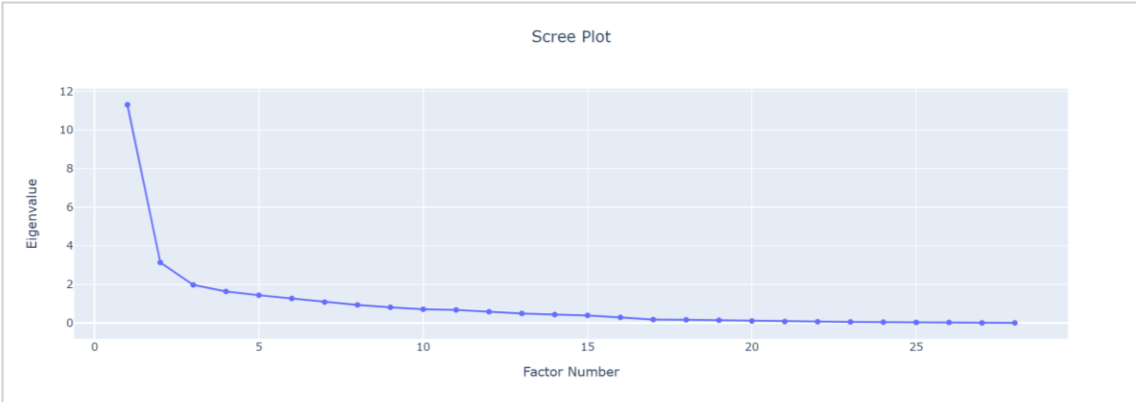

Figure 57: Eigenvalue scree plot for "swimming pool" analysis. Factor Analysis interface.

The scree plot, seen in Figure 57, indicated that 2 to 3 factors would be a good selection for factor analysis. Factor analysis was run with 3 factors to see how much variance the third factor explained. Factor 1 explained 34.2% of variance, Factor 2 explained 14.0% of variance, and Factor 3 explained 6.3% of variance. Thus Factor 3 is not very important and is ignored for this example.

Figure 58: Bar graph of parameter weightings for Factor 1.

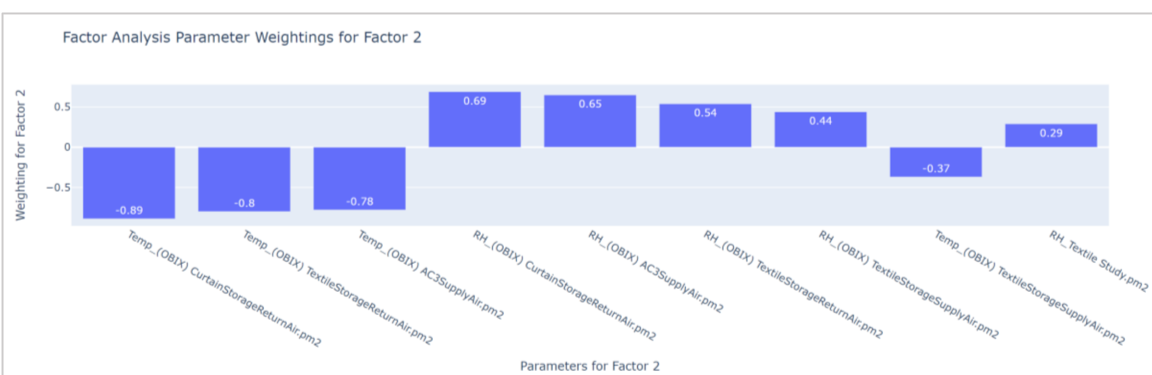

Figure 59: Bar graph of parameter weightings for Factor 2.

Figures 58 and 59 show the bar graphs for the weightings of parameters for Factors 1 and 2. Note that parameters can be negatively correlated to a factor. One sees that the relative humidity of the four rooms being examined as well as the exterior dew point, dry bulb temperature, and wet bulb temperature are all closely related to Factor 1. This indicates that exterior weather is closely linked with the interior room parameters. The temperature of HVAC system air is inversely associated with Factor 2 while the relative humidity of the HVAC system air is positively associated with Factor 2. The relationship does not depend on whether it is HVAC return or supply air. This indicates that Factor 2 pertains to HVAC air.

Figure 60: Visualization on floorplan of "swimming pool" factor analysis results for Factor 1.

Finally, the results are visualized on a floorplan, shown in Figure 60 for Factor 1. All polygons are labeled by the user after saving the image to aid understanding. Note the usage of polygons to represent exterior parameters not visible on the floorplan. Rooms are split to show weightings for both temperature and for relative humidity. This visualization aims to show spatial relationships, but the HVAC systems, which run throughout several rooms at once, are more difficult to visualize; thus, they are shown on the side along with exterior weather data. From this visualization, conservators can quickly see the weightings of parameters associated with Factor 1.

6.7 General Discussion

Each of the interfaces produced by this thesis facilitates better conservation at Winterthur. The Bounds and Swing Analysis for One File interface allows the conservator to examine individual files and thus check on the control of specific rooms. The Bounds and Swing Analysis for Multiple Files interface allows conservators to check groups of rooms and visualize comparisons. The Cross-Correlation interface allows conservators to understand how closely related rooms are. Used in conjunction, the Cross-Correlation interface and the Bounds and Swing Analysis for Multiple Files interface give the conservator a thorough understanding of the relationships amongst rooms: both a number for how related two rooms are and an understanding of what similarities they share. The Factor Analysis interface presents a novel opportunity for Winterthur's conservators to identify underlying factors relating rooms and other parameters, such as exterior weather data.

Interface design is successful in many ways. The choice of Dash for interface development means that deployment of the interface is easy and can be done remotely, as it is a browser-based system. The layout style of the interfaces is sleek and uncrowded. The user can see steps grouped together in bounding boxes and understand what actions generate specific outputs, with informative blue text providing feedback. Interfaces are separated into four separate interfaces which allows the user to know exactly which interface they are using, eliminating confusion. As well, this speeds up processes as the user will not be attempting to run several interfaces at once.

This work seeks to balance user control of the interface while not allowing too much freedom. For instance, the user can enter whatever bounds they desire for a bounds analysis, but only numbers may input, subtly guiding the user to enter numbers. As well, flexibility to re-run processes ensures the user can be as efficient as possible when working with the interface. However, interface design can continue to be developed. The flow of how to navigate the interfaces is top to bottom, and while that might be intuitive for some users, more can be done to ensure there is no confusion. The biggest source of confusion for users is likely to be understanding what additional processes must be re-run to change outputs from one process. Future work includes making changes to the interface layout after conservators identify frequent areas of confusion and additional desired features.

Visualization results provide conservators with ample material to better understand the microclimate system at Winterthur. Times series, bar charts, heatmaps, scatter plot matrices, and visualization on a floorplan were all implemented to present the results of the analyses used in the interfaces. All of these facilitate a better understanding for a diverse audience of viewers, and are customizable, using the toolbar provided in the interface. Conservators can now efficiently produce up-to-date visuals of microclimate monitoring data from Winterthur. However, the color scale in cross-correlation heatmaps automatically adjusts to the highest and lowest correlation values; this makes comparing heatmaps from different sessions difficult as their color scales can be different. A constant color scale would ensure that heatmaps are comparable across analysis sessions.

The results in Section 6.6 showcase how the Factor Analysis interface can be used to confirm intuitions. Factor analysis of the rooms surrounding the “swimming pool” reveals that exterior weather is highly associated with those rooms. This example demonstrates how the interface can be used by conservators to better understand issues such as those laid out in Section 4.3.

In sum, this thesis completed its objectives. Intuitive human-machine interfaces were designed for Winterthur to translate the results of analyses of its microclimate monitoring data to a diverse audience. Bounds and Swing Analysis interfaces show basic relationships within and between datasets while the Cross-Correlation interface allows conservators to understand what sensors and rooms are highly correlated with each other. This thesis implemented factor analysis to highlight trends between monitored variables and other parameters. This thesis develops new methods of visualization, specifically visualization on a floorplan, to highlight the spatial relationships between microclimate monitoring variables. Finally, this methodology has been formalized so that it can be applied to other monitoring situations, such as other museums or historic structures.

Chapter 7: Conclusions and Future Work

7.1 Conclusions

This thesis aims to facilitate better conservation efforts at Winterthur Museum, Garden, and Library. Temperature and relative humidity must remain controlled in order to best preserve the museum's important collections. Winterthur is home to 175 microclimate monitoring sensors that collect temperature and relative humidity values every fifteen minutes. Data has been collected in this manner since 2013. Understanding this vast amount of information is vital to conservation efforts but incredibly difficult to do without the proper tools. Four interfaces have been created to enable conservators to analyze data files in various ways and visualize results for easy action.

The Bounds and Swing Analysis for One File interface permits the user to analyze one file at a time, gaining a thorough understanding of how temperature, relative humidity, and dew point are changing over time in a room. The conservator can set the desired range for bounds or swing, ensuring flexibility. Running bounds and swing analyses are the first courses of action in understanding a data file. When a user has run the more complicated analyses in other interfaces, they can return to bounds and swing analyses to better understand those results at the most basic level. Bounds and swing analyses indicate how well microclimate is being controlled within a room.

The Bounds and Swing Analyses for Multiple Files interface analyzes several files at once. This interface allows the user to visualize the bounds and swing analyses of many rooms at once on a floorplan. While analyzing one file at a time allows for greater detail, looking at multiple files reveals relationships, especially when visualized. Identifying relationships enables conservators to determine potential causes of those relationships or lack of them. A lack of a relationship between two rooms that should be closely correlated is just as much of an important result as a relationship between two closely correlated rooms.

The Cross-Correlation interface determines the exact correlation between various files and provides three methods of visualization. This gives conservators several methods to identify and explain relationships between rooms. Heatmaps and scatter plot matrices quickly show the relationships between all rooms, while visualization on the floorplan allows for spatial understandings of the correlations. Conservators are given an exact correlation number for relationships, quantifying relationships observed during bounds and swing analyses.

The Factor Analysis interface facilitates understanding of the underlying relationships between variables by identifying unobserved factors. Conservators who use this interface can determine whether there is a hidden factor relating several variables, such as sunlight exposure or high exterior winds. Action can then be taken to address those underlying variables and improve control of the measured parameters, temperature and relative humidity.

Visualization of the results for all of the analyses utilized in this thesis is very important. Failure to properly understand how temperature and relative humidity are being controlled can lead to faster deterioration of priceless museum collections. Through visualization, quick insights into microclimate control can be made, allowing for the best preservation of museum artifacts. When managing large datasets, such as that of Winterthur's microclimate monitoring sensors, tables or single values are insufficient when it comes to explaining the relationships present. Various traditional graphing methods are used, such as time series, heatmaps, and bar charts, as well as novel methods for visualization on a floorplan. Visualizing the results of analyses on floorplans has not been done before and serves to further conservators' understanding of microclimate control at Winterthur.

Design principles were used to create the layout and structure of the interfaces. Feedback to the user is very important so that the user knows when processes are running and have been completed. It is also important that the user can easily find the feedback, so it was made identifiable using the color blue. Processes are grouped together visually with bounding boxes so that the user is able to intuitively understand what steps go together. The ability for the user to go back and adjust things without having to start from the very beginning is also important in the thesis. Processes are divided up within the code so that users can easily adjust parameters without having to restart, and so that no one process takes too long. The ability of the user to customize the floorplan to their needs is also an important aspect of the interfaces. The user can choose to visualize just a portion of a floorplan or the whole structure; the interfaces allow either as the user can upload any image and draw polygons wherever they like. These interfaces aim to balance customizability and control, guiding the user through processes but also allowing the freedom to adjust an analysis or visual to specific needs.

All of these interfaces work in conjunction, and each fill a unique need in aiding conservators' understanding of what is going on at Winterthur. The methodology of this thesis can be used by other institutions to develop their own interfaces for analyzing and visualizing sensor monitoring data. These institutions could be other conservation institutions that have the same need to manage microclimates in order to preserve museum items, or they could be other sensor monitoring situations that yields large amounts of data, such as structural health monitoring.

Structural health monitoring examines loads, stresses, and displacements on the structure. By being able to put this data in conversation with environmental factors such as weather, as this thesis does, the analyst gains a better understanding of what factors are causing variations. Perhaps the concrete is deflecting the most on the portion of a bridge that is most exposed, thus facing the most weather. The relationships between deflections and weather can be revealed and confirmed through factor analysis and visualized through mapping onto a plan of the bridge. The methods for visualization and the creation of interfaces presented in this thesis are transferable and would allow those who monitor structures to easily see changes and examine relationships, just as a museum conservator does.

The work of this thesis fills an important gap in the existing field of microclimate monitoring. It provides conservators with a versatile toolbox with which to tackle protecting the collections housed at Winterthur Museum, Garden and Library. By moving microclimate monitoring away from merely identifying trends after the fact to recognizing relationships between variables, this thesis enables Winterthur's conservators to take action before problems occur and gain a better understanding of the microclimates they regulate.

7.2 Future Work

There are several areas of future work that should be explored. In general, improvements can continue to be made to the efficiency and layout of the interfaces of this thesis based on conservator feedback. As well, future work should be focused on both improving the methods of this thesis and adapting those methods to other monitoring situations.

The Factor Analysis interface and visualization can be improved in several ways. The process of selecting data to be used within the Factor Analysis interface is not completely foolproof, as it does require selection of files with overlapping date ranges, as analysis cannot occur with empty values. This is due to the nature of factor analysis, but future work should aim to help the user better understand what files can be analyzed together. As well, future work should focus on methods to incorporate data such as the number of windows, the amount of sunlight, or the number of people in a room over time into factor analysis. With the ability to analyze these data sources, factor analysis will become an incredibly powerful tool for identifying what issues are affecting the control of the microclimate in certain rooms. For instance, an identified factor may be highly correlated with sunlight, and thus the temperature of rooms also associated with that factor are perhaps influenced by sunlight. The conservator may note that none of the rooms associated with that factor have curtains, and thus can take action to limit sunlight in those rooms.

Future work should also consider adjusting the analyses performed here. The Parallel Analysis method for determining the number of factors for factor analysis, presented in Section 5.2.3, should be studied further to see if it provides clearer direction for users during Factor Analysis than interpreting a scree plot themselves, as Ledesma et al. suggest [54]. As well, cross-correlation is currently separated into temperature correlation and relative humidity correlation. Future work should include the ability to examine correlation between temperature and relative humidity. For Bounds and Swing

Analysis for Multiple Files, the user should be able to visualize multiple datasets on the same time series graph.

Future studies should investigate the use of cross-correlation results to optimize the number of sensors in a structure. Reducing the number of needed sensors is much more economical that purchasing a microclimate monitoring sensor for every room. Cross-correlation, as laid out prior, will indicate which rooms are always related. A model can then be trained to predict what is happening in Room B based on the data coming from Room A. If this model is confirmed to be successful, then a sensor is only needed in Room A. There is a precedent for building models for sensor optimization such as in Akbarzadeh et al. [58] and Yi et al. [59]. A museum can save money and reduce the amount of data to be worked with through sensor optimization.

Future work should include investigating other ways to incorporate floorplan visualization. A time series visualization video of the whole structure on floorplan throughout a selected time period could reveal daily trends and highlight differences between different sides of the structure. Future work should also include the ability to save a visualized floorplan to the user's computer via a button and to add a title to the image. The final colored floorplan would also look more sophisticated if once a user has drawn a polygon, the circles representing the corners disappeared. As well, if the value associated with a polygon appeared on the polygon, the user would not have to rely on the subjective measure of a color scale to determine the value. A floorplan could highlight the rooms that are involved with each factor and then when the user hovers over the polygon, the weightings of all parameters associated with that room would show as well. Future work should also address how to best show HVAC and exterior weather data, such as in the results in Section 6.6, so that the user does not have to add labeling after the fact.

Future work should seek to utilize the methodology of this thesis for other monitoring situations, whether in a museum context or a structural health monitoring context. This thesis provides an effective method for analyzing and visualizing large datasets and so should be applied to similar situations. This will improve understanding of those monitoring systems and promote action to remedy any issues the results reveal. In the case of museums, this ensures best preservation of collections; for structural health monitoring, the stability of the structure is at stake.

Chapter 8: Appendix

8.1 Python Code

All Python code for the interfaces and analyses can be found at https://github.com/ekeim1/Winterthur-Interfaces.

Chapter 9: References

[1] D. Camuffo, E. Pagan, A. Bernardi, and Becherini, “The impact of heating, lighting and people in re-using historical buildings: a case study,” *Journal of Cultural Heritage*, vol. 5, no. 4, pp. 409–416, Oct. 2004, doi: 10.1016/j.culher.2004.01.005.

[2] E. Lucchi, “Review of preventive conservation in museum buildings,” *Journal of Cultural Heritage*, vol. 29, pp. 180–193, Jan. 2018, doi: 10.1016/j.culher.2017.09.003.

[3] S. P. Corgnati, V. Fabi, and M. Filippi, “A methodology for microclimatic quality evaluation in museums: Application to a temporary exhibit,” *Building and Environment*, vol. 44, no. 6, pp. 1253–1260, Jun. 2009, doi: 10.1016/j.buildenv.2008.09.012.

[4] G. Pavlogeorgatos, “Environmental parameters in museums,” *Building and Environment*, vol. 38, no. 12, pp. 1457–1462, Dec. 2003, doi: 10.1016/S0360-1323(03)00113-6.

[5] L. D. Pereira, A. R. Gaspar, J. J. Costa, and G. Pereira, “The importance of long-term hygrothermal assessment of museum spaces: method and application in a permanent exhibition in a historical building,” 2019. http://revista.arp.org.pt/en/artigos/2018005.html (accessed Oct. 17, 2019).

[6] F. Sciurpi, C. Carletti, G. Cellai, and L. Pierangioli, “Environmental monitoring and microclimatic control strategies in ‘La Specola’ museum of Florence,” *Energy and Buildings*, vol. 95, pp. 190–201, May 2015, doi: 10.1016/j.enbuild.2014.10.061.

[7] D. Camuffo, A. Bernardi, G. Sturaro, and A. Valentino, “The microclimate inside the Pollaiolo and Botticelli rooms in the Uffizi Gallery, Florence,” *Journal of Cultural Heritage*, vol. 3, no. 2, pp. 155–161, Apr. 2002, doi: 10.1016/S1296-2074(02)01171-8.

[8] K. Gysels et al., “Indoor environment and conservation in the Royal Museum of Fine Arts, Antwerp, Belgium,” *Journal of Cultural Heritage*, vol. 5, no. 2, pp. 221–230, Apr. 2004, doi: 10.1016/j.culher.2004.02.002.

[9] D. Camuffo et al., “Environmental monitoring in four European museums,” *Atmospheric Environment*, vol. 35, pp. S127–S140, Jan. 2001, doi: 10.1016/S1352-2310(01)00088-7.

[10] “A Practical Introduction to Factor Analysis: Exploratory Factor Analysis,” *UCLA Institute for Digital Research and Education*, 2019. https://stats.idre.ucla.edu/spss/seminars/introduction-to-factor-analysis/a-practical-introduction-to-factor-analysis/ (accessed Oct. 17, 2019).

[11] N. B. Gallagher, B. M. Wise, and Stewart, “Application of multi-way principal components analysis to nuclear waste storage tank monitoring,” *Computers & Chemical Engineering*, vol. 20, pp. S739–S744, Jan. 1996, doi: 10.1016/0098-1354(96)00131-7.

[12] F.-J. Garcia-Diego and M. Zarzo, “Microclimate monitoring by multivariate statistical control: The renaissance frescoes of the Cathedral of Valencia (Spain),” *Journal of Cultural Heritage*, vol. 11, no. 3, pp. 339–344, Jul. 2010, doi: 10.1016/j.culher.2009.06.002.

[13] M. Zarzo, A. Fernández-Navajas, and F.-J. García-Diego, “Long-term monitoring of fresco paintings in the cathedral of Valencia (Spain) through humidity and

temperature sensors in various locations for preventive conservation," *Sensors*, vol. 11, no. 9, pp. 8685–8710, 2011, doi: 10.3390/s110908685.
[14] A. B. Costello and J. Osborne, "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis," *Practical Assessment, Research, and Evaluation*, vol. 10, p. Article 7, 2005, doi: https://doi.org/10.7275/jyj1-4868.
[15] D. A. Norman, *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books, 2013.
[16] "eClimateNotebook." https://www.eclimatenotebook.com/.
[17] M. King, R. Napolitano, M. Cooper, M. Descroders, J. Wickens, and B. Glišić, "Tool Development for Monitoring Relative Humidity and Temperature in Museum Environments," forthcoming, 2020.
[18] E. R. Tufte, *The Visual Display of Quantitative Information*, Second Edition. Cheshire, CT: Graphics Press, 2001.
[19] H. E. Silva, G. B. A. Coelho, and F. M. A. Henriques, "Climate monitoring in World Heritage List buildings with low-cost data loggers: The case of the Jerónimos Monastery in Lisbon (Portugal)," *Journal of Building Engineering*, vol. 28, p. 101029, Mar. 2020, doi: 10.1016/j.jobe.2019.101029.
[20] J. Valach et al., "Everything is data - Overview of modular system of sensors for museum environment," presented at the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 2015, vol. 40, pp. 439–442, doi: 10.5194/isprsarchives-XL-5-W7-439-2015.
[21] C. Boncaina, P. Baggio, F. Cappelletti, P. Romagnoni, and A. G. Stevan, "The Scrovegni Chapel: The results of over 20 years of indoor climate monitoring," *Energy and Buildings*, vol. 95, pp. 144–152, May 2015, doi: 10.1016/j.enbuild.2014.12.018.
[22] "About Winterthur | Winterthur Museum, Garden & Library." http://www.winterthur.org/visit/about-winterthur/ (accessed Oct. 17, 2019).
[23] "The House within the House: Imagining the Original Winterthur," *Winterthur Museum & Library Blog*, Oct. 19, 2011. http://museumblog.winterthur.org/2011/10/19/the-house-within-the-house-imagining-the-original-winterthur/ (accessed Oct. 17, 2019).
[24] "The House | Winterthur Museum, Garden & Library." http://www.winterthur.org/visit/museum/house/ (accessed Oct. 17, 2019).
[25] "Visions of Grandeur: Winterthur Transformed | Winterthur Museum & Library Blog." http://museumblog.winterthur.org/2014/04/04/visions-of-grandeur-winterthur-transformed/ (accessed Oct. 17, 2019).
[26] J. L. Fairbanks, "The Architectural Development of Winterthur House," *Winterthur Portfolio*, vol. 1, pp. 80–105, 1964.
[27] "Virtual Tour | Winterthur Museum, Garden & Library." http://www.winterthur.org/visit/tours/virtual-tour/ (accessed Oct. 17, 2019).
[28] Detailed floorplans obtained from Winterthur Conservators.
[29] "Winterthur Topo Map in New Castle County Delaware," *TopoZone*. https://www.topozone.com/map-print/ (accessed Oct. 17, 2019).
[30] USGS, "Science Explorer," *USGS*. https://www.usgs.gov/search-map?search=winterthur%20delaware (accessed Oct. 17, 2019).

[31] T. S. of Delaware, “Delaware Geography,” *https://delaware.gov*. https://delaware.gov/topics/facts/geo.shtml (accessed Oct. 17, 2019).
[32] “PEM2 Datalogger,” *Image Permanence Institute*. https://store.imagepermanenceinstitute.org/environmental-monitoring/pem2-datalogger (accessed Apr. 22, 2020).
[33] National Climatic Data Center, “Engineering Weather Data,” *Engineering Weather Data*. http://web.utk.edu/~archinfo/EcoDesign/escurriculum/weather_data/weather_data_summ.html (accessed Oct. 18, 2019).
[34] J. Wickens and M. Cooper, “Winterthur Site Visit,” Jan. 27, 2020. Performed by Elizabeth Keim and Rebecca Napolitano.
[35] “PyCharm: the Python IDE for Professional Developers by JetBrains,” *JetBrains*. https://www.jetbrains.com/pycharm/ (accessed Apr. 14, 2020).
[36] “Dash Documentation & User Guide | Plotly.” https://dash.plotly.com/ (accessed Apr. 14, 2020).
[37] W. Reinhart, *dash-floorplan*. 2020. https://github.com/wfreinhart/dash-floorplan.
[38] “pandas - Python Data Analysis Library,” *pandas*. https://pandas.pydata.org/ (accessed Apr. 14, 2020).
[39] “NumPy,” *NumPy*. https://numpy.org/ (accessed Apr. 14, 2020).
[40] “base64 — Base16, Base32, Base64, Base85 Data Encodings — Python 3.8.2 documentation,” *Python Documentation*, Apr. 13, 2020. https://docs.python.org/3/library/base64.html (accessed Apr. 14, 2020).
[41] “Google Drawings,” *Google*. https://support.google.com/docs/answer/179740?co=GENIE.Platform%3DAndroid&hl=en.
[42] *plotly: dash-sample-apps/apps/dash-oil-and-gas/assets/s1.css*. Plotly, 2020.
[43] W. Kenton, “Cross-Correlation,” *Investopedia*. https://www.investopedia.com/terms/c/crosscorrelation.asp (accessed Oct. 17, 2019).
[44] “pandas.DataFrame.corr,” *pandas 1.0.3 documentation*, 2014 2008. https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.corr.html (accessed Apr. 15, 2020).
[45] M. Rahn, “Factor Analysis: A Short Introduction, Part 1,” *The Analysis Factor*, Sep. 10, 2012. https://www.theanalysisfactor.com/factor-analysis-1-introduction/ (accessed Oct. 17, 2019).
[46] IBM, “KMO and Bartlett’s Test,” *IBM Knowledge Center*. .
[47] “factor_analyzer package — factor_analyzer 0.3.1 documentation.” https://factor-analyzer.readthedocs.io/en/latest/factor_analyzer.html#module-factor_analyzer.factor_analyzer (accessed Apr. 12, 2020).
[48] “Validity of Correlation Matrix and Sample Size,” *Real Statistics*. http://www.real-statistics.com/multivariate-statistics/factor-analysis/validity-of-correlation-matrix-and-sample-size/ (accessed Apr. 14, 2020).
[49] D. Ludecke, D. Makowski, M. S. Ben-Shachar, I. Patil, and S. Hojsgaard, “Structural Models (EFA, CFA, SEM...),” *github*. https://easystats.github.io/parameters/articles/efa_cfa.html (accessed Apr. 14, 2020).

[50] B. A. Cerny and H. F. Kaiser, “A Study of A Measure of Sampling Adequacy For Factor-Analytic Correlation Matrices,” *Multivariate Behavioral Research*, vol. 12, no. 1, pp. 43–47, 1977, doi: 10.1207/s15327906mbr1201_3.
[51] R. H. Hoyle and J. Duvall, “Determining the number of factors in exploratory and confirmatory factor analysis,” *ResearchGate*, Dec. 2003. https://www.researchgate.net/publication/245904294_Determining_the_number_of_factors_in_exploratory_and_confirmatory_factor_analysis (accessed Apr. 14, 2020).
[52] R. B. Cattell, “The Scree Test For The Number Of Factors,” *Multivariate Behavioral Research*, vol. 1, no. 2, pp. 245–276, Apr. 1966, doi: 10.1207/s15327906mbr0102_10.
[53] M. Rahn, “Factor Analysis: A Short Introduction, Part 4-How many factors should I find?,” *The Analysis Factor*, Nov. 16, 2012. https://www.theanalysisfactor.com/factor-analysis-how-many-factors/ (accessed Apr. 14, 2020).
[54] R. D. Ledesma, P. M. Valero-Mora, and G. Macbeth, “The Scree Test and the Number of Factors: a Dynamic Graphics Approach,” *Spanish Journal of Psychology*, vol. 18, no. 11, pp. 1–10, Jun. 2015, doi: 10.1017/sjp.2015.13.
[55] H. Abdi, “Factor Rotation in Factor Analyses,” in *Encyclopedia of Social Sciences Research Methods*, M. Lewis-Beck, A. Bryman, and T. Futing, Eds. Thousand Oaks, CA: Sage, 2003.
[56] R. Napolitano and B. Glišić, “Open-source software for diagnostics of existing masonry structures using physics-based modeling, point clouds, and machine learning tools,” *Princeton University*, Aug. 2019.
[57] “Spreadsheet Software - Microsoft Excel,” *Office*. https://products.office.com/en-us/excel (accessed Apr. 17, 2020).
[58] V. Akbarzadeh, C. Gagne, M. Parizeau, M. Argany, and M. A. Mostafavi, “Probabilistic Sensing Model for Sensor Placement Optimization Based on Line-of-Sight Coverage,” *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 2, pp. 293–303, Feb. 2013, doi: 10.1109/TIM.2012.2214952.
[59] T.-H. Yi, H.-N. Li, and M. Gu, “Optimal sensor placement for structural health monitoring based on multiple optimization strategies,” *The Structural Design of Tall and Special Buildings*, vol. 20, no. 7, pp. 881–900, 2011, doi: 10.1002/tal.712.