
# AFRL-RH-WP-TR-2021-0082

# MULTI-HAZARD MONITORING NETWORK

**Thomas M. Peters, Oliver Stroh, Geb Thomas**

**The University of Iowa**

**DECEMBER 2021**
**Final Report**



**AIR FORCE RESEARCH LABORATORY**
**711TH HUMAN PERFORMANCE WING,**
**AIRMAN SYSTEMS DIRECTORATE,**
**WRIGHT-PATTERSON AIR FORCE BASE, OH 45433**
**AIR FORCE MATERIEL COMMAND**
**UNITED STATES AIR FORCE**

# NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the Air Force Research Laboratory Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RH-WP-TR-2021-0082 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

DURAN.CHRISTIN.MARIE.1286379393 Digitally signed by DURAN.CHRISTIN.MARIE.1286379393 Date: 2022.02.08 12:08:49 -05'00'

CHRISTIN DURAN, DR-III, PhD
Human Optimization Branch
Airman Systems Directorate
711th Human Performance Wing
Air Force Research Laboratory

MCKINLEY.RICHARD.A.1262239710 Digitally signed by MCKINLEY.RICHARD.A.1262239710 Date: 2022.02.11 09:14:44 -05'00'

R. ANDY MCKINLEY, DR-III, PhD
Core Research Area Lead
Cognitive and Physical Performance
Performance Optimization Branch
Airman Biosciences Division

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE *(DD-MM-YY)* | 2. REPORT TYPE | 3. DATES COVERED *(From - To)* |
|---|---|---|
| 15-12-21 | Final | 2018 - 2021 |

**4. TITLE AND SUBTITLE**
Multi-Hazard Monitoring Network

**5a. CONTRACT NUMBER**
FA8650-18-C-6935

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
*Thomas M. Peters, Oliver Stroh, Geb Thomas

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**
Legacy RHM/USAFSAM

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
*The University of Iowa
145 N. Riverside Drive
Iowa City, IA 52242

**8. PERFORMING ORGANIZATION REPORT NUMBER**
S-158-112-001

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Air Force Materiel Command
Air Force Research Laboratory
711th Human Performance Wing
Airman Systems Directorate
Airman Biosciences Division
Performance Optimization Branch
Wright-Patterson AFB, OH 45433

**10. SPONSORING/MONITORING AGENCY ACRONYM(S)**
711 HPW/RHBC

**11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)**
AFRL-RH-WP-TR-2021-0082

**12. DISTRIBUTION/AVAILABILITY STATEMENT**
Distribution Statement A: Approved for public release.

**13. SUPPLEMENTARY NOTES**
Report contains color. AFRL-2022-0256, cleared 20 January 2022

**14. ABSTRACT**

Personal sampling is a primary method used to ensure that a working population is protected from chemical and physical hazards. However, analyses of personal samples are often expensive, resulting in use of a few personal exposure measurements of a subset of multiple hazards to represent exposures for large working populations. This work developed a wireless sensor network to assess multiple hazards in real-time at many locations in the workplace. The sensor network was deployed at Robins Air Force Base to assess noise, particulate matter, carbon monoxide, total volatile organic compounds, temperature and relative humidity in a hangar where large cargo planes were being refurbished. Collected data were visualized through time series and heat maps. In conclusion, this project demonstrated the usefulness of the sensor network in a real-world environment.

**15. SUBJECT TERMS**
gases, vapors, aerosols, noise, chemical exposure, occupational exposure, low-cost sensors

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT: | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON (Monitor) |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Christin Duran |
| Unclassified | Unclassified | Unclassified | SAR | 31 | 19b. TELEPHONE NUMBER *(Include Area Code)* N/A |

CONTENTS

## List of Figures



## List of Tables

# 1.0 Summary

Personal sampling is the primary means to ensure that a working population is protected from chemical and physical hazards such hazards. However, analyses of personal samples are often expensive, resulting in use of a few personal exposure measurements of a subset of multiple hazards to represent exposures for large working populations. This work developed a wireless sensor network to assess multiple hazards in real-time at many locations in the workplace. The sensor network was deployed at Robins Air Force Base to assess noise, particulate matter (PM), carbon monoxide (CO), total volatile organic compounds (TVOCs), temperature and relative humidity in a hangar where large cargo planes were being refurbished. Collected data were visualized through time series and heat maps. In conclusion, this project demonstrated the usefulness of the sensor network in a real-world environment. Future work will optimize the number of network notes for a given occupational space and compare results to traditional personal exposure measurements.

# 2.0 Introduction

Workers are exposed to a variety of chemical and physical hazards. Aerosols and gases are important chemical hazards, resulting in a wide array of adverse cardiopulmonary health effects. Exposures to high levels of noise and heat stress are common in many trades. Industrial hygienists use personal sampling as the primary means to ensure that a working population is protected from such hazards. For aerosols and gases, personal sampling of breathing zone air results in a time weighted average (TWA) measure of personal exposure that can be compared to an occupational exposure limit (OEL).

However, analyses of personal samples are often expensive, resulting in use of a few personal exposure measurements of a subset of multiple hazards to represent exposures for large working populations. Such sparse measurements are unlikely to adequately characterize the large variability in exposures between- and within-workers typical of most workplaces. Understanding variations in exposure, however, is critical to identifying the determinants of exposure and devising effective risk mitigation strategies.

Direct-reading instruments (DRIs), sensors that provide near real-time output of hazard concentration or intensity, can be applied to assess exposure variability through task (temporal) monitoring or hazard mapping (depiction of concentrations measured spatially as contours on a map). However, DRIs have found limited use in industrial hygiene practice because traditionally they have been relatively expensive (>$4,500 each). Recently, researchers have taken advantage of advances in sensor technology by simultaneously assessing temporal and spatial variability in environmental exposures through distributed networks of inexpensive ($10 to $500) sensors. Applied to workplaces, distributed monitoring with inexpensive sensors may provide a way to assess exposure variability with unprecedented resolution applicable to risk evaluation and mitigation.

Dr. Peters at The University of Iowa has developed a distributed monitoring network of low-cost sensors for multiple chemical and physical hazards (Thomas, Sousan, Tatum, et al., 2018). During a previous effort, this network was deployed for eight continuous months in a heavy-vehicle machining and assembly center. Preliminary work was conducted to estimate personal exposures by integrating network measurements with worker tracking information. The network data was also used to examine the spatiotemporal variability of these multiple hazards.

In this project, we developed a revised version of the low-cost sensor monitor based on lessons learned from the first deployment. Then, the network was deployed at one USAF facility and operated for several weeks. Finally, data analysis was conducted to examine the variability in network data.

The time-series data obtained at many sites within a workplace have numerous exciting uses, including identification of the determinants of exposure (e.g., tasks, processes, and/or control measures), integration into production meetings as an air quality metric to avoid fouling of products and machines, and assessment of factory-wide hazards, such as dust explosions. The distributed network can be used as a surveillance system to trigger additional sampling (including other hazards or aerosol chemical speciation) when threshold levels are exceeded. Additionally, we propose that these data can

be used to predict worker exposures, for all workers and all time scales at minimal additional cost and without the need for costly and burdensome personal sampling equipment. This research also has great potential for research to practice because hazard maps will facilitate rapid dissemination of hazard information between employees, management, and regulators. Increased knowledge of exposure variability will improve the surveillance and epidemiological study, and make possible dose metrics (exposure × time) appropriate for the study of chronic disease.

# 3.0 Methods and Results

## Task 1. Select a USAF facility to deploy the UIowa multi-hazard monitoring network.

In consultation with the project officer, we selected two locations to deploy the network: the flight line at Nellis Air Force Base (AFB); and maintenance hangars at Robins AFB.

## Task 2. Revise the UIowa multi-hazard monitor.

*Sensors.* This work focused on selecting sensors for contaminant endpoints, including PM, noise, CO, TVOCs, temperature and humidity. A sensor was also included to monitor global positioning system (GPS) location.

We reviewed the sensors in our previous multi-hazard monitor that were used in heavy-vehicle manufacturing facilities (Version 1; **Table 1**). We found that for PM measurements, the Sharp GP sensor was lacking in repeatability and sensitivity. It was thus replaced in Version 2 with the PMS7003 from Plantower. The PMS7003 was selected as having favorable performance in peer-reviewed journal articles, although its performance under high concentrations is unknown, however. Future work includes testing of newly available PM sensors or development of a custom sensor based on camera technology that would facilitate cleaning.

The noise sensor used in the Version 1 was a custom sensor made at the UIowa. We found good agreement with reference sensors and therefore retained it without change in Version 2. There are however new things that could be done developmentally with the noise sensor. The current range for the sensor is from 65 dBA to 100+ dBA. Further work to extend the low range would be beneficial to accurately assess dose. The sensor could also be miniaturized into a stand-alone package that could be integrated into a monitor or worn on a person.

The CO sensor used in Version 1 functioned well and was retained in Version 2. TVOC was a new endpoint for us in Version 2. We selected the SGP30 from Adafruit. The SGP30 is a metal oxide that combines multiple sensors on one chip to measure TVOC and equivalent calculated carbon dioxide ($eCO_2$) values. This sensor measures $eCO_2$ concentration within a range of 400 to 60,000 parts per million (ppm) and TVOC concentration within a range of 0 to 60,000 parts per billion (ppb). The sensor responded to nail polish, which is high in VOCs. Further tests were not conducted. Tests of other VOC sensors, including a newer version of the SGP30, may provide an improved sensor.

The $CO_2$ sensor used in Version 1 (CO2-D1, Alphasense) was replaced in Version 2 with the SGP30, which measures 'electronic' $CO_2$. We felt $eCO_2$ was adequate as $CO_2$ was not a primary target. Ozone (O3-B4) was dropped from Version 2 because it was not a primary target for this project. Temperature and relative humidity sensors from Version 1 were retained in Version 2 (HT22dd).

**Table 1. Sensors selected to measure various endpoints for this project**

| Endpoint | Version 1 (previous work) | Version 2 (this project) | Notes |
|---|---|---|---|
| Particulate Matter* | GP2Y1010AU0F, Sharp | PMS7003, Plantower | PMS7003 not evaluated under high loading conditions<br>Possible new sensors available but all have cleaning issue; we have a solution based on camera technology that could overcome these issues |
| Noise* | Custom, U. Iowa | Same | Upgrades could include decreasing minimum range; repackage into smaller form-factor |
| Carbon monoxide* | CO-B4, Alphasense | Same | Upgrade could be to change to smaller sensor |
| Total volatile organic compounds* | N/A | SGP30, Adafruit | Newer sensors have emerged that may have better performance |
| Carbon dioxide | CO2-D1, Alphasense | SGP30, Adafruit | SGP30 reports 'electronic' $CO_2$ |
| Ozone | O3-B4, Alphasense | not applicable | |
| Temperature | HT22, Adafruit | Same | |
| Humidity | HT22, Adafruit | Same | Have had issues with humidity sensor after time |
| GPS location | N/A | NEO-6M, MakerFocus | Serves to locate node and sync clocks |

*Asterisk indicates primary target.

*Monitors (nodes)*. We integrated all of the selected sensors shown in **Table 1** into the monitor shown in **Figure 1**. This monitor can be deployed in a roving or stationary installation in Aim 2 field studies. The monitor draws power directly from on-board lithium ion batteries. These batteries can be charged from moving machines (e.g., a forklift truck) or from the wall with an alternating current / direct current transformer. The monitoring system consists of an ultra-low-cost computer (Raspberry Pi 4 Model B, Raspberry Pi Foundation, Cambridge CB4 0DS, United Kingdom) with a simple interface to wirelessly collect data, perform network diagnostics, and store locally on a MySQL database. This computer powers a microprocessor (Teensy 3.5, PJRC, Sherwood, OR) which stores all data received, powers a radio transceiver, and an organic light-emitting diode (OLED) screen. The OLED display shows current connectivity between the sensing units and this central hub, as well as the current time. Sensing nodes connect wirelessly connect to the radio transceiver to transmit air quality data from the sensors listed in **Table 1**. These sensors are mounted on a small, custom circuit board with a microprocessor (Teensy, PJRC, Sherwood, MN) which collects and locally stores all information to provide data redundancy. The microcontroller can read both analog and digital signals, as well as handle common communication protocols such as $I^2C$ and SPI allowing most commercially available sensors to be compatible with this system. The wireless transceiver on both sides is controlled using the SPI communication protocol, utilizing the 915 MHz band subject to the Industrial, Scientific, and Medical exemption which allows unlicensed operation of radio communication.

**Figure 1: Client monitor**

*The device is powered by batteries seen in the top right corner. The variety of sensors included connect to the microcontroller in the bottom right, which sends a signal via radio frequency to the server.*

*Server*. We amalgamated multiple monitors into a distributed sensing architecture by connecting them wirelessly to a central server. As shown in **Figure 2**, the server consists of a Teensy microcontroller connected to a Raspberry Pi. The Teensy interfaces wirelessly with the clients via radio communication, which then sends the data locally to the Raspberry Pi. The Pi stores a MySQL database for the data, as well as software to allow the user to add comments to the database as needed. SD cards are included in all steps of the process of data transfer – on each client, on the server's Teensy, and in the Raspberry Pi. This ensures that there will be no issues in retaining necessary data.

The server stores the monitor identification, timestamp, sensor readings, and error information in a database. Should the communication between the monitor and server break, or if the server fails, the data is redundantly stored at each monitor. An alarm is raised when any monitor becomes unreachable. During development, a network of six monitors were thoroughly tested in the laboratory. After the small-scale system is able to successfully transfer data to the database, we were ready to deploy and test the full-scale network.

**Figure 2: The exterior of the server**

*The OLED screen on the left displays status messages. USB, Ethernet, and HDMI ports on the right connect the Raspberry Pi to external devices.*

We diverted the project during the COVID pandemic to take a deep dive on monitoring and estimation of noise hazard maps. This project is described in **Appendix A**. This work assesses the effectiveness of three interpolation functions: a traditional Kriging model, a physics-based model, and a hybrid model that combines the two. The effectiveness of each was tested with sound data collected in four different settings. Over 10,000 model simulations were constructed from the sample data, each simulation representing different combinations and numbers of sensor training points. For each scenario, the effectiveness of each model was compared using the root mean square error at the locations between (interpolated) and outside (extrapolated) the simulated sensor locations. The hybrid model consistently reported the lowest, or tied for lowest, error in each setting overall. The Kriging model performed best for the densest networks (highest number of sensors), but the hybrid model performed within 10% of the dense Kriging network's error with less than a third of the sensors required. The hybrid model proved to be highly effective at estimating sound levels, and future work may be done utilizing the model to estimate values of other hazards such as exposure to gases and aerosols.

## Task 3. Deploy the monitoring network in a USAF facility.

*Nellis Air Force Base.* We supported the 711$^{th}$ HW field team to test the network system at Nellis AFB. First, we sent the nodes and server to the 711$^{th}$ HPW in Dayton, OH for testing. We altered the equipment due to problems with communications. This altered equipment was taken to Nellis AFB for testing on the flight line.

We had mixed results in the field. There was trouble with timestamps across the board. Field testing showed that the client's time was not correctly being updated via GPS and that some of the cables were being destroyed through repeat opening and closing of the case. To mitigate the GPS issue, code was modified to rely on the real-time clock built into the micro controller when GPS signal is unavailable. Upon connection to GPS satellites, the microcontroller will update its

timestamp based on the GPS signal and continue to update the microcontrollers time with the GPS if the signal is available.

To help reduce the risk of cables physically breaking, all components were moved to a custom designed printed circuit board (PCB). As many of the sensors as possible were moved directly to the PCB, with the remaining sensors being connected using custom made JST connectors. These connectors are more robust than the previous solution of soldering wire from pin to pin, as well as preventing any accidental unplugs due to the design of JST connectors.

Testing for the server found limited range, issues with the microcontrollers clock not updating, and not supporting the target number of monitors for Robins Air Force Base Test. The range issue was addressed in several ways. First, premade radio antennas designed for the 915 MHz band were used to improve signal strength and physical robustness of both the clients and server. Secondly, a custom PCB was designed for the server with a large grounding plane to provide more stable power to the system. Third, new radio chips were used as the initial server's radio chip ended up providing poor range due to a damaged chip.

To handle the timing issue, the client timing code was modified and used to utilize both the real-time clock and GPS to update the time. To handle 12 monitors with one server, which was the target for Robins, parameters in the radio server's application programming interface were adjusted with testing in-between to determine what would provide balance between reliable connectivity and the number of monitors. The code was also optimized to not run unnecessary functions unless radio communication with the clients was established.

*Robins Air Force Base.* We travelled with 711th HPW researchers to deploy the hazard network during a week-long field trip to Robins AFB. We travelled to the facility on Sunday. On Monday we met with management to present the protocol for sampling. On Tuesday through Thursday, we deployed the network (12 nodes and a server) around USAF cargo planes. At some locations we collocated reference equipment (**Table 2**). For each monitor, we worked with facility management to identify the nearest location to the optimal site feasible. All monitors were placed on tripods at a height of 1 to 2 meters. In the first week, we operated the network in several locations as specified in **Table 2**. The monitors were left in place in Hangar 81 for another two weeks. This length of time was to collect data sufficient to demonstrate the functionality of the novel framework.

### Table 2. Description of tests conducted

| Day / Time | Description | Details |
|---|---|---|
| Tues 08:30 – 10:00 | Boot up of system, collocation in Hangar 81 | C17 – waiting for parts |
| Tues 10:00 – 14:30 | Full network deployment in Hangar 81 | C17 – waiting for parts |
| Wed 08:00 – 08:30 | Boot up of system, collocation in Hangar 81 | C17 – waiting for parts |
| Wed 08:30 – 14:30 | Full network deployment in Hangar 81 | C17 – waiting for parts |
| Thu 08:00 – 09:45 | Full Network deployment in Hangar 81, applied sound calibrator to monitors for 10 seconds each | C17 – waiting for parts |
| Thu 10:00 – 1330 | Moved network deployment to Hangar 83 | C17 – lots of activity |
| Fri 08:00 – 13:30 | Full network deploy in C5 Hanger 3, 10:00 applied sound calibrator to monitors for 10 seconds each | C5 – lots of activity with break in middle |
| Fri 13:30 – 14:30 | Network deploy in Hangar 81 | C17 – train staff to operate network; no activity |

Upon arrival, connectivity between the 12 clients and server appeared satisfactory. The nodes were able to be placed in a way to encompass the C17/C5 aircraft being worked on allowing generation of heat maps and exposure data for the aircraft maintainers. After a few hours of testing though, the radio server froze with data no longer being collected. A two hour software reset was implemented on the server to work around this issue, causing a loss of approximately 20 seconds worth of data per reset but allowing long-term running with minimal oversight.

## Task 4. Analyze the data from the monitoring network.

We analyzed some of the massive amounts of data from Robins AFB. We designed software to analyze the hazard levels from the network and evaluate whether hazards are evenly dispersed throughout the facility. This software produces hazard maps of each contaminant.

We implemented these algorithms in the free R statistical language and MATLAB. Using data from day 4 in C5 Hanger 3, which had the most activity, the sound instruments from the monitors were compared to the reference dBadge 2. This data is displayed below in **Table 3**.

**Table 3. Comparison of sound measurement by test instruments to the reference dBadge 2**

| Monitor ID | Sensor | Day | Average Absolute Error (dBA) |
|---|---|---|---|
| 2 | Sound | 4 | 1.15 |
| 4 | Sound | 4 | 5.47 |
| 5 | Sound | 4 | 1.03 |
| 7 | Sound | 4 | 1.97 |
| 8 | Sound | 4 | 1.4 |
| 9 | Sound | 4 | 0.92 |
| 11 | Sound | 4 | 2.17 |
| 12 | Sound | 4 | 2.13 |

The time series for one monitor is also shown below in **Figure 3**. The y-axis shows the sound throughout the test period on day 4. As shown, sound levels above 100 dBA were captured around 13:00. At this noise level, it only takes 15 minutes to reach the maximum noise dosage per the guidelines set by the National Institute for Occupational Safety and Health (NIOSH) recommended exposure limit.

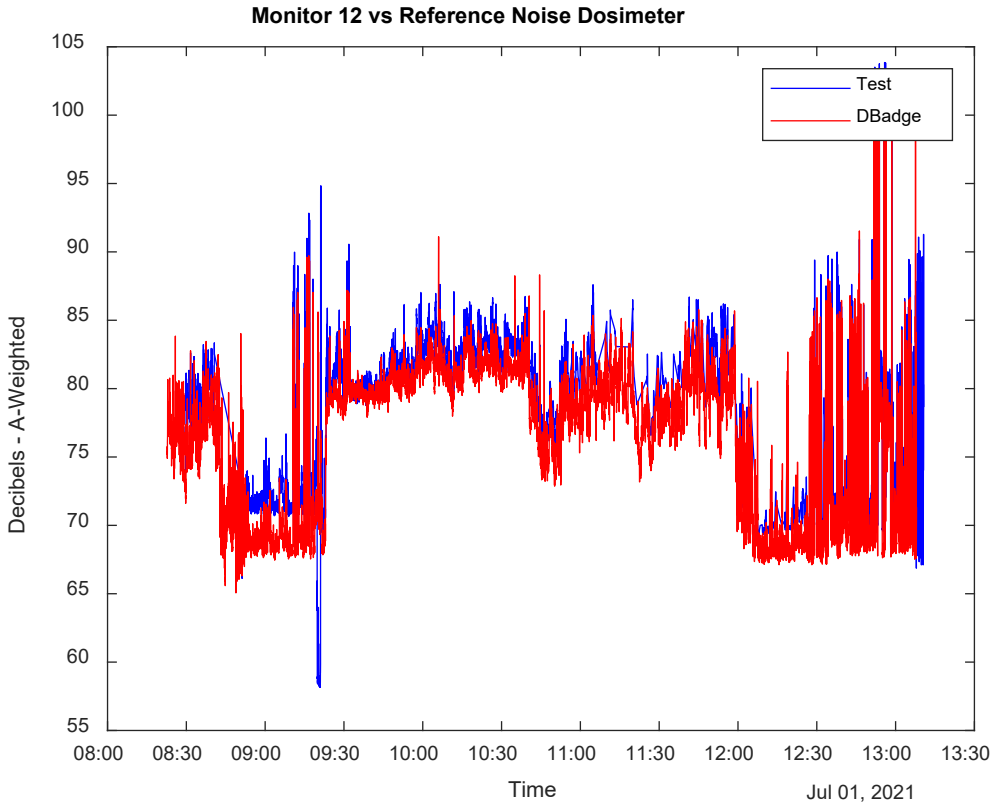



**Figure 3: Time series of noise exposure captured by both the test sensor network and the reference instrument (dBadge 2)**

These data also can be visualized through a heat map. **Figure 4**, show below, shows a heat map of the noise levels in C5 Hanger 3 at 12:53 PM with an outline of a C5 imposed on top. This shows two locations where noise exposures over 100 dBA were recorded by the sensor networks, as well as providing an estimate on how the noise disperses throughout the rest of the hanger. These maps can be generated real time to assist in rapid response to potentially hazardous

- 

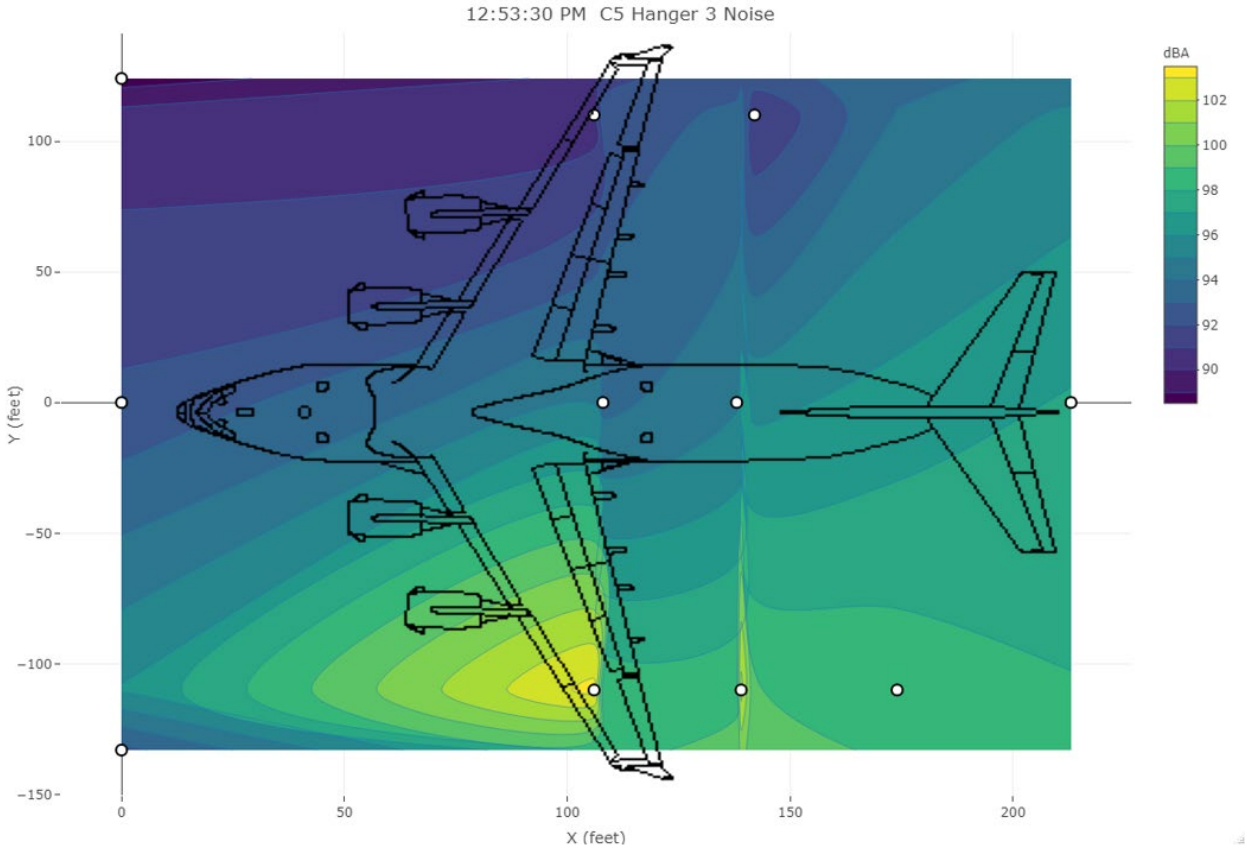



**Figure 4**: Noise map on Day 4 in C5 Hanger 3 with an outline of a C5 superimposed on the map

### Task 5. Reporting. We will provide quarterly reports and a final project report.

We provided quarterly reports throughout the project. This report serves as the final project report.

## 4.0 Conclusions and Future Work

Our multi-hazard network worked as expected, wirelessly streaming data to a central server in a USAFfacility. Our initial results show that network data compare favorably to that from the reference instruments. We have used the network to produce maps of multiple hazards present in aircraft maintaining facilities.

Future work includes the following:

- For each sensor type, develop calibration functions and conduct a loading test to determine long-term performance under simulated workplace conditions, which will be used to determine the frequency with which sensors need to be replaced. We will compile accuracy, precision, bias, and long-term performance data into a single table from which informed decisions can be made to select an optimal sensor suited to meet the workplace characteristics (e.g., hazard type, anticipated exposure levels and variability, number of monitors needed) given constraints on resources (funds for capital expenditures).
- Develop web-based software to allow remote access the database and monitors.
- On the server, store metadata, such as user-defined calibration factors, diagnostic information, and maintenance information, which can be pushed to the monitors.

Develop software to determine optimal sensor coverage, number, and sensor type for given situations in the military workplace.

# Appendix A. Combining physics-based and Kriging models to improve estimation of noise exposure


## ABSTRACT

Traditionally, worker exposure to occupational hazards is measured by equipping workers with personal wearable equipment. Hazard mapping is an alternative method of estimating workers' exposure to a variety of occupational hazards. Occupational hazards in an environment are mapped by placing stationary or mobile sensors throughout a facility and estimating hazard levels between the sensors. Complex environments may require many sensors to produce an accurate map, but sensitive interpolation functions can reduce the number of sensors required. This work assesses the effectiveness of three interpolation functions: a traditional Kriging model, a physics-based model, and a hybrid model that combines the two. The effectiveness of each was tested with sound data collected in four different settings. Over 10,000 model simulations were constructed from the sample data, each simulation representing different combinations and numbers of sensor training points. For each scenario, the effectiveness of each model was compared using the root mean square error at the locations between (interpolated) and outside (extrapolated) the simulated sensor locations. The hybrid model consistently reported the lowest, or tied for lowest, error in each setting overall. The Kriging model performed best for the densest networks (highest number of sensors), but the hybrid model performed within 10% of the dense Kriging network's error with less than a third of the sensors required. The hybrid model proved to be highly effective at estimating sound levels, and future work may be done utilizing the model to estimate values of other hazards such as exposure to gases and aerosols.


## INTRODUCTION

Wireless sensor networks can inexpensively monitor multiple environmental and occupational hazards over large areas (Zuidema, Stebounova, Sousan, et al., 2019). These hazards can include the presence of dangerous gasses, airborne particles, and noise. Tailoring the spatial and temporal patterns of sensor sampling allows the designer to refine the system so that it can produce a continuously updated hazard map that accurately reflects the hazard levels between sources of contaminants and noise (Lake, Zhu, et al., 2015). Coupling these hazard maps with worker location tracking allows individual worker exposure levels to be estimated, avoiding the more cumbersome exposure sampling method that relies on mounted sensors directly on workers (Pavón, Sigcha, et al., 2018). The accuracy of the hazard maps depends on the density of the sensors, setting up a natural design constraint of cost versus accuracy (Koehler, Volckens, 2011). Sophisticated interpolation algorithms enable a design to use fewer sensors while maintaining accuracy.

Hearing loss associated with hazardous noise levels is one of the most common illnesses caused by workplaces in the United States, according to a study by NIOSH. About 22 million US workers are exposed to hazardous noise each year (Tak, Davis, Calvert, 2009). A major step in reducing work-related hazardous noise exposure is to more effectively monitor sound levels. As with other workplace environmental hazards, noise levels can be monitored with stationary sensors, and effective interpolation methods can reduce the number of sensors required in any environment.

There are several approaches to interpolation, each providing different strengths and weaknesses. The Kriging method (Koehler, Peters, 2013) utilizes Gaussian process regression to build a highly flexible,

accurate, and computationally lightweight predictive model, especially in geostatistical implementations (Williams, Rasmussen, and Edward, 2006). When the sampling pattern is dense, a Kriging model can sensitively infer many different hazard distribution curves from the pattern observed at the sampled locations. A drawback of the Kriging model is that it is naïve to the type of information that it is modeling and will simply follow the pattern of the data. Outlier data can significantly skew the model, especially when there are few inputs. Also, Kriging models can behave erratically beyond the domain of the sampled data points, so extrapolation can be problematic compared to interpolation.

An alternative estimation method uses a physics-based model. In ideal circumstances, noise decreases inversely with distance from the sound source, following the inverse-proportional law. This law can be utilized as a linear regression model to estimate sound pressure level at a given location. The physics-based model is accurate, even with fewer inputs, because it includes underlying knowledge of the distribution of noise. This advantage is lost when the environment deviates from the ideal, because it doesn't account for the presence of sound reflections, ambient noise, or baffling.

The concept of a hybrid model combining physical processes with statistical models has been previously introduced, such as in the "physics-informed Kriging" model (Yang, Tartakovsky, Tartakovsky, 2018), which generates data from a physics-based computer simulation, then uses Kriging to estimate model parameters.

The combination of a physics-based model and statistical model like Kriging may be extended to monitoring a variety of other hazards such as gases or particulate matter in future work.

## METHODS

The study compared the three models in four experimental conditions. One experiment was conducted outdoors in a flat field without buildings or obstructions, and little outside noise. Three experiments were conducted inside of an indoor research facility. The indoor experiments varied in size of room and initial sound level. Table A1 shows the conditions for each experiment.

**Table A1. The four experimental conditions**

| Experiment name | Location | Noise Level 5 feet from source (dBA) |
|---|---|---|
| Loud outdoor | Outdoors | 101.5 |
| Large indoor | Indoor, large room | 95.2 |
| Small indoor | Indoor, small room | 95.5 |
| Quiet indoor | Indoor, large room | 88.1 |

### Sound Monitor

Sound pressure level was measured with a custom sensor we had previously implemented in a multi-sensor wireless network (Zuidema, Stebounova, Sousan, et al., 2019), seen below in Figure A1. The sound sensor consists of components that cost about $30. It provides similar readings to a Type 2 reference sound level monitor that cost $1,800 (Hallett, Tatum, et al., 2018). The sensor measures the A-weighted sound level in decibels, which describes sound pressure level weighted to the level of human

hearing, every two seconds. These sensors have an accuracy of +- 2 dBA. In addition to the sound level meter (SLM), the monitor records information for various other hazards, including CO, $O_3$, temperature and relative humidity, and VOCs.

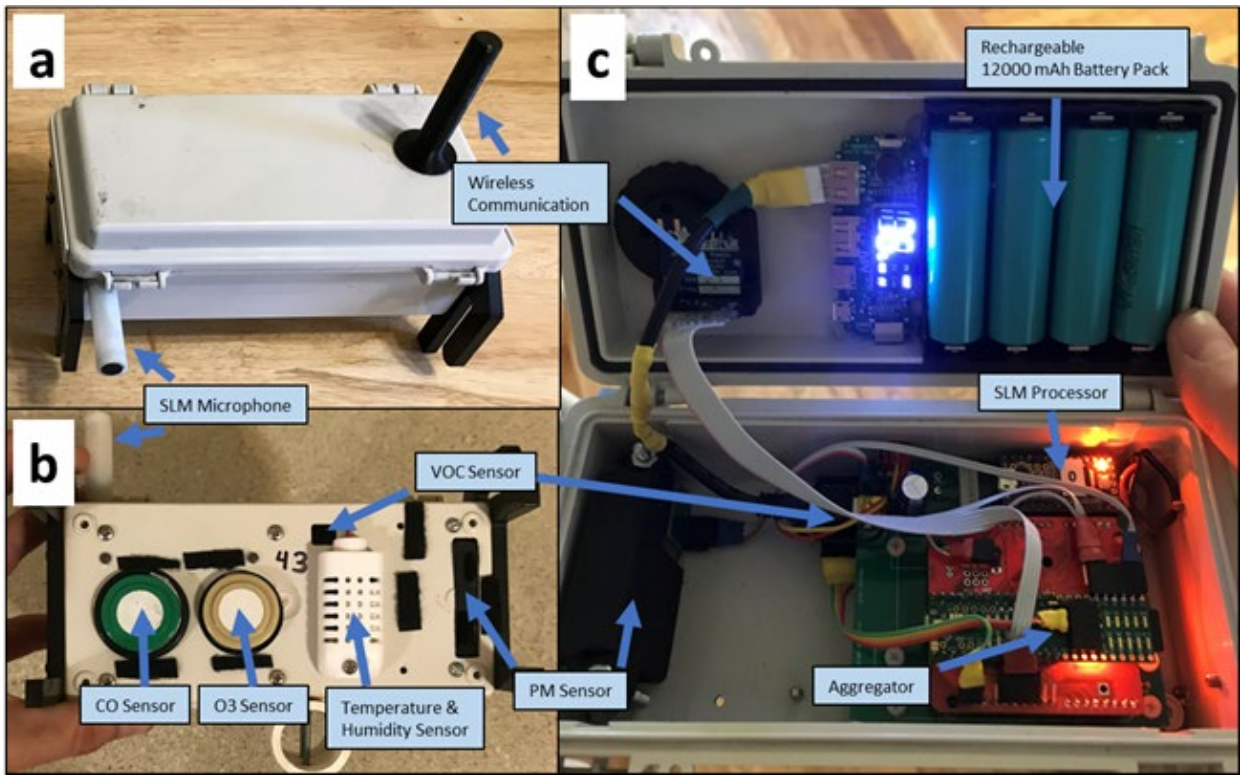



**Figure A1: The monitor used in this experiment. The sound level meter (SLM) is part of a multi-hazard wireless sensor device, which includes a variety of other sensors.**

## Data Collection

For each experiment, a sound amplifier and speaker (G20, Crate) driven by a white noise generated was placed approximately two feet above the ground. The sound monitor collected sound level readings at various distances directly in front of the speaker. Sampling intervals started five feet from the sound source and increased with distance, starting with two-foot increments, then five-foot, ten-foot, and for the loud outdoor experiment – 20-foot increments. At each sampling distance, five samples were collected, each averaged over a 10-second interval.

## Models

### Physics-based model

The physics-based model derives from equation 1, the theoretical propagation of sound pressure level (Hansen, 2001).

$$L_{P2} = L_{P1} - 20 * \log_{10}(\frac{r_2}{r_1}) \quad (1)$$

Here $L_{P1}$ and $L_{P2}$ are the sound pressure level at two positions, and $r_1$ and $r_2$ are the distance from the source at which the sound pressure is measured. Given one input measurement, say $L_{P1}$ at distance $r_1$, equation 1 predicts the sound pressure level $L_{P2}$ at any other position, $r_2$. Separating the $r$ variables, the equation can be expressed as

$$L_{P2} = L_{P1} - 20 * \log_{10}(r_2) + 20 * \log_{10}(r_1) \quad (2)$$

This indicates that the sound at any distance, $L_{P2}$, can be calculated from the features of another reference point, $L_{P1}$ and $r_2$. Combine the terms of the reference point into a single variable, $b$.

$$b = L_{P1} + 20 * \log_{10}(r_1) \quad (3)$$

This term may be estimated from multiple points by:

$$\hat{b} = \frac{\sum(L_{P1}^i + 20 * \log_{10}(r_1^i))}{n} \quad (4)$$

The superscript $i$ represents the sample index from a total $n$ samples.

## *Kriging model*

The Kriging model is based on a Gaussian process regression, which uses a Gaussian process to estimate values based on prior covariances. The Gaussian process can be expressed as a normal distribution of mean 0, and standard deviation represented by a covariance matrix determined by a given kernel function and input data, $\boldsymbol{X}$ and $\boldsymbol{X}_*$.

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{f}_* \end{bmatrix} \sim N(0, \begin{bmatrix} K(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I}_n & K(\boldsymbol{X}, \boldsymbol{X}_*) \\ K(\boldsymbol{X}_*, \boldsymbol{X}) & K(\boldsymbol{X}_*, \boldsymbol{X}_*) \end{bmatrix}) \quad (5)$$

where $\boldsymbol{X}$ is the training data input vector, $\boldsymbol{y}$ is the training data response vector, $\boldsymbol{X}_*$ is the prediction data input, $\boldsymbol{f}_*$ is the prediction output, $\sigma$ is the variance of the measurement noise, $\boldsymbol{I}_n$ is the $n \times n$ identity matrix, and $K(\cdot,\cdot)$ is a kernel function.

In this case, K represents the squared exponential kernel, expressed in equation (6).

$$K(\boldsymbol{X}, \boldsymbol{X}_*) = \alpha^2 \exp(-\frac{(\boldsymbol{X} - \boldsymbol{X}_*)^2}{2l^2}) \quad (6)$$

Where $l$ is the length-scale, which is a user-input parameter that determines how aggressively the function adjusts to new data.

Equation 4 can be rewritten in terms of $\boldsymbol{L}_p$ and $\boldsymbol{r}$ as:

$$\begin{bmatrix} \boldsymbol{L}_{P1} \\ \boldsymbol{L}_{P2} \end{bmatrix} \sim N(0, \begin{bmatrix} K(\boldsymbol{r}_1, \boldsymbol{r}_1) & K(\boldsymbol{r}_1, \boldsymbol{r}_2) \\ K(\boldsymbol{r}_2, \boldsymbol{r}_1) & K(\boldsymbol{r}_2, \boldsymbol{r}_2) \end{bmatrix}) \tag{7}$$

where $\boldsymbol{L}_{P1} = \left[L_{p1}^1, \cdots, L_{p1}^n\right]^T$ and $\boldsymbol{r}_1 = [r_1^1, \cdots r_1^n]^T$ are the training response and input, respectively. $\boldsymbol{r}_2$ is the set of locations at which the sound pressure levels are predicted. Given this formulation, the distribution of $\boldsymbol{L}_{P2}$ is:

$$\boldsymbol{f}_*|\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X}_* \sim N(\bar{\boldsymbol{f}}_*, cov(\boldsymbol{f}_*)) \tag{8}$$

Or, rewritten with the relevant variables:

$$\boldsymbol{L}_{P2}|\boldsymbol{r}_1, \boldsymbol{L}_{P1}, \boldsymbol{r}_2 \sim N(\mu_2, \sigma_2^2) \tag{9}$$

where

$$\begin{aligned} \mu_2 &= K(\boldsymbol{r_1}, \boldsymbol{r}_2)[K(\boldsymbol{r}_1, \boldsymbol{r}_1) + \sigma^2 \boldsymbol{I}_n]^{-1} \boldsymbol{y} \\ \sigma_2^2 &= K(\boldsymbol{r}_2, \boldsymbol{r}_2) - K(\boldsymbol{r}_1, \boldsymbol{r}_2)[K(\boldsymbol{r}_1, \boldsymbol{r}_1) + \sigma^2 \boldsymbol{I}_n]^{-1} K(\boldsymbol{r}_1, \boldsymbol{r}_2) \end{aligned} \tag{10}$$

Where $\mu_2$ represents the predicted mean, and $\sigma_2^2$ represents the prediction variance.

## Hybrid model

Finally, the hybrid model uses elements of both the physics-based model and the Kriging model. The key difference between the proposed hybrid model and the Kriging model is that the training data for hybrid model is the difference between the physics-based model and the training points themselves. In this way, the training of the hybrid model takes both the data and physics model into consideration. More specifically, the prediction result for the hybrid model is as follows:

$$\boldsymbol{L}_{hybrid} = K(\boldsymbol{r}_1, \boldsymbol{r}_2)[K(\boldsymbol{r}_1, \boldsymbol{r}_1) + \sigma^2 \boldsymbol{I}_n]^{-1}\left(\boldsymbol{y} - \hat{L}_{p1}\right) + \hat{L}_{p2} \tag{11}$$

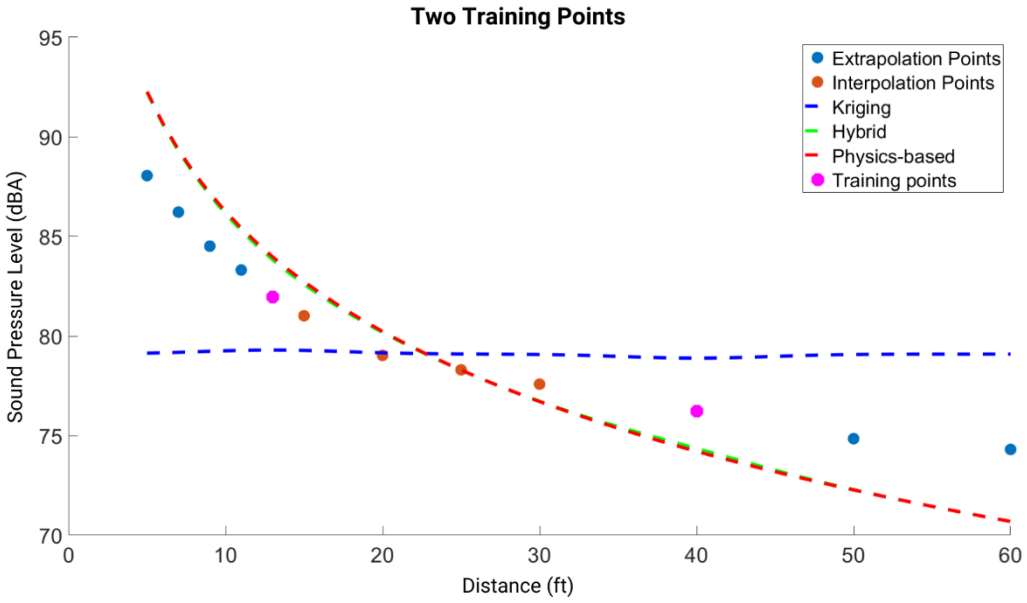
where $\hat{L}_{p1}$ is the data estimate predicted by the physics-based model, i.e., plugging $\hat{b}$ and $\boldsymbol{r}_1$ into Equation 3. $\hat{L}_{p2}$ is the prediction result for $\boldsymbol{r}_2$ by using the estimated physics-based model. Clearly, the $\boldsymbol{L}_{hybrid}$ considers the data $y$ and physics-based model simultaneously.

To demonstrate how the models fit different inputs, Figure A2 (a)-(d) illustrates all three models for data sampled from the quiet indoor experiment. Each graph displays the three model fits when 2, 4, 6, and 8 points are selected as training points for the model. The other data from the test are indicated as either extrapolation points if they are outside the domain of the training points, or interpolation points if they are inside the domain of the training points.

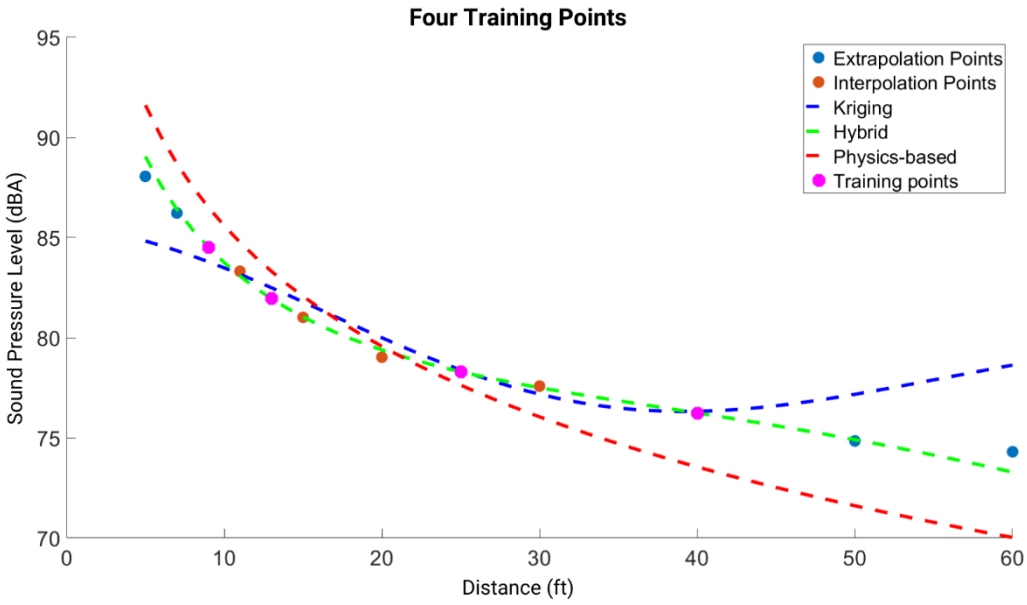
Figure A2 shows how the different numbers of training points affect the various models' fits. In Figure A2 (a), the Kriging model is essentially a flat line at the mean of the two points, because there is too little information for it to distinguish a pattern. As more training points are available for the Kriging model, it adjusts to the shape of the data. In Figure 2 (d), the Kriging model intersects all training points. It is very close to most of the test points, except for the most distant extrapolation point. When extrapolating, the model tends to regress towards the mean.

The physics-based model has the same approximate shape in all circumstances, because it rigidly assumes the same data distribution. The steepness of its curve varies with the training points. For example, the falloff in Figure 2 (d) is less sharp than in the other figures.

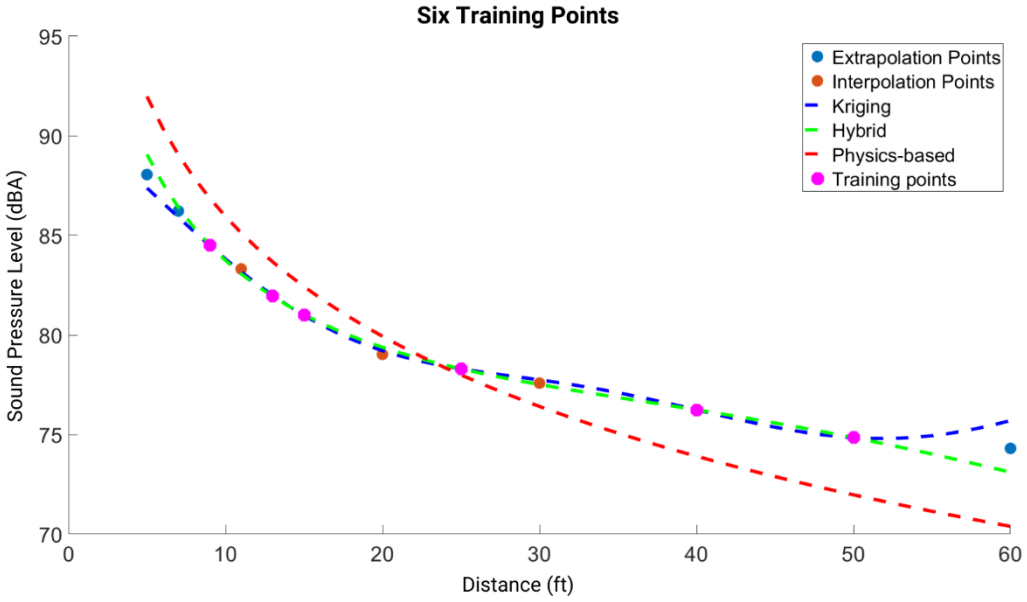
When there are few training points, the hybrid model closely follows the physics-based model, as in Figure 2 (a). In Figure 2 (b) the hybrid model clearly demonstrates the same rough shape as the physics-based model, but with more flexibility, more closely following data. By Figure 2 (d), when there is more data available, the hybrid closely resembles the Kriging model, except in the extrapolation regions, where it follows the shape of the physics-based model.



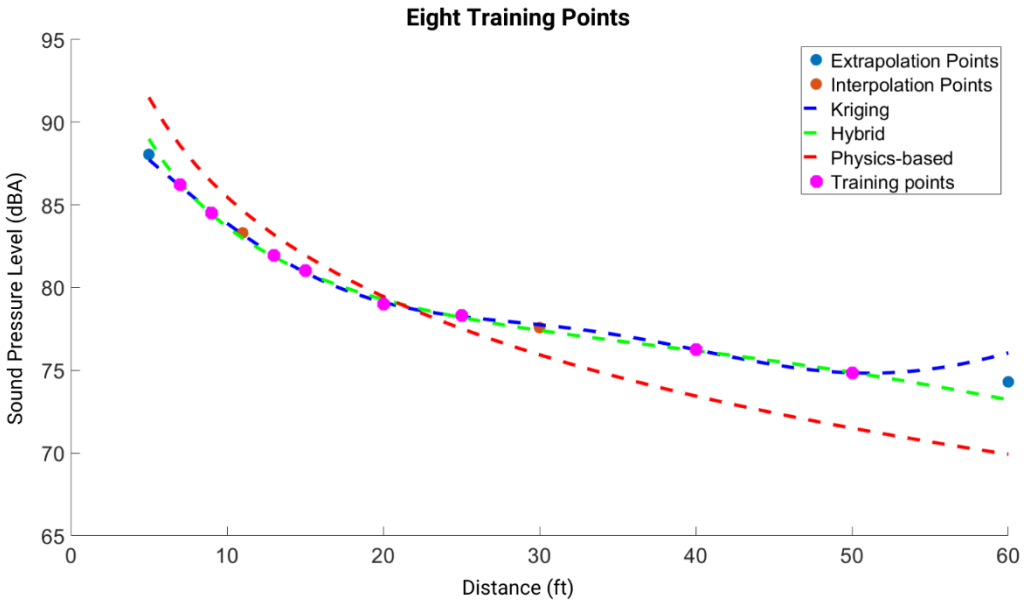

(a): Two training points.




(b): Four training points.




(c): Six training inputs.




(d): Eight training inputs.

**Figure A2 (a)-(d). The models fit on sample data from the quiet indoor data set, simulating 2, 4, 6 and 8 monitoring locations providing the model inputs. The training points are magenta, the interpolated points are red, and the extrapolated points are blue.**

## Model Assessment

The quality of fit for each model was measured as the root mean square error (RMSE) on test points. RMSE was averaged among all test points for a given set of data. In order to thoroughly evaluate the model capabilities, it is helpful to have a variety of training and testing points.

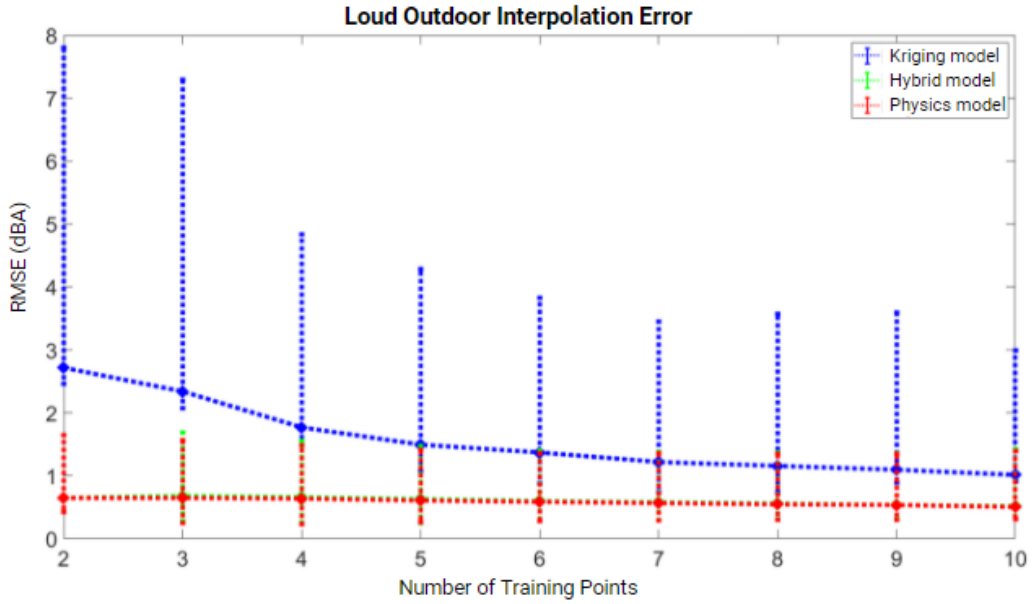
Each model was fit to the test data in a series of model evaluation trials. In each trial, a given data set was divided into a set of $n$ training points used to fit the data. The remaining data samples were used as test locations to evaluate the model fit. Thus, each trial represents an individual arrangement of sensors at the training locations; the quality of the model fit for each trial is measured as the model error at the remaining positions, or test points. Every possible combination of training points were considered in a separate trial. For example, with 12 data points, when testing the models with 3 training points, there are $\binom{12}{3}$ or 220 possible arrangements of training points. The trials also differed in the number of locations used as inputs. This represents scenarios in which more or fewer monitors were used to measure the sound.

For each experiment, error is assessed as either interpolation error or extrapolation error. Interpolation error represents RMSE that was tested on points within the bounds of the training points, while extrapolation error represents RMSE tested on points outside of the bounds of training points.

## RESULTS

Figures present the average RMSE values in dBA for each of the three models with various numbers of training points. The blue markers represent the Kriging model. The green markers represent the hybrid model. The red markers represent the physics-based model. The colored circles show the mean RMSE, and the vertical error bars show the $10^{th}$ and $90^{th}$ percentiles of the RMSE for the given model and number of inputs. For each experiment, there are two figures: interpolation error and extrapolation error. The interpolation error is the average RMSE for the test points between the input locations for each trial. The extrapolation error is the RMSE for test points closer than the nearest input position or farther than the farthest input position.

Figure A3 illustrates the three model fits for the loud outdoor data set. The physics-based and hybrid models have low and nearly identical errors, less than 1 dBA, regardless of the number of input locations, for both the interpolated and extrapolated data. The confidence intervals never exceeding 2 dBA. The Kriging model is less accurate, with RMSE of about 3 and large confidence intervals, particular with few training points. The RMSE slowly improves as number of training points increases, reaching RMSE of about 1 for the interpolation error and about 4 for the extrapolation error.



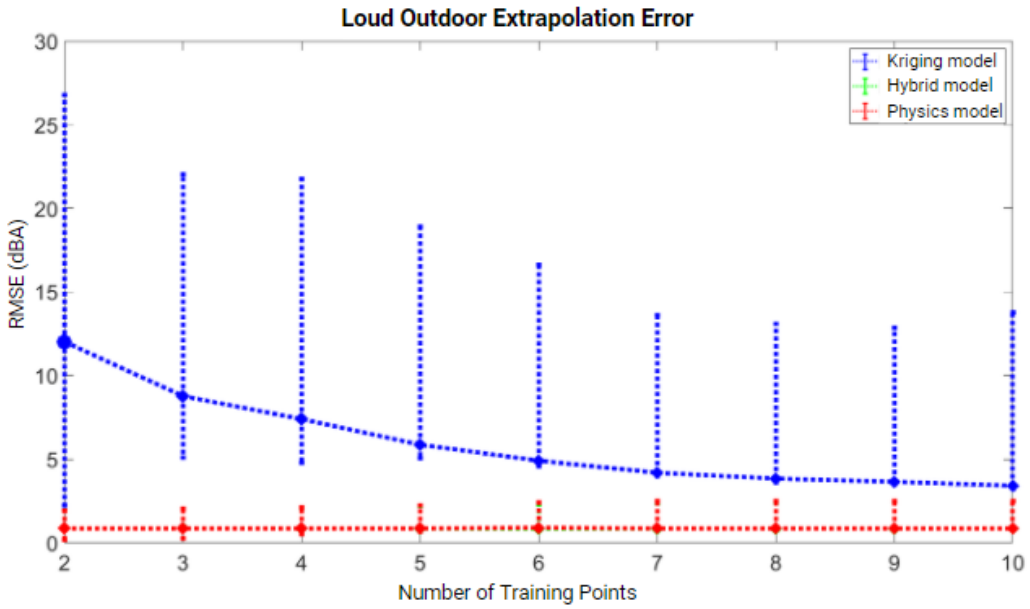



**Figure A3. Average RMSE (dBA) at the interpolated (a) or extrapolated (b) testing points for the three models using the loud outdoor data set. The error bars represent the 10%-90% confidence interval across every permutation of training points, given the particular number of inputs. The top shows the error just of the interpolated point, the bottom for the extrapolated points.**

Figure A4 displays similar plots for the large indoor data set. The pattern is more complex than the outdoor data set. The interpolation error is less than 1.5 RMSE with 2 training points for all three models. Both the hybrid and Kriging models become more accurate as number of training points increase. The physics-based model is most accurate with few training points, but becomes less accurate with more training points. The extrapolation data is similar, but has a higher RMSE.

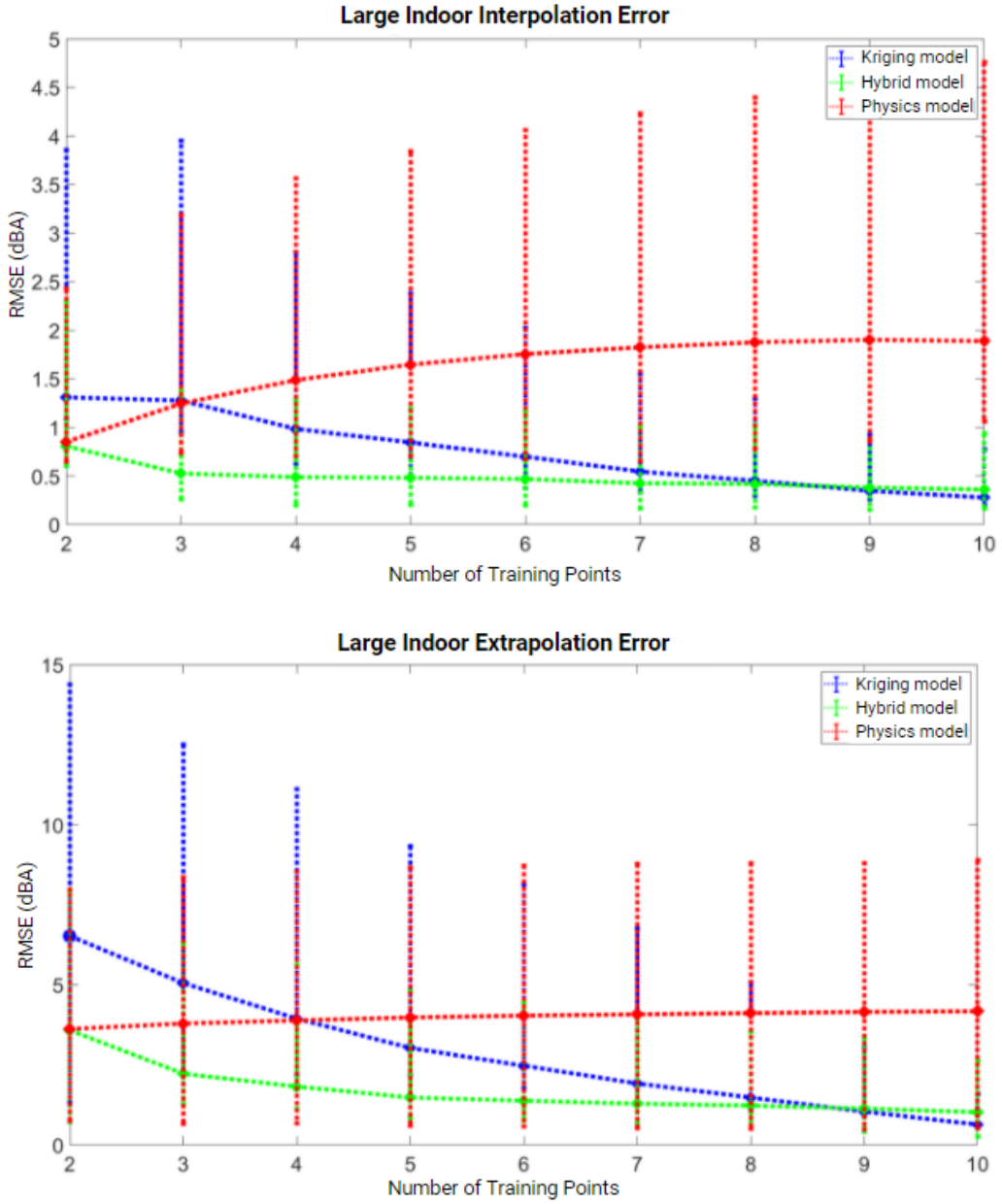



**Figure A4. RMSE vs number of training points for each type of model for the large indoor experiment.**

Figure A5 provides similar data for the small indoor data set. The patterns are similar to the large indoor data set. Again, the hybrid model is the most accurate in each condition. The physics model is the most accurate with few training points and decreases with more points. The data-driven model is least accurate with few points and improves with more training points. The physics model tends to be better than the hybrid model when extrapolating data, until there are many training points.

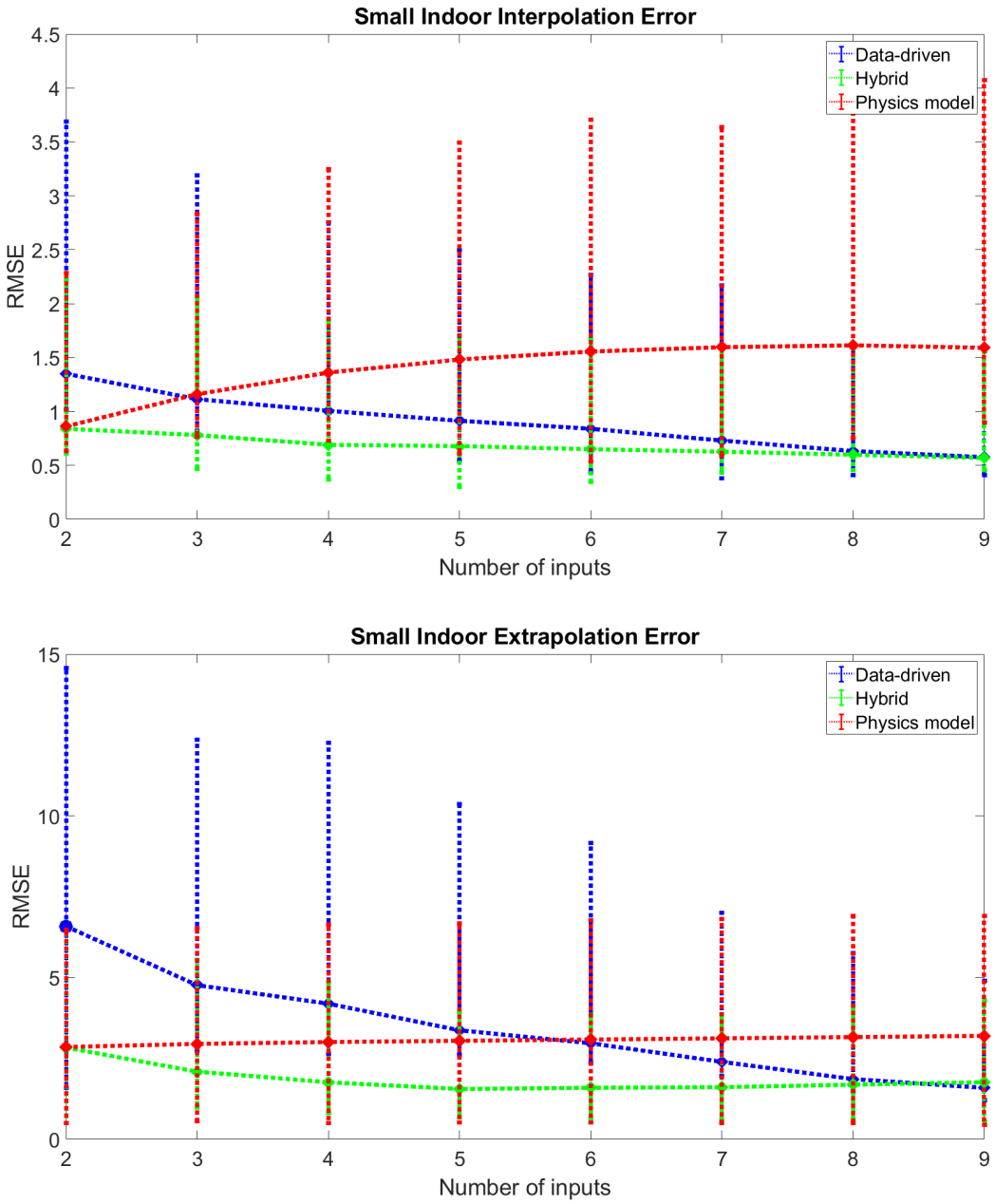



**Figure A5. RMSE vs number of training points for each type of model for the small indoor experiment.**

Figure A6 shows the errors for the quiet indoor data set. The pattern is consistent with the other indoor training sets. Again, the hybrid model is among the most accurate in all conditions. The physics model is initially among the most accurate, but is quickly surpassed by the data-driven model with just three or four training points.

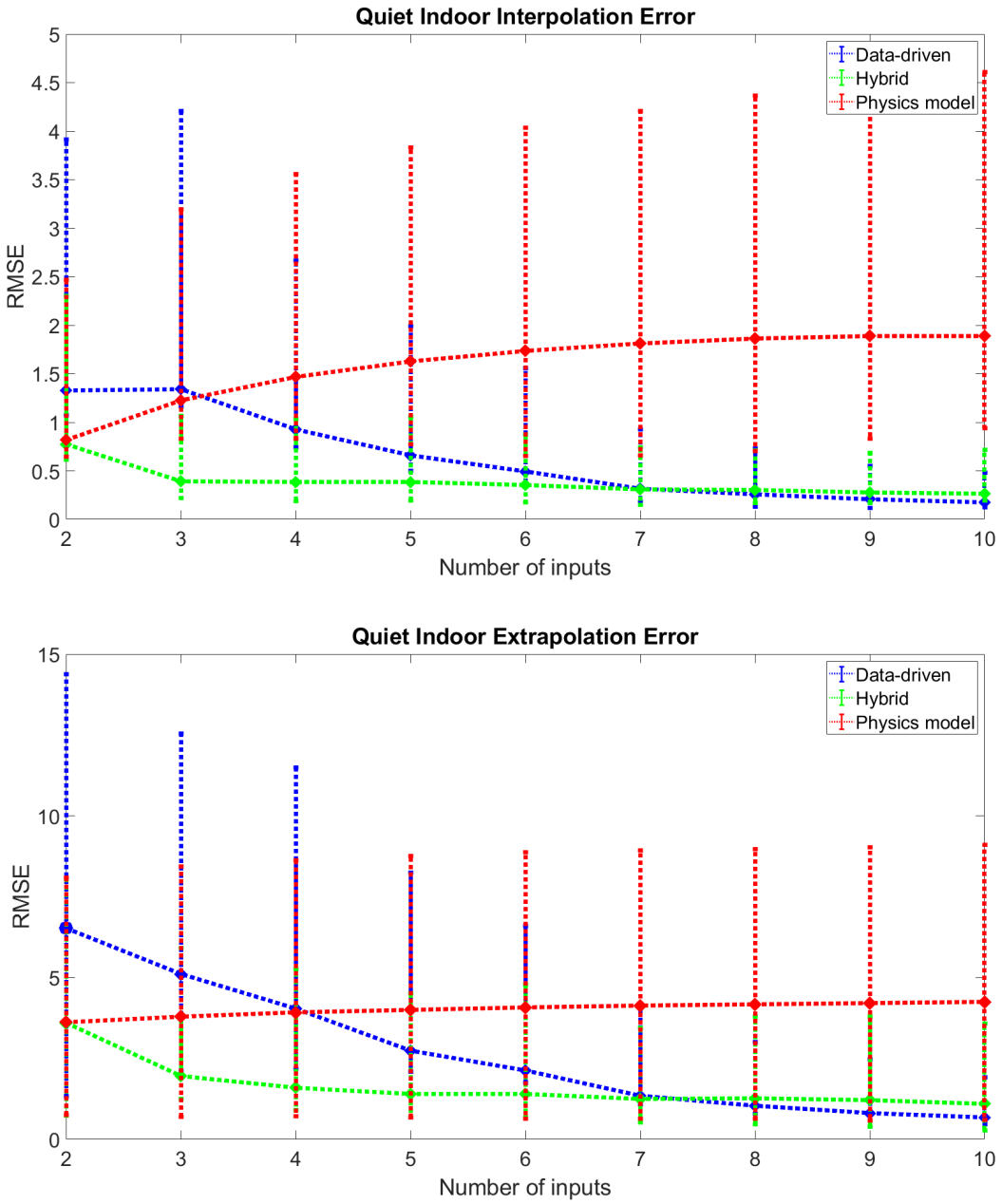



**Figure A6. RMSE vs number of training points for each type of model for the quiet indoor experiment.**

## DISCUSSION

The study compared the accuracy of three models: a Kriging model, a physics-based model built on the inverse-proportional law, and a hybrid model that uses Kriging with information from the physics-based model. Interpolation and extrapolation were evaluated separately. It is likely that most sensor networks will be designed to emphasize interpolated values, but placing sensors at the outer boundaries of

every region of interest can be expensive, so the stability of a model's extrapolation of data is important to consider where sensor deployments are limited. This paper hypothesizes that Kriging will be highly effective in a dense wireless network (e.g., when there are many sensors), but the hybrid model will have the highest accuracy with fewer sensors. Data sparsity is one of the pitfalls in hazard mapping (Koehler, Volckens, 2011), and the hybrid model may help to overcome that pitfall.

The hybrid model shows the most flexibility of the three models by consistently showing the lowest or within a 10% margin of the lowest RMSE in all conditions. In many cases, especially with fewer training points, significantly outperforming one or both of the other models.

For the interpolated outdoor, loud data, which closely follows the assumptions of the physics model (Figure 4), the hybrid and physics-based models perform very well, with both models having an average of less than 1 dBA of error in all conditions and confidence intervals never exceeding 2 dBA. The Kriging model performs relatively poorly for lower number of inputs, and better at higher number of inputs, with the average model error moving from about 3 dBA to 1 dBA RMSE. The large confidence interval on the Kriging model suggests that the model sometimes fits very poorly. This outdoor, loud experiment was specifically designed to be favorable to the physics model by including little to no background noise or reflected sound. Not surprisingly, the physics-based model performed well and the hybrid model, which uses the physics-based model as input, did not deviate far from the physics-based model because the residual errors were so small. The physics-based and hybrid model performance was excellent with just two inputs and did not improve much with more inputs.

Figures 5-7 show the results of the indoor data sets. In these conditions, reflection of noise from walls in the room cause the sound to decrease irregularly, making it more difficult for the physics-based model to predict. In these cases, a similar pattern is evident: the hybrid and physics-based models start strong while the Kriging model starts poorly with fewest training points. And as training points increase, the Kriging and hybrid models improve while the physics-based model either worsens or remains the same. At ten training points, the Kriging model typically has the lowest RMSE, although by a narrow margin compared to the hybrid model. For the extrapolation results, the Kriging model tends to be less accurate than both the physics-based and hybrid models, especially as the range of extrapolation increases. This is because the algorithm tends to regress toward the mean in the absence of new data. The physics model presumes that the extrapolated data will tend to diminish with distance from the source.

When provided with between 3 and 7 inputs, the hybrid model is accurate in each circumstance, demonstrating the flexibility that makes this model attractive. When using five or fewer sensors, the hybrid model is at least 50% more accurate than the Kriging model for interpolation, and even more so for extrapolation. In densely sampled conditions, when there are 8 or more inputs for the indoor experiments, the hybrid and Kriging models have RMSE errors within 5% of one another. When the conditions follow the assumptions of the physics-based model, as in the outdoor experiment, the hybrid model offers substantial improvement over the Kriging model even at the highest numbers of inputs and is quite accurate even with only two to three sensors.

## CONCLUSION

The hybrid model consistently demonstrated low RMSE for interpolation and extrapolation in the variety of conditions presented, especially with fewer training points. The model was shown to perform at least as well as both the Kriging and physics-based models at their best, demonstrating high flexibility. In almost all conditions and numbers of inputs, the hybrid model improved upon the physics-based model. When the Kriging model performs at its best, with number of sensors of 9-10, the hybrid model performed within 10% accuracy with only 3-4 sensors required.

Implementing the hybrid model into a hazard mapping process can reduce costs by reducing the number of monitors required for effective mapping. In the one-dimensional case, the hybrid model achieved results similar to those of the Kriging model with just one third of the number sensors. In a two-dimensional mapping situation, this savings may change like the square of the one-dimensional case, yielding nearly an order of magnitude improvement. The hybrid model may also generalize to other hazard monitoring applications, including different hazards, or additional sources of hazards, offering increased opportunities to reduce cost and increase efficiency.

# REFERENCES


Hallett, L., M. Tatum, G. Thomas, S. Sousan, K. Koehler and T. Peters (2018). "An inexpensive sensor for noise." Journal of occupational environmental hygiene **15**(5): 448-454.

Hansen, C. H. (2001). "Fundamentals of acoustics." Occupational Exposure to Noise: Evaluation, Prevention and Control. World Health Organization: 23-52.

Koehler, K. A. and T. M. Peters (2013). "Influence of analysis methods on interpretation of hazard maps." Annals of occupational hygiene **57**(5): 558-570.

Koehler, K. A. and J. Volckens (2011). "Prospects and Pitfalls of Occupational Hazard Mapping:'Between These Lines There Be Dragons'." Annals of Occupational Hygiene **55**(8): 829-840.

Lake, K., J. Zhu, H. Wang, J. Volckens and K. A. Koehler (2015). "Effects of data sparsity and spatiotemporal variability on hazard maps of workplace noise." Journal of occupational and environmental hygiene **12**(4): 256-265.

Pavón, I., L. Sigcha, P. Arezes, N. Costa, G. Arcas and J. López (2018). Wearable technology for occupational risk assessment: Potential avenues for applications. Occupational safety and hygiene VI, CRC Press**:** 447-452.

Tak, S., R. R. Davis and G. M. Calvert (2009). "Exposure to hazardous workplace noise and use of hearing protection devices among US workers—NHANES, 1999–2004." American journal of industrial medicine **52**(5): 358-371.

Thomas, G. W., S. Sousan, M. Tatum, X. Liu, C. Zuidema, M. Fitzpatrick, K. A. Koehler and T. M. J. S. Peters (2018). "Low-cost, distributed environmental monitors for factory worker health." **18**(5): 1411.

Williams, C. K. and C. E. Rasmussen (2006). Gaussian processes for machine learning, MIT press Cambridge, MA.

Yang, X., G. Tartakovsky and A. Tartakovsky (2018). "Physics-informed kriging: A physics-informed Gaussian process regression method for data-model convergence." arXiv preprint arXiv:1809.03461.

Zuidema, C., L. V. Stebounova, S. Sousan, A. Gray, O. Stroh, G. Thomas, T. Peters and K. Koehler (2019). "Estimating personal exposures from a multi-hazard sensor network." Journal of Exposure Science & Environmental Epidemiology.

# LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| AFB | Air Force Base |
| CO | carbon monoxide |
| $CO_2$ | carbon dioxide |
| dBA | A-weighted sound level in decibels |
| DRI | direct reading instruments |
| $eCO_2$ | equivalent calculated carbon dioxide |
| MHz | megahertz |
| NIOSH | National Institute for Occupational Safety and Health |
| $O_3$ | ozone |
| OEL | Occupational Exposure Limits |
| OLED | organic light-emitting diode |
| PCB | printed circuit board |
| PM | particulate matter |
| ppb | parts per billion |
| ppm | parts per million |
| RMSE | root mean square error |
| SLM | sound level meter |
| TVOC | total volatile organic compound |
| TWA | time weighted average |
| VOC | volatile organic compound |