

Experiential learning of data acquisition and sensor networks with a cloud computing platform

Eric Mahoney,^{a†} Colleen Chau,^{a†} Qiyin Fang^{a,b*}

^aMcMaster University, School of Biomedical Engineering, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1.

^bMcMaster University, Department of Engineering Physics.

[†]Contributed equally

ABSTRACT

Experimental data acquisition and statistical data analysis are core components in photonics undergraduate curriculum. Although it focuses on experimental data, the content is usually delivered by a lecture-based format. This disconnect is partially due to the contents are delivered at the early years of the program when experimental data acquisition techniques have not yet been introduced. In a second-year data acquisition and applied statistics course, we have designed an experiential learning module that covers the fundamental content of data acquisition and statistical analysis. This module has a single physical experimental setup that is continuously measuring environmental parameters (temperature, humidity, light, imaging, etc.) with a set of multi-modality sensors in an Internet-of-things (IoT) big data platform. Different types of sensors measuring the same parameters are also used for cross-validation purposes. The data is streamed to a cloud computing platform, allowing each student to acquire their own subset of data, then perform processing and analysis. The capability of remote access to a physical sensing experiment provides the students with hands-on learning opportunities on a managed complex data acquisition system. The platform provides a set of powerful visualization tools to allow a multi-dimension view of complex data streams (e.g. time-lapse of statistical distribution). Such IoT data acquisition platform allows key concepts to be demonstrated, applied, and tested including error propagation, distribution and test of distribution, correlation and cross-validation, data rejection, and signal processing. This experiential learning module has been demonstrated to be more effective in achieving related learning objectives through quantitative graduate attribute measurements as well as qualitative feedback.

Keywords: data acquisition, statistics, data analysis, data processing, experiential learning, cloud computing, Internet-of-things, IoT, sensors

1. Introduction

Most undergraduate engineering curriculum requires applied statistics. A key aspect of applied statistics is closely associated with experimental data acquisition, processing, and analysis. Typically, statistics for engineers is taught through a traditional lecture-based format and provides students with the background and tools necessary for application in engineering concepts. However, students often find it difficult to apply the tools to practical engineering problems because statistics is taught as a theory heavy math course that has limited relationships to their engineering speciality [1].

In the McMaster Engineering Physics program, the first statistics course that the students take is Eng Phys 2W03, “Applied Statistics for Engineering” with a focus on experimental data acquisition and analysis. This course introduces second year engineering physics students to estimation of true value, probability density functions, analysis of variance, experiment design, and application of statistical analysis. The 3-credit unit course consists of three-hour lectures and a tutorial every week over the span of one semester. In the past years, theoretical concepts were delivered through lectures and Matlab and Microsoft Excel was taught in tutorials for data processing and plotting (including curve fitting). The assignments were analytical problems from the textbook (John R. Taylor: “An Introduction to Error Analysis: The study of uncertainties in physical measurements,” 2nd Edition, 1997, University Science Books).

Often in early level engineering classes, students have not yet learned data acquisition instrumentation. As a result, the data acquisition and analysis are taught on paper using hypothetical experiments and simulated data sets. The lack of opportunity to connect statistics with other engineering courses causes students to view the course as an entirely separate topic rather than an essential part of engineering practice [1].

Experiential learning is a process through which the learner constructs knowledge, skill, and value from direct experiences. At the undergraduate level for engineers, experiential learning is delivered through course projects, internships, and capstone projects. However, these types of vehicles for delivery are usually introduced in the senior years when students have the fundamental theoretical knowledge to use during the experience. Engineering students should be introduced to real problems and application of theoretical knowledge as early as possible since they build the critical skills that are vital to the engineering profession [2]. Experiential learning can also provide the means of learning through action as students who chose to study engineering enjoy “doing things” which should translate to better learning outcomes [3].

To address these challenges, we developed an experiential learning module that uses a smart environmental sensing system providing students hands-on experiences to apply the statistics concepts and theories to process and analyze real engineering data. The smart sensor system is capable of continuously measuring multiple environmental factors (temperature, humidity, light, images) with different sets of sensors. The data acquisition is achieved by an Arduino based platform and experimental control and data communication is managed by a Raspberry Pi. The data is continuously uploaded to a cloud data platform, PI Vision, created by our industry partner, OSIsoft, where the students can view and download the data for analysis.

2. Data Acquisition System (DAQ)

The smart sensor station system that has been developed can be described as a sensing module that periodically measures the local environment, a relay module that sends the data to a remote server, and a webpage-based user interface used to view and export data as shown in Figure 1.

2.1 Sensing Module

The sensing module of the DAQ system consists of an Arduino UNO and several sensors for climate monitoring. The sensor set includes an ambient light sensor, two temperature sensors with different precisions, and a humidity sensor. The Arduino UNO is used to measure each group of sensors through its inputs once every 30 seconds. The ambient light sensor is measured as an analog input between 0 and 3.3V where the voltage measured is proportional to the intensity of light in the room. The temperature and humidity sensors are connected as digital inputs and the digital signal is processed to determine the temperature and relative humidity. Two sets of sensors were built into two separate sensor stations. The DAQ system is controlled by a Raspberry Pi (v3 Model B+) equipped with a CMOS camera module to capture pictures of the environment at the moment of each measurement.

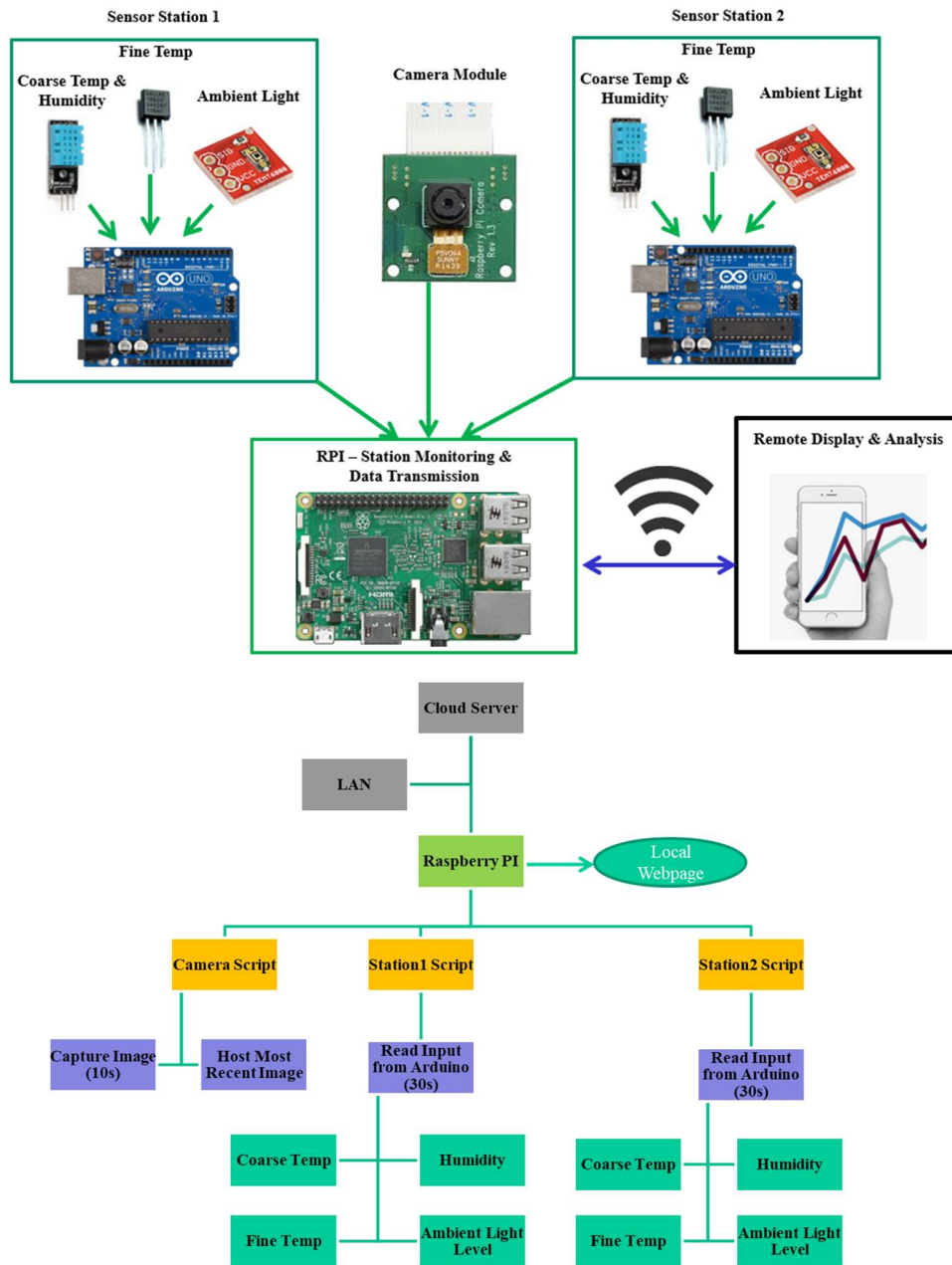


Figure 1: Sensor station system schematics and process flow

The Raspberry Pi serves to distribute power from the supply to the sensor modules, formats the data retrieved from the sensor modules, and relays the results to the cloud based OSISOFT PI Vision server over the internet. The Raspberry Pi runs a python script that allows it to read the sensor results from each Arduino using serial communication; then stream the data to the server as a formatted message that identifies the station, the time the data was captured, the data type, and the measurement result. The data can then be viewed from the PI Vision display in real time and exported for analysis from a computer or wireless device.

2.2 Cloud Data Storage and User Interface

The OSIsoft PI Vision is a web enabled user interface that is accessed remotely over the internet as shown in Figure 2. This enables multiple concurrent users to access results from a browser. After logging in, the display can be configured to show all of the data collected or it can show data for fixed time frames in the form of graph indicators. The data can then be exported from the display for further processing and analysis. Alternatively, gauges and dials can be added to the display to show the most recent results reported by each sensor module. These indicators can be arranged over a schematic such as a floor plan or a front panel display to provide additional context to the data being collected. The display can be easily reconfigured or updated to include additional indicators as additional DAQ modules are added to the system.

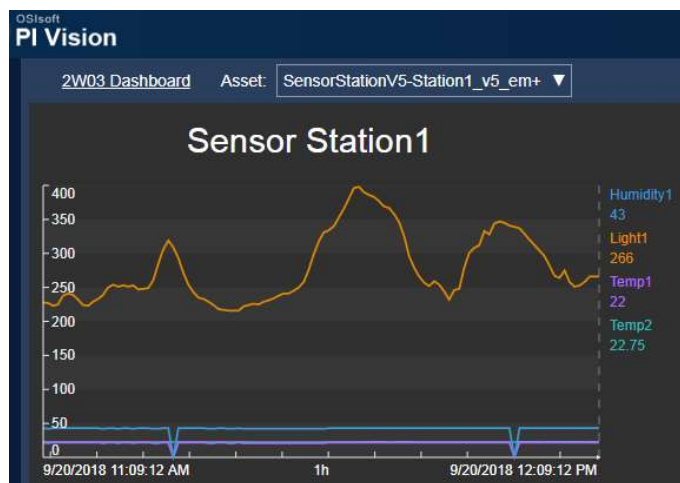


Figure 2: OSIsoft PI Vision display

3. Learning Objectives and Outcomes

The DAQ system was placed facing a window in a room in the Burke Science Building (BSB) at McMaster University, where all Eng Phys 2W03 students performed their introductory Electromagnetism lab. Hence all students in this course can physically see, but not adjust, the system (Figure 3). Building the unit for the students allows separation of the challenges in electronics hardware so that they can focus on the data acquisition aspects, which is the emphasis of the course. In addition, the PI Vision cloud data platform gives all students access to the data using one data acquisition module, significantly reducing costs associated with labs, such as experimental hardware, maintenance, and space. PI Vision also gives students the flexibility to visualize and analyze the data on their own time, from any device with internet access. The design allows the development of additional lectures that cover basic smart sensor technology and, most importantly, how to deal with real experimental data including noise, systematic errors, random errors, and multi-sensor validation (from different temperature and humidity sensors), correlation, system response, etc. Abstract concepts (e.g. distributions of random error, correlations between different sensors (e.g. light vs temperature), erroneous data rejection, digital signal processing, etc.) introduced in the lecture can now be associated with real data sets. At the same time, students still learn to use Matlab and Excel as they are required to process, plot, and analyze data.

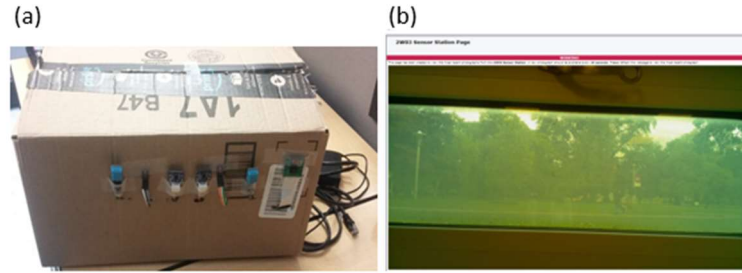


Figure 3: DAQ sensor system (a) and camera view of the window (b)

The following concepts were taught with the DAQ module:

- Plotting data with error bars
- Understanding differences between different types of sensors of the same sensing parameter
- Plotting sensor data distributions
- Using Chi-squared test of distributions
- Developing hypothesis to explain data trends and hypothesis testing
- Determining outlier rejection
- Making conclusions about whether variables are correlated and how correlated variables can be used to predict each other
- Calculating and determining if variances are caused by random or systematic errors

3.1 Assignments and Tutorials

The DAQ system was mostly introduced and discussed in tutorials and assignments so that students would obtain theoretical background knowledge in the lectures. In the first tutorial, students were introduced to the DAQ station. The system was brought to the tutorial session and students were able to physically see the components as the component specifications were being reviewed. The students would be able to view the system anytime in the lab where it is constantly acquiring data, if they wish to see the setup again. The data acquisition process and equipment were also explained to them. They were then shown how to extract data from the PI Vision dashboard. In subsequent tutorials, the assignments were brought up and the students would self-grade their assignments which will be discussed in the next section of the paper.

The assignments associated with the DAQ system were developed as case studies and followed the flow of the course. The DAQ assignments allowing the students to acquire selected sets (e.g. time frame) of data and work with the data set. Since the DAQ module is acquiring data 24/7, assignments can be designed so that different data sets can be used for different students, preventing plagiarism and promoting higher critical thinking.

The learning outcomes for each case study are listed Table 1 below.

Table 1: Case Study Learning Outcomes

Case Study 1	Case Study 2	Case Study 3
<ul style="list-style-type: none"> • Set up remote access, data download • Preprocess the data in Excel or Matlab • Plot selected data with error bars • Understand the differences between different types of sensors of the same sensing parameter 	<ul style="list-style-type: none"> • Plot sensor data distributions • Chi-squared test of distributions • Develop hypothesis to explain data trends and hypothesis testing • Outlier rejection 	<ul style="list-style-type: none"> • Correlation between different sensing modalities • Variances

3.2.2 Case Study 1

In the first case study, students learn how to export data from PI Vision and either pre-process it in Excel or import it directly into Matlab. They are then asked to calculate the running average and running standard deviation for a 1-hour window over a 12-hour period. Depending on their student numbers, students were split into two groups, so that each group used a different 12-hour data set. Each of the 4 sensors were plotted as a running average and must include standard deviations with error bars. The students were then asked to give an explanation for the shape of the error bars observed in each series and learn the statistical significance associated with each type of sensor.

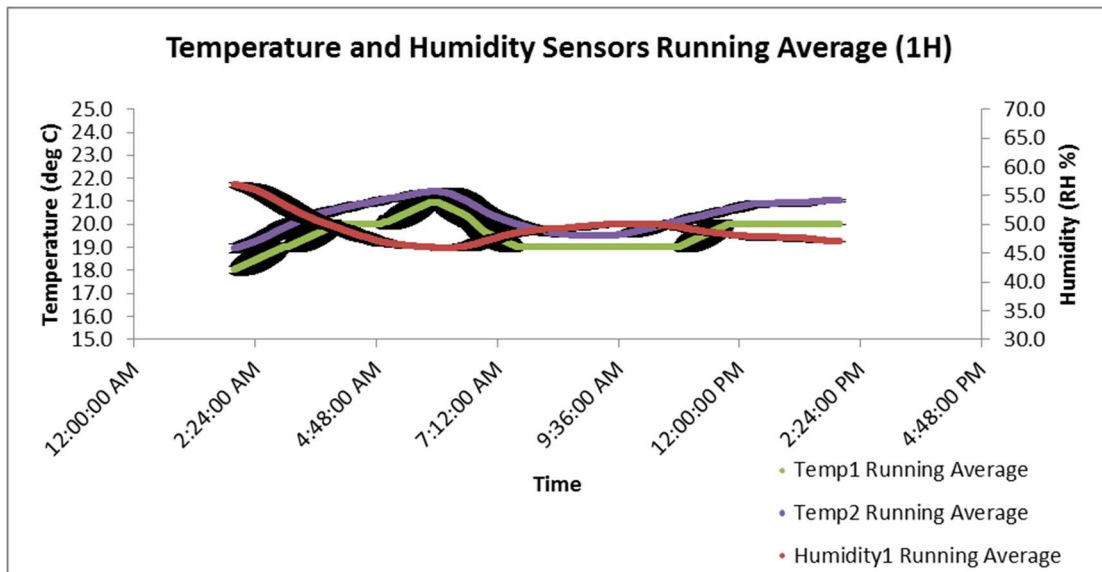


Figure 4: Temperature and humidity sensor running average plot with standard deviations

The plot of the data is shown in Figure 4. For regions where the slope looks to be relatively flat, the data is not changing over a long period of time. Therefore, the standard deviation must be small. Where the slope increases, there is a greater difference between values over time, so the standard deviation must increase.

For the coarse temperature sensor, there is an egg-shaped error where the temperature changes with greatest error observed when the average is between two discrete temperature values. Error appears to be close to zero when the values are equal to discrete temperature readings. The coarse temperature sensor can only predict temperature in 1-degree increments. To plot an average of 1/2 degree, half of the values must be above and half must be below the average, this is where the error is greatest. When a reading is equal (or almost equal) to a discrete temperature, the temperature reading has been ~constant for 1 hour, so the error is small.

3.2.3 Case Study 2

For the second case study, students were asked to plot histograms of two humidity sensors in the same figure and assume the data follows a normal distribution. They then had to prove or disprove the assumption using a reduced Chi-squared test. If there was any outlier data that they had rejected, they must explain why.

The plots associated with case study 2 is presented in Figure 5. Figures 5a and 5b shows the short and long term humidity sensor plots. After outlier rejection, Figures 5c and 5d show the normal distribution and Chi-square tests. The expected counts are determined by looking at the range of data, average, and standard deviation. Based on the large Chi-squared value, the distribution is determined not normally distributed. Data outliers were defined as data that was greater than 3 standard deviations from the average or if readings of 0% RH are accompanied by 0 on the other sensors.

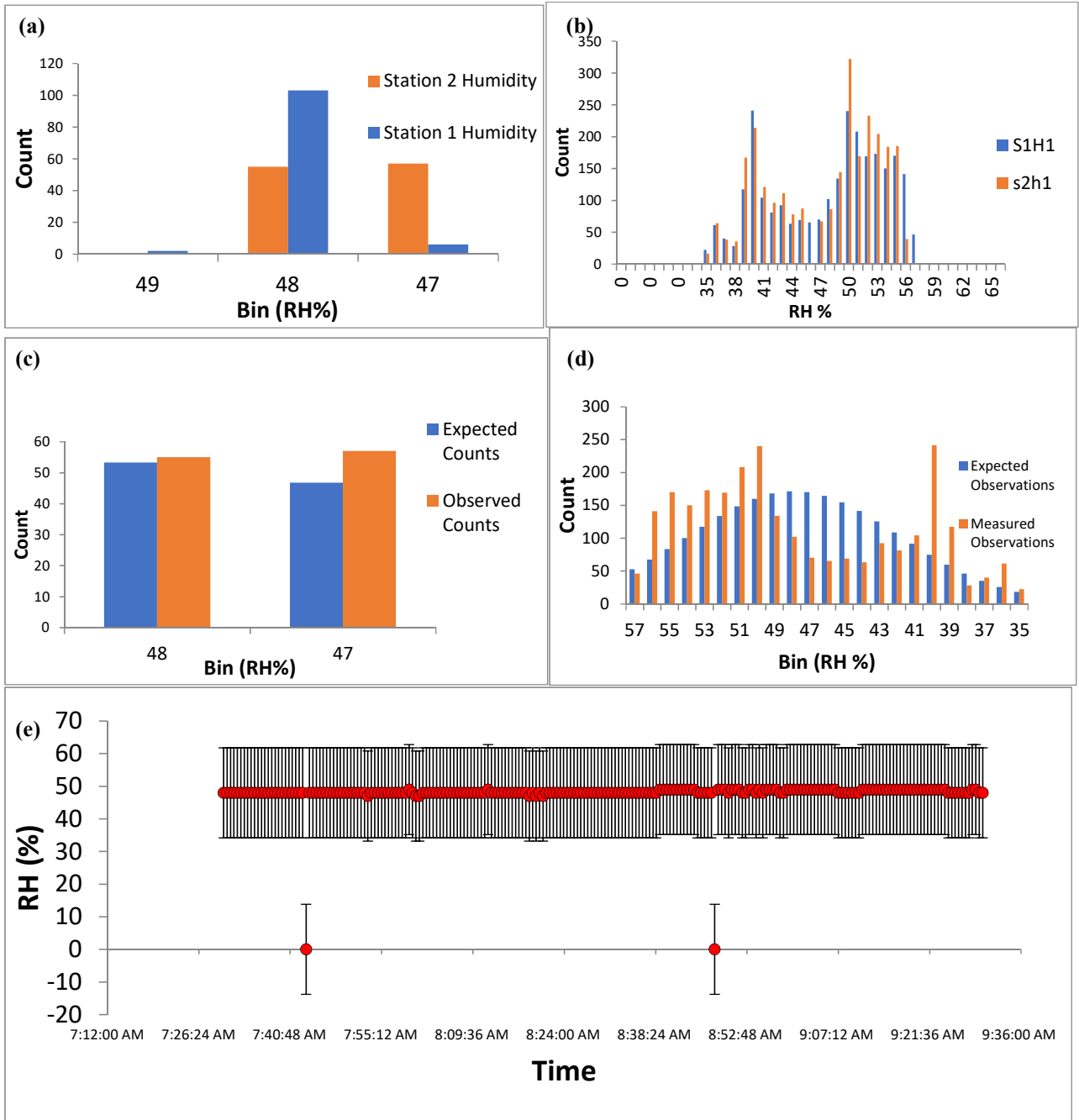


Figure 5: Short term relative humidity (RH) measurements for two sensor stations (a), and long-term relative humidity measurements for two sensor stations (b). Short term observed vs. expected counts for sensor station 1 after outlier rejection (c) and long term observed vs. expected counts for sensor station 1 after outlier rejection (d). Station 1 relative humidity outlier rejection (e).

3.2.4 Case Study 3

Case study 3 tasks the students with producing correlation plots for a 24-hr period to determine if the measured values are correlated with one another. They also compare the fine and course temperature sensors in the same manner. They are then asked to plot the data as a 15-minute moving average with standard deviations for each of the sensors. A

trendline is added to determine the R^2 value and trendline equation. The final exercise is to determine the variance for temperature and light levels and whether the observed variance is random or systematic.

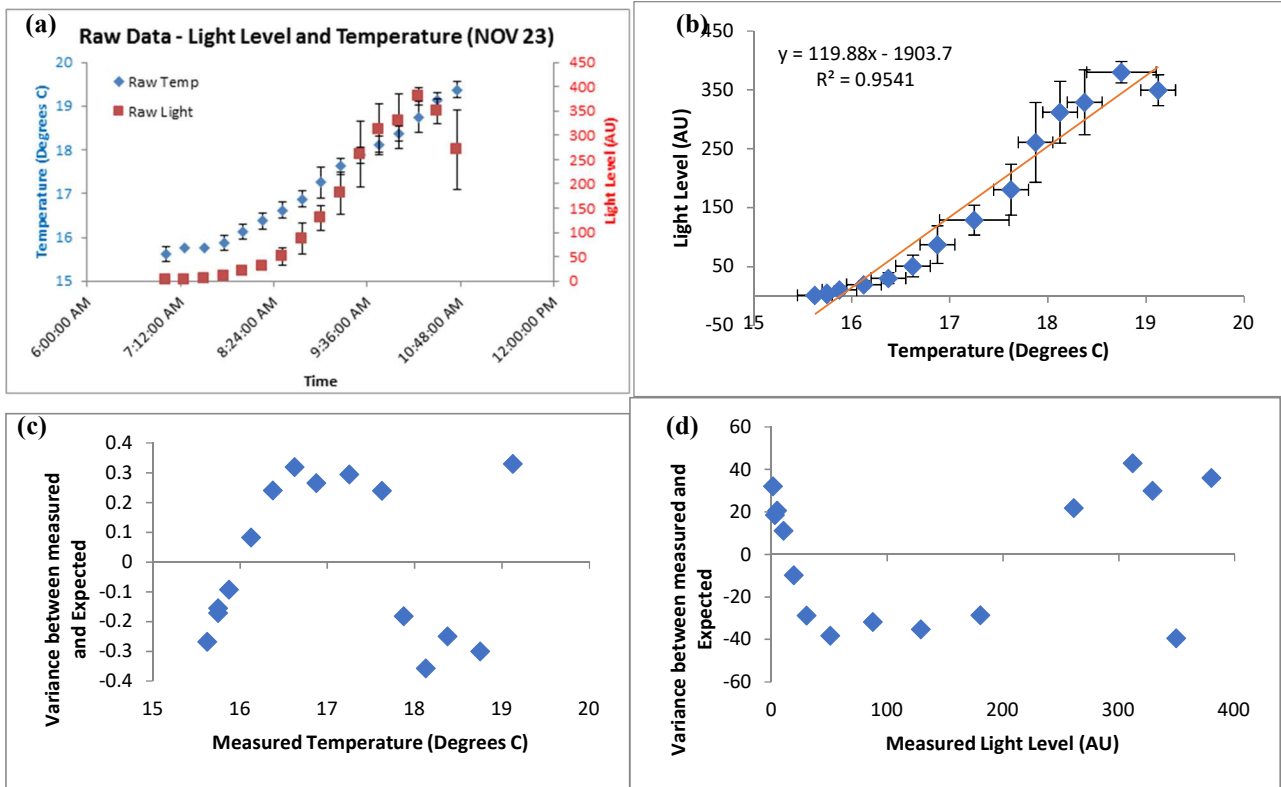


Figure 6: Temperature and light level over a 4-hour period (a); correlation of temperature and light level (b); variance between measured and expected temperature (c); and variance between measured and expected light level (d).

Figure 6b shows that when the data is binned into 15 minute averages and after outlier removal, the temperature and light level trends seem to be correlated with each other as indicated by the R^2 value. This case study shows the importance of allowing the students to see the experimental setup and the conditions that the system was in. One would expect that temperature and humidity would correlate well with each other and temperature would also correlate with the light level. However, we see little correlation between temperature and humidity because the DAQ system was placed inside a building that is regulated by HVAC circulation. There is a loose correlation between the temperature and light levels. The DAQ system is placed at the window where it is exposed to the difference of light levels throughout the day. Despite the HVAC system in the building, the temperature near the glass window will change throughout the day as the sun warms the material, hence the slight correlation.

Having dual temperature sensors and different precision temperature sensors allows for verification and comparison between the same types of sensors. Both the coarse and fine temperature sensors are highly correlated, showing that the variances for temperature sensors indicate systematic error. The camera allows the students to observe the lighting in the window to compare with light level sensor readings throughout the day.

3.3 Self – Grading of Assignments

In traditional assignments where the students solve analytic problems, one issue is that they would do the homework but usually will not look at them afterwards. Since identifying their own mistakes is an important learning method, the students brought their completed assignment to tutorial to mark along with the TA, based on a given rubric. The intent is for the students to review their own assignments and reinforce the learning.

3.4 Evaluation Structure

The previous year’s evaluation structure is compared to this year’s evaluation structure in Table 2. The intent of the DAQ module is to replace the theoretical questions with hands-on data acquisition and provide real data sets to work with. The material developed for the module does not add extra work to the course as the assigned questions from the textbook are reduced to accommodate for the DAQ module. The overall load of the course is not significantly changed.

Table 2: Evaluation Structure Comparison

	Previous Year’s Evaluation structure without DAQ	Revised Evaluation Structure with DAQ
Assignments	40%	10%
Mid-term exam 1	20%	20%
Mid-term exam 2	-	20%
DAQ module – hands-on data processing	-	10%
Final exam	40%	40%

3.5 Metrics

The metrics used to evaluate the learning outcomes with the new DAQ module was by comparing the midterm and exam scores from previous years and through the Canadian Engineering Accreditation Board (CEAB) graduate attribute measurement in knowledge base competence.

Although the questions in the midterms and exams vary from year to year, the difficulty level and concepts tested are comparable. In the first midterm, the class average was $78 \pm 12\%$, which is significantly better than the average grade of $66 \pm 15\%$ during 2014-2015 using similar questions. We also saw significant increase in final exam grade: $75 \pm 14\%$, over $55 \pm 15\%$ in 2014-2015. However, the final exam questions are different between the two years so the results are not directly comparable.

The Canadian Engineering Accreditation Board (CEAB) accredits undergraduate engineering programs. One of their accreditation evaluation criteria is graduate attributes which means that the institution must demonstrate that the graduates of a program possesses competency in certain areas. The indicator for knowledge base competence in natural sciences and engineering fundamentals was chosen to evaluate the performance outcomes of the effectiveness of the DAQ. The 2018 assessment is compared to the last assessment done in 2014. The data presented in *Figure 7* indicates that there was no statistical significance between the two years.

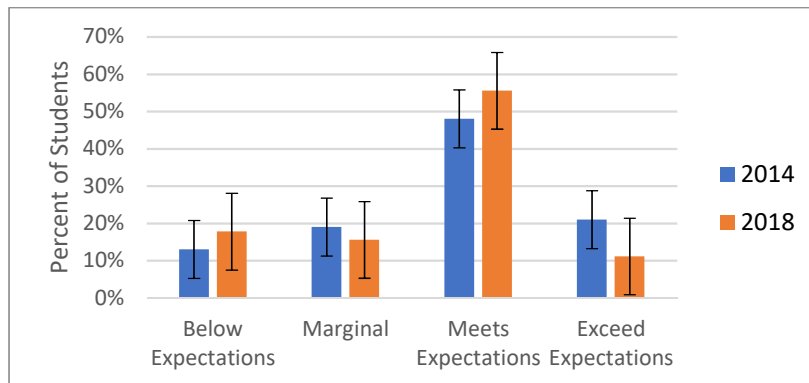


Figure 7: Eng Phys 2W03 graduate attributes assessments

These measurements do not show a systematic and statistically significant measurement on student’s learning outcomes. Without a control group, it reduces the ability to make casual conclusions. In order for this to be a robust pedagogy

research experiment, we would need to run a control group that does not include the DAQ module alongside a group that uses the DAQ module. We need to take into account the student's experience about the addition of the new module as this module is meant to be an experiential learning experience. We would also need a way to measure the carry over of knowledge to other courses and projects. However, based on the midterm and exam data that we have, we can hypothesize that the DAQ module may improve engineering students' fundamental tool box in math and statistics through a hands-on experience that allows them to see and understand the experimental conditions based on exam scores. We hope that they will be able to link the material learned in this course to their future courses and labs.

4. Conclusion

As technological advances are made and are accessible to students and educational institutions, our methods of teaching are also able to evolve to more efficiently teach our future engineers. With this DAQ system, students perform and learn data acquisition, processing, and statistical analysis through IoT big data platform. The experiential learning module that we have created encourages a hands-on approach to statistics which is what engineers need to see their theory put into practice. As a result, students achieved higher test results and can link the content learned in this course to other courses, laboratories, and hopefully in their professional careers.

Acknowledgement

The PI Vision software is provided by OSIsoft. The authors acknowledge Mr. Christian Foisy and Ms. Erica Trump from OSIsoft for their extensive technical support to the PI Vision platform. EM acknowledges a McMaster Faculty of Engineering Dean's PhD Excellences Award and the Ontario Graduate Scholarship. CC acknowledges the support from the Paul R. MacPherson Institute for Leadership, Innovation & Excellence in Teaching at McMaster University and its Graduate Student Educational Developer (SED) Program.

References

- [1] S. Bisgaard, "Teaching statistics to engineers," *The American Statistician*, vol. 45, no. 4, pp. 274-283, 1991.
- [2] A. Hajshirmohammadi, "Incorporating Experiential Learning in Engineering Courses," *IEEE Communications Magazine*, vol. 55, no. 11, pp. 166-169, 2017.
- [3] R. J. Wilson, "'WHAT DOES THIS HAVE TO DO WITH US?'" TEACHING STATISTICS TO ENGINEERS," in *ICOTS-6, the Sixth International Conference on Teaching Statistics*, Cape Town, 2002.