SYSTEMS ANALYTICS APPROACH USING WIRELESS SENSOR NETWORK TECHNOLOGIES AND BIG DATA VISUALIZATION FOR CONTINUOUS ASSESSMENT OF AIR QUALITY IN A WORKPLACE ENVIRONMENT

Judith Molka-Danielsen
Molde University College, Norway
j.molka-danielsen@himolde.no

Veronika Olešnaníková
University of Žilina, Slovakia
veronika.olesnanikova@gmail.com

Peter Šarafín
University of Žilina, Slovakia
peter.sarafin@gmail.com

Róbert Žalman
University of Žilina, Slovakia
robozalman@gmail.com

Per Engelseth
Molde University College, Norway
Per.Engelseth@himolde.no

Abstract

The indoor air quality in workplace buildings, e.g. air temperature, humidity and levels of carbon dioxide (CO2), play a critical role in the perceived levels of workers' comfort and in reported medical symptoms. CO2 can act as an oxygen displacer, and in confined spaces humans can have, for example, reactions of dizziness, increased heart rate and blood pressure, headaches, and in more serious cases loss of consciousness. Specialized organizations can be brought in to monitor the work environment for limited periods. However, new low cost wireless sensor network (WSN) technologies offer potential for more continuous assessment of workplace air quality. Central to effective decision making is the systems analytics approach and visualization of the big data, as collected through WSN for control of the workplace environment. This paper presents a case study of a systems analytics approach for monitoring and visualization of air quality using big data as collected with WSN technologies. The case workplace environment is a university lecture hall. This small case study demonstrates a proof of concept that the monitoring and visualization approach has added value for decision makers. In addition, this study discusses the implications for alternative visualization approaches of BD. We also identify other potential applications of WSN technologies and visualization of BD in other workplace environments; for example, for worker safety in high risk industries (e.g. oil and gas industry, carbonated beverage workers, etc.) and for quality of goods in supply chain management.

Please cite as: Molka-Danielsen, J., Olešnaníková, V., Šarafín, P., Žalman, R. & Engelseth, P.: Systems Analytics Approach Using Wireless Sensor Network Technologies and Big Data Visualization for Continuous Assessment of Air Quality in a Workplace Environment (2016). Paper presented at NOKOBIT 2016, Bergen, 28-30 Nov. *NOKOBIT*, vol. 24, no. 1, Bibsys Open Journal Systems, ISSN 1894-7719.

1. INTRODUCTION

Indoor air quality is important for worker satisfaction, safety and health. Measures of air quality that are used in general assessment of air quality include: CO2 levels, temperature, and relative humidity. Specialized organizations can be brought in to measure air quality in designated test sites for limited periods. However, there can be variability in measurements due to season or periodic organization activities. Our research is a small, proof of concept, case study for a systems analytics of "Big Data" (BD) for determining air quality in a university lecture hall. We explore a visualization approach that can be replicated in other workplace settings. Recent low cost wireless sensor network (WSN) technologies give opportunities for organizations to set up for longer-term or continuous air quality measurement of workplace spaces. The small size and self-sufficient characteristics of the sensor devices allow for their flexible placement. They can be used, for example, to measure the air quality in a redesigned building, in an area of reported problems, or during a period of increased industrial activities. The research problem addressed in this study is to gain understanding of how low cost WSN technologies and visualization approaches of BD can be used to better monitor and with the objective to improve air quality in workplace buildings.

In this study we classify the continuous monitoring of air quality data in the lecture hall as an application of BD. We characterize the BD by degree of structure (or variety), volume and velocity. The BD collected in this study by the WSN sensors is highly structured (of low degree of variety) in that data is collected internally by the organization (e.g. not by social media for this case) and data quality measures can be compared to national and international quality standards. The data volume is dependent on the length of data collection period (e.g. 4-days). The characteristic of velocity means how fast the data is produced and how fast the data must be processed to make decisions. The collection velocity is based on our predetermined interval of measurement (e.g. 1 data point per collection factor per minute). The BD is of value if it can be presented in such a way as to contribute to intelligence analytics in a decision point. In our findings section we provide a historical graphical presentation of the measured data for the study period.

We point out that there are other potential applications of WSN technologies and visualization of BD in other workplace environments beyond this small study. For example WSN and visualization can be used for worker safety in high risk industries (e.g. oil and gas industry, carbonated beverage workers, etc.) and for quality of goods in supply chain management. In such environments it would be ideal to also view the graphical presentation of the data in real-time, such that in other workplace environments (e.g. high risk industries) BD could be monitored for spikes in air quality measurements.

In this case study the time-frame for decision making about air quality in university lecture hall can take place a week after a data-collection period is completed. The interpretation of the BD can result in a decision to change in ventilation system settings. Alternatively, in a high risk industry, the time-frame for decision making perhaps must be made within minutes. The analysis made in a short time-frame can result in a more immediate decision to evacuate a building. In the discussion section of our paper we elaborate further on the design considerations for visualization of the BD. This paper concludes with suggestions for further research.

2. LITERATURE REVIEW

2.1. Measurements Of Air Quality

Various indoor air quality studies in Norway have examined the air quality in modern energy efficient buildings. The parameters measured have included air temperature, relative humidity, and CO2 concentration. Some have also measured particulate matter concentration (Heiselberg, 2016; Norbäck, Nordström & Zhao, 2013). Earlier studies have shown that office workers frequently report health issues of fatigue, heavy-headedness, and eye irritations (Skyberg and Skuleberg, 2003). Norbäck et al (2013) state that in some buildings the ventilation systems are controlled manually (constant flow), and in other newer energy efficient buildings ventilation systems can be controlled by temperature and CO2 demand-controlled sensors (variable flow) systems. In such buildings sensors are used to measure temperature and

CO_2, and the ventilation flow is increased when CO_2 levels rise above 899 ppm, or 22.0 °C. The Norwegian Work and Environment Act (Arbeidstilsynet, 2012) recommends that the standard value upper limit for CO_2 concentration in an indoor environment of 1000 ppm (parts per million). Indoor temperatures should be maintained between 20.0 - 22.0 °C. Indoor temperatures above 22.0 °C are reported to be unpleasant and that working efficiency falls with higher temperatures. Relative humidity indoors will change depending on the relative humidity outdoors. Ideally it should be between 30% - 60% relative humidity indoors. However, many buildings do not try to adjust humidity (by either adding moisture or dehumidification) (Mycoteam, 2015).

There are some research studies that examine relationships between workplace air quality and worker satisfaction or performance. One study conducted in Norway by Solberg, Cao, Alonso and Burud (2016), finds that low relative humidity in recently built airtight and auto-mechanically ventilated buildings, is associated with lower performance of office workers. Another study by Shahzad, Brennan, Theodossopoulos, Hughes, and Calautit (2015) examines energy efficiency and user comfort in comparing Norwegian cellular offices versus British open plan workplaces. In that study workers in the Norwegian offices had control over the thermal system, and reported higher comfort and satisfaction as compared to the ratings by the British workers. In the British workplace the workers could not open windows, and were in airtight and auto-mechanically controlled environment. The British workplace reported lower energy consumption. These studies highlight the point that building management of ventilation can have an effect on worker performance or satisfaction.

A prior study conducted at Molde University College (HiMolde) of the air quality of five selected offices was conducted in 2015 by Mycoteam AS (2015). Mycoteam AS provides professional assessment of indoor air quality. They carry out inspections and analysis of buildings, and provide reports that can be a basis for repair of existing damage, or can help to understand the cause of damages and can offer advice to prevent future problems. The measures at HiMolde included registration of CO_2, air temperature, and relative humidity. Among the sampled offices were four occupied that received reports of "heavy air" and one empty office. The offices were monitored for one week. None of the offices exhibited measures outside of the acceptable limits. The CO_2 measures were within the range of 400-500 ppm for the occupied offices. In several of the offices an upper limit of 23.0 °C and in one 24.5 °C was noted. The workers' report of "heavy air" may be related to the temperature. The University Personnel Director at HiMolde noted that they could only sample a very limited area, and had expressed a wish to test other areas. In particular, they wished to test the air quality in a much used lecture room. This request motivated this study to develop an alternative low cost approach to monitoring and flexible analysis of work-space air-quality in the university.

2.2. Wireless Sensor Network Technology

There are two approaches to detect CO_2 using sensors. The first one is measurement using method of wavelength absorption, which is one of the properties of chemical compound. This method is called NDIR (not dispersive infrared) (Garcia-Romeo, et al, 2012). The second method is based on changes of electrical charge of chemical reaction measuring. This reaction is a result of air contact (CO_2 particles) with particles in the sensor. The most used detection method is using NDIR sensors (Jonqwon Kwon, et al, 2009), but the price of sensors is quite high (between 100 and 1000 euro). In this study, we use the second method, MQ-135 sensor, which provides monitoring of air quality and with suitable settings it can be used as sensor for CO_2 detection.

The MQ-135 sensor is designed for the use in air quality control equipment for buildings / offices. It is suitable for detection of NH_3, NO_3, alcohol, Benzene, smoke and CO_2 (Furturlec, 2016).

The sensor units used in this study are depicted in Figure 1. All sensors can function indoors or outdoors. The functions of the different units are listed as follows:

- gas detector: measuring CO_2 and NO_3 in given area
- temperature, relative humidity and pressure measurement

Figure 1. WSN sensor units

Other types of sensors that are not used in this case are acoustic emission detectors. These for example can measure the intensity of noise in the workplace.

2.3. Big Data

The concept of "big data" was introduced by the META Group (Laney, 2001) where they discussed how enterprises would use data to increase business opportunities. They characterized data by "volume" (how much), "velocity" (speed of data into and out of the organization) and "variety" (range of data sources or data types). These became known as the 3Vs of big data. Other "Vs" have been studied, such as "veracity" (quality of data source) and "variability" (inconsistency of data source).

Not restricting themselves to words starting in "V", in a research study by Rozados and Tjahjono (2014), they produced a list of 52 mainstream sources of Big Data (BD) across supply chains. This data varies by degree of structure, volume and velocity. The dimension of 'structured-ness' indicates three types of data: core transactional data, internal systems data, and other data. They indicated that greater enterprise resources are generally spent on the preparation, management and use of the core transactional data. However, there is little empirical evidence regarding the effect of management and use of the unstructured big data in SCM. Unstructured BD can be sourced in large volume and rapidly acquired from GPS-enabled or social-media-enabled applications. Recent approaches in visualization and geo-analytics have allowed data generated outside of the main enterprise to contribute to knowledge-sharing activities, collaboration and integration of the supply chain network. Traditionally there has been a trade-off, that by increasing structuration, data becomes more applicable for internal use by a firm. Yet big data involves a layered continuum picture from largely controllable internal data, through intermediately controllable supply chain data, and to less controllable environmental data. While we only examine internal data generated within the firm for this study, future research is needed to develop approaches for the intelligent analysis of BD that are potentially generated at all parts of a supply chain.

Hilbert (2016, p.139) points out that the full name of BD is "Big Data Analytics" because the promises of BD is not about only dealing with data in greater quality and quantity and from greater variety of sources, but to be able to use the presentation of the data for intelligent analysis and decision making. Additionally, Kamran and Haq (2013) point out, the new role of data in SCM is to change data into actionable information within the role of business analytics and use software such as Google Refine to prepare data and Circos Software to represent SCM data. In general, for data to be useful at decision making points, such as in planning and forecasting, it must leverage the humans' visualization system and their ability to spot patterns, trends and outliers. In the Findings section we present the visualization approach of the BD for the case study and discuss the implications for other workplace applications.

3. METHOD

Our case study approach is initially informed by interviews with: the University Personnel Director, the Technical Management of the teaching rooms, and a representative of the Building Management team (Statsbygg). The personnel director described the need for monitoring of air quality to a greater extent than was achieved by an external organization. She expressed a wish to monitor all lecture rooms. Our study was limited to demonstration of concept, as we had only one example of each monitoring device. The technical management informed us that the ventilation system was not turned on during the quiet period (first data collection period) in the summer. It was on during the successive two measurement periods after semester start. The building management informed us that the ventilation system is controlled automatically using temperature and CO2 sensors to trigger the system. The system is of the type called, demand-controlled sensors (variable flow) systems as described above, and uses a temperature of 22.0 °C to trigger the system on.

Data was collected for one frequently used lecture room that had previous reports of heavy air, according to the personnel director. There are no windows to the external environment located in the lecture room. Measurements of temperature, CO2 and humidity were collected during an empty period (July 30th-31st), two days after opening (August 17th-18th), and during days when used for lectures (August 21st-22nd). Each of the data collection periods were 48 hours. The three measurement periods are also identified as collection sets A, B and C. Data collection during the empty period (set A) was used to gain baseline measures and a graphical representation. The active periods during data collection after semester start were between the periods of 09:00 – 15:00.

Devices were placed indoors on a ledge near the back of the room. They were not be placed close to the air conditioning vents and doors. They were placed at an ideal height of 1,5 – 2 meter above the floor. The devices are self-sufficient. They are powered from 5V source or from the batteries while they are deployed in the outdoor environment. In this case study, the unit was supplied by the adapter from the electrical network all the time.

4. FINDINGS

Baseline measures for the lecture room were collected on July 30th-31st (Set A) and are depicted in Figure 2. The CO2 measures were below 899 ppm, and so well within an acceptable range. Because the Figure 3 and Figure 4 present the graphical presentation of the data collected on August 16th-18th (Set B) and August 21st-22nd (Set C) respectively. The varieties of data are only three data factors for analysis as assessed over time. It is possible to include a solid threshold bar where thresholds are exceeded. However, in Figure 2, the threshold for temperature is continuously exceeded, while the other two factors (CO2 and humidity) never exceed the threshold.

Two software products were used for the graphical presentation of the data – LiveGraph and Highcharts. LiveGraph is used for the real-time graphical visualization. It can automatically re-examine the data file at short time intervals (between 100 Hz and once per hour) and update the graphs on the screen in real-time. Highcharts is a product of Norwegian company Highsoft. It is a charting library written in JavaScript, offering an easy way of adding interactive charts. Besides the common chart types Highcharts currently supports a wide range of advanced chart types (e.g. scatter, angular gauges, area range, area spline range, column range, bubble, funnel, waterfall, meteogram, polar chart, synchronized and multiple axes). Both tools are free to be used for non-commercial purposes.

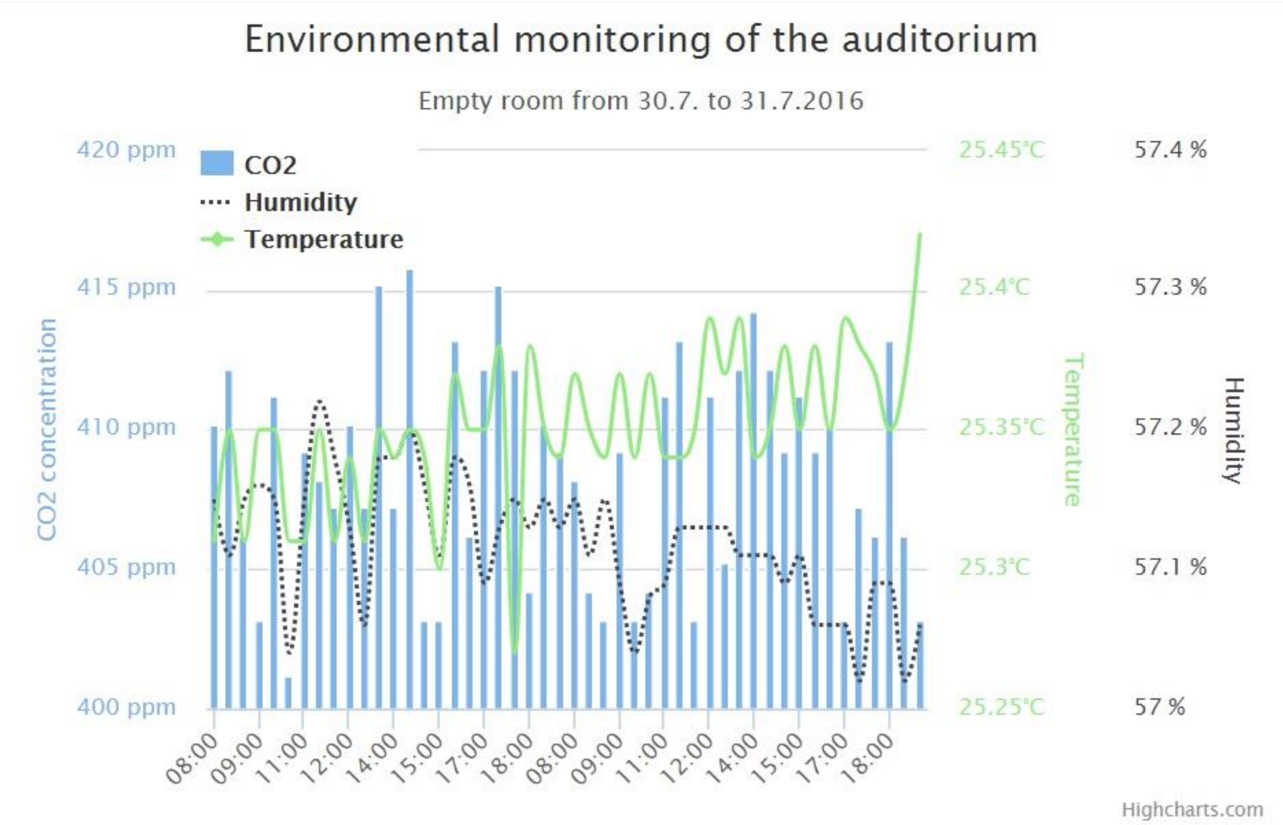

Figure 2. Baseline measures of the air quality in the auditorium during non-use period.

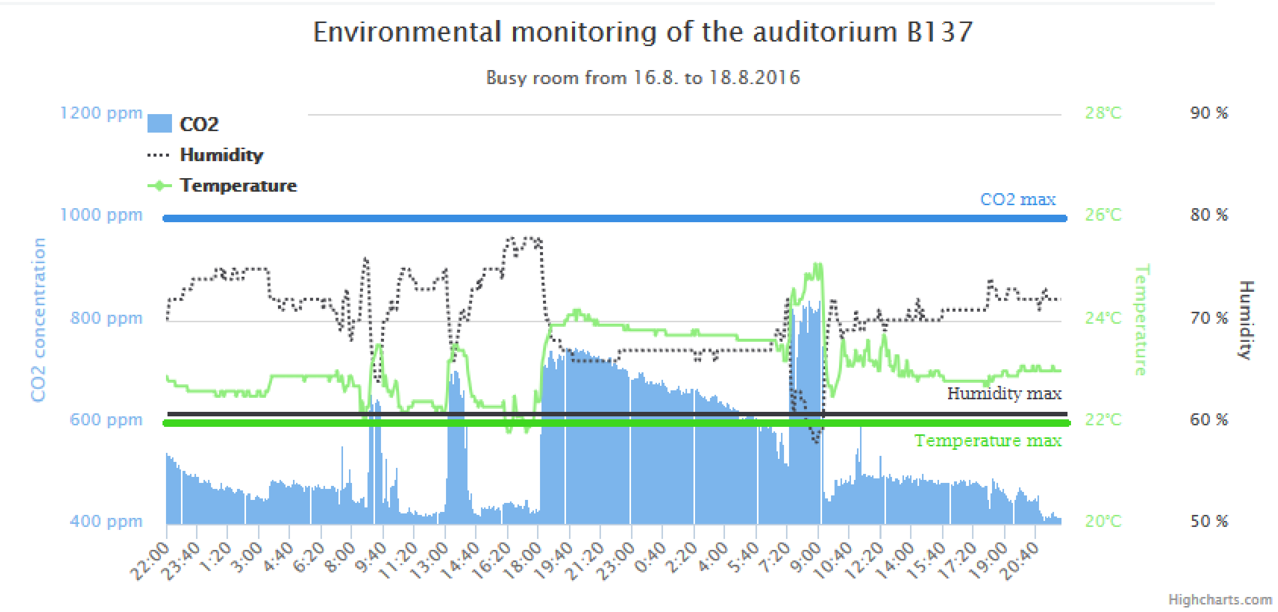

Figure 3. Air quality measures in the auditorium August 16th-18th.

For data set A, the room was not in use during the entire 48 hour period. From data set B, the lecture room was occupied as stated in the Table 1. The auditorium has capacity for 120 persons. The measurements were taken from 16th (10:00 PM) until 18th (10:00 PM) of August.

As the measurement started on 16th of August at 10:00 PM, the temperature, humidity and CO2 level did not change dramatically until 7:30 AM next morning. This can be explained as a fact that nobody visited the auditorium for the long time and air conditioning was turned off. We are informed that at 7:00 AM the air conditioning starts working and it is disabled at 5:00 PM. Obtained results show that there was scheduled to be room-use activity beginning at 9:00 AM, 1:00 PM and 6:00 PM. However, on the 18th, due to personnel meetings, the actual use of the room started from 07:30 AM. Comparing gathered data to the occupancy of the room described in Table 1, we can observe the dependencies. The rise of the CO2 level between 7:00 AM and 6:00 PM is caused by the groups of people accessing the room, and in some cases in addition to the planned use as specified in the time schedule. As the air conditioning is not working after 5:00 PM, the accumulated CO2 level after this time has slowly descending nature.

Date	**Time**	**Occupancy** (persons present / capacity of room)
17th August 2016	13:15 – 14:00	53/120
18th August 2016	07:30 – 15:00	40-50/120
22nd August 2016	09:15 – 15:00	20-40/120

Table 1. Occupancy of the auditorium during the measurements

The activity in the selected room started on 18th of August at 7:30 AM. The obtained results state the approximated end of the activity to 3:00 PM. Most notable regarding the CO2 levels recorded as seen in Figure 3, were those taken 18th August between 07:00 AM and 09:00 AM. In this period the ventilation system was just turned on and there would have been an arrival of personnel for the scheduled use that was expected to begin at 9:00 AM. This would mean there was a rise of CO2 before the system had a time to become effective. It would indicate that the ventilation system was functioning well if given a time to impact the air quality. Other lesser spikes of CO2 correspond with other active use times of the room, according to the schedule.

The activity in a room is not dependent only on the level of CO2 but also on the level of relative humidity and temperature. The weather can also affect the measured results. During the measurement there was sunny weather and the average temperature was 16.8°C on 17th August and 15.6°C on 18th August (www.yr.no).

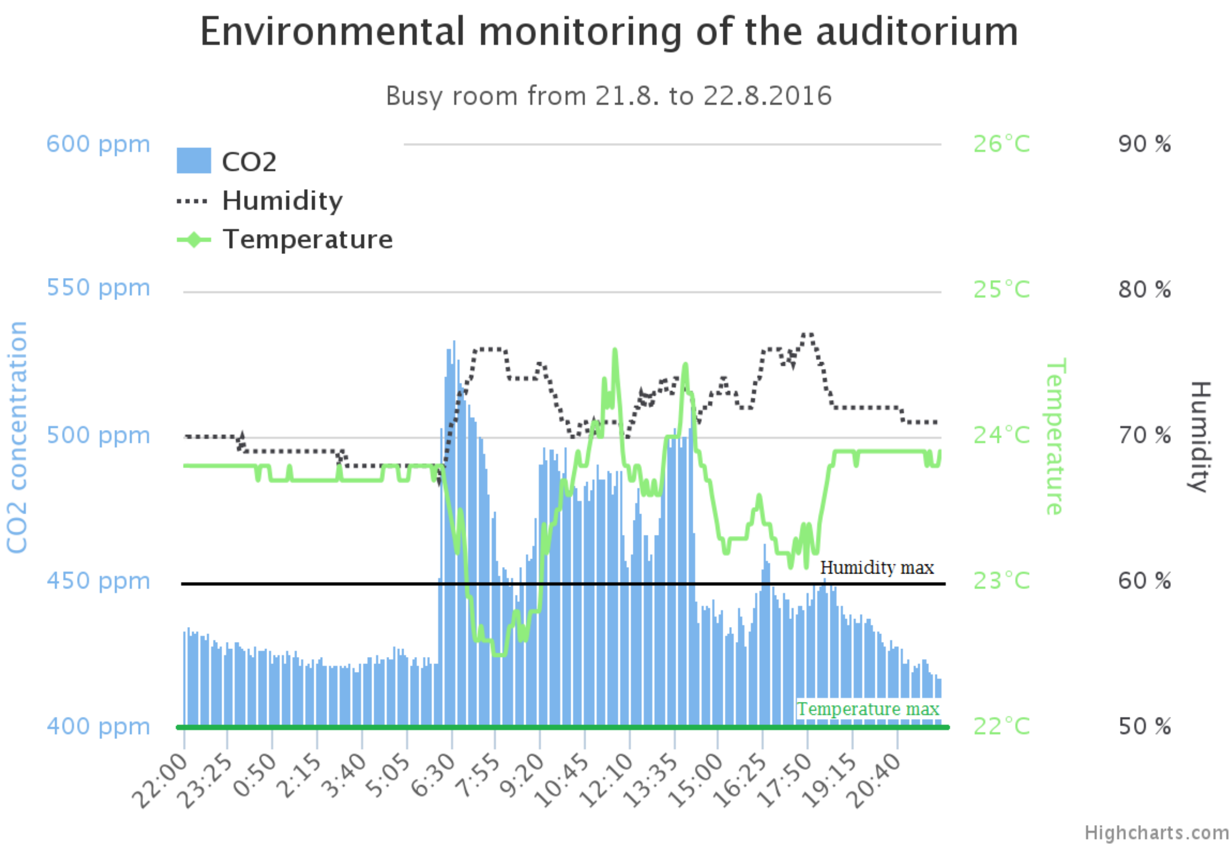

Figure 4. Air quality measures in the auditorium August 21st-22nd.

The data set C was gathered from August 21st until August 22nd. The recorded data are visualized in the Figure 4. The measurement started at 10:00 PM. The level of CO2, temperature and relative humidity in the auditorium did not change significantly during the night. In the morning the activity started, so all measured features had changed markedly. The room should be occupied on 22nd until 3:00 PM as stated in Table 1 but it was idle from 2:00 PM. There was no other activity in the monitored room so the CO2 level remained unchanged until the end of measurement. The CO2 concentration did not exceed the maximum value (1000 ppm). When considering the data obtained in previous days, it can be seen that the number of people present in the monitored room is strongly affecting the CO2 levels, so the maximum obtained CO2 level did not exceed 540 ppm. During the measurement there was sunny weather and the average temperature was 16.2°C on 22nd August (www.yr.no).

In brief, the findings from the three data sets show that the air quality in the room was affected by the numbers of persons simultaneously occupying the room. The ventilation system although using sensors to monitor its actions, was also turned off during the nights or periods of non-scheduled use time, most likely to save on energy costs. One decision that building management could make is to consider turning on the ventilation systems 1 hour earlier, or to change the automatic thresholds for the system to automatically turn on, e.g. increasing thresholds during the night, but lowering them again in time-frames prior to usage periods.

5. DISCUSSION

The challenge of graphical presentation of the data in a meaningful way is normally a big challenge for BD analytics. In this case study, however, the challenge does not appear to be very great because we are limiting the scenario to one lecture room (volume) and to the measurement of three internal climate factors (limited variety). The speed of incoming data (velocity) is also pre-determined, as we are measuring climate factors at one minute intervals. The challenges of volume and variety are most relevant to discuss here. These factors are somewhat related. As we increase variety of measures, we also increase the volume of BD to be considered in decision making.

For all data sets (A, B and C) we can use static images and threshold bars to indicate when attention is needed to adjust the ventilation system. However, such systems should be scalable to measure multiple work spaces (e.g. 100 rooms), to be able to signal when thresholds are exceeded and should have the ability to identify the location the rooms that have problems. The visualization in a scaled up case, could have displays for each monitored room in multiple windows. Alternatively, the display could only be shown for rooms with exceeded thresholds. While the measures would continue to come in and be processed, the display tool could be programmed to only open when these exceptions occur.

While in the lecture hall environment we are only measuring one gas (CO2). Alternatively, in industrial settings other dangerous gases may need to be measured. For example, in more complex industrial cases, with a greater variety of data (e.g. hydrogen sulfide – H2S, carbon monoxide – CO, etc.), the visualization system would need to be designed to provide real-time presentation of data that would indicate degree or urgency and warnings that may be needed to make more timely decisions.

The graphical presentations used in our study do leverage the ability of management to make sense of the volume of data that are collected from the lecture room sensors. Our visualization is more useful than viewing the numerical data in a stream as it is collected from the sensors. For example, Figure 5 presents only 15 minutes of data collected from one sensor. Such a presentation of data does not help management to see the big picture. Such a visualization is not effective for the present case, and would not be effective for monitoring of multiple spaces simultaneously.

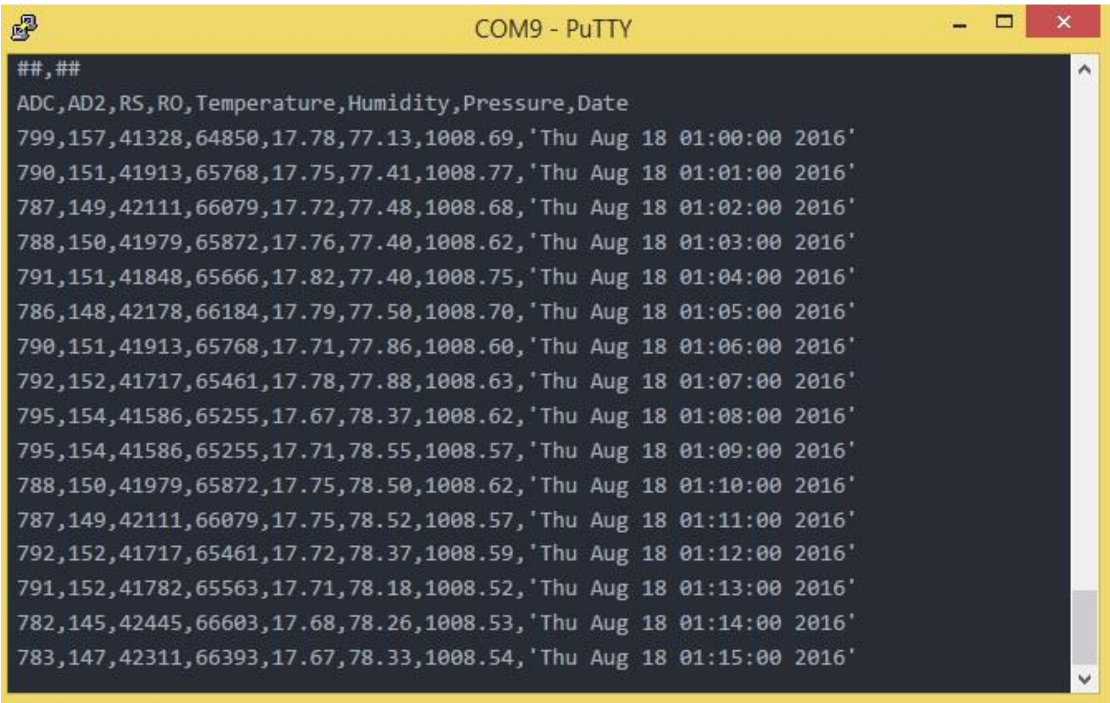

Figure 5. Measures of temperature, humidity and air pressure for 15 minutes on August 18th.

We also explored the possibility of presenting data using other graphical representations. One alternative considered was to overlap data collected across all the sampled 24 hour periods. Such a presentation could be used to look at difference between quality measures at time intervals across days of the week. Such a presentation would not be useful in this case, but it could be imagined that in an industrial setting that one objective might be to associate air quality with special activities (e.g. washing pipes on Tuesdays).

6. CONCLUDING REMARKS

This paper presented a systems analytics approach for monitoring and visualization of air quality using big data as collected with WSN technologies in the case study of one university lecture hall. The findings of this small case study have shown that not surprisingly, the number of persons simultaneously located in the room do have an impact on CO2 levels. However, the university building management may decide to turn on ventilation systems further in advance of the arrival of students. In brief, we demonstrated a proof of concept that the monitoring and visualization approach can be used by management to make decisions about worker environment conditions. We discussed alternative visualization approaches of BD are needed for different circumstances. In cases where more factors (variety) of data are needed, the visualization system may need to "select" to present the most critical data that exceeds thresholds, and down prioritize the presentation of other data.

We recognize that there many potential applications of WSN technologies and visualization of BD in industrial workplace environments. For example, in another research study we are measuring air quality in an industrial setting for a logistics business in the shipping industry. The workers in this location can be exposed to sharp but brief rises in CO2 levels as the workers are located near parts cleaning facilities. It is important for management to monitor work-spaces, in case any areas are not receiving proper ventilation. In workplace environments where quick decisions need to be made, continuous and near real-time views of big data may be needed. We suggest that future studies also examine the challenges of presenting real-time visualization of big data.

References

Arbeidstilsynet, (2012). Veiledning om klima og luftkvalitet på arbeidsplassen. Veiledning, best nr. 444 (2012).

Futurlec, (2016) MQ-135 datasheet, https://www.futurlec.com/Datasheet/Sensor/MQ-135.pdf

Garcia-Romeo D., Fuentes H., Medrano N., Calvo B., Martinez P. A., Azcona C., (2012). A NDIR-based CO2 monitor system for wireless sensor networks, Circuits and Systems (LASCAS), 2012 IEEE Third Latin American Symposium on , vol., no., pp.1,4, Feb. 29 2012-March 2 2012

Heiselberg, P.K. (Ed.) (2016). CLIMA 2016 – proceedings of the 12th REHVA World Congress: volume 7, Aalborg: Aalborg University, Department of Civil Engineering. Accessed from vbn.aau.dk on July, 19, 2016.

Hilbert, M. (2016). Big Data for Development: A Review of Promises and Challenges, Development Policy Review, 34(1), 135-174.

Jonqwon Kwon, Gwanghoon Ahn, Gyusik Kim, Chun Kim, Hiesik Kim, (2009). A study on NDIR-based CO2 Sensor to apply Remote Air Quality Monitoring System, ICROS-SICE International Joint Conference 2009

Kamran, A and Haq, S.U., 2013. Visualizations and analytics for supply chains, IBM developer Works, Technical report Feb. 26th, Accessed 03.02.14 at: http://www.ibm.com/developerworks/library/ba-visualizationsanalytics/

Laney, D. (2001). "Application Delivery Strategies", META Group, META Delta, File 949, 6 February.

Mycoteam, (2015) "Kontroll av inneklima I utvalgte rom" Report 2015 (ref: 201504051), 6. mai.2015.

Norbäck, D., Nordström K., and Zhao, Z. (2013). "Carbon dioxide (CO2) demand-controlled ventilation in university computer classrooms and possible effects on headache, fatigue and perceived indoor environment: an intervention study" Int. Arch Occup Environ Health (2013) 86:199-209.

Rozados, I.V. and Tjahjono, B., 2014. Big Data Analytics in Supply Chain Management: Trends and Related Research, 6th International Conference on Operations and Supply Chain Management, Bali, 2014.

Solberg, K. V., Cao, G., Alonso, M. J., and Burud J. P. (2016) "A Field Study of Indoor Climate in an Energy Efficient Building in Norway" In Heiselberg, P. K. (Ed.) (2016). CLIMA 2016 - proceedings of the 12th REHVA World Congress: volume 7. Aalborg: Aalborg University, Department of Civil Engineering.

Shahzad, S. S., Brennan, J., Theodossopoulos, D. Hughes, B., Calautit, J. K. (2015) "Energy efficiency and user comfort in the workplace:Norwegian cellular vs. British open plan workplaces", Energy Procedia 75, 807–812.

Skyberg, K. and Skuleberg, (2003). "Symptoms prevalence among office employees and associations to building characteristics", Indoor air 13(3) 2003, 246-252.