Deploying and Analyzing Air Quality Sensors in Mongolian Gers

Lehi Sttenio Alcantara
Brigham Young University

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Engineering Commons

BYU ScholarsArchive Citation
https://scholarsarchive.byu.edu/etd/8908

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.
ABSTRACT

Deploying and Analyzing Air Quality Sensors in Mongolian Gers

Lehi Sttenio Alcantara
School of Technology, BYU
Master of Science

The purpose of this research is to develop best practices for deploying air quality sensors in a remote location such as Mongolia. I discussed the architecture and design constraints when collecting remote air quality sensors data, the challenges that emerge while implementing a sensor-based network in a remote location such as Mongolia. The tradeoffs of using different architectures are described.

I observed the usage of electrical heaters in modified gers in remote locations and conclude how effective they are in reducing PM$_{2.5}$ levels by analyzing air quality data and go through the process of cleaning up the data and removing humidity from low-cost sensors used to deploy in a remote location such as Mongolia so that the PM$_{2.5}$ reading is more accurate.

In order to help many humanitarian efforts dealing with better air quality in developing countries, an air quality sensor was designed to keep low cost as much as possible. The cost is about $200 to build, which is cheaper than other low-cost sensors, yet provides more functionality (e.g., CO$_2$ sensing) and used cellular connectivity to upload data in real-time. This sensor has implications beyond Mongolia. The sensor can be used anywhere WiFi connectivity is not available, such as parks, bus stops, and along roadways, breaking the constraints that other low-cost sensors have. Removing the need for WiFi is a necessary step in allowing ubiquitous air quality sensing.

The contributions in this thesis are: First, I presented the challenges one should consider while deploying air quality sensors in developing countries. Second, since Mongolia offers a unique environment and constraints, I shared experiences in deploying sensors in a remote location like Mongolia. This experience goes beyond air quality sensors and can inform anyone who is deploying sensors in remote areas. Third the analysis of the PM$_{2.5}$ on the gers gives us better insights as to whether modifying gers with insulation and using electrical heaters as opposed to burning coal to heat up the gers makes a difference in regard to better air quality in the gers.

Keywords: wireless sensor networks, remote sensing, system applications and experience, data analysis, internet of things, air quality
ACKNOWLEDGMENTS

I would like to thank all IT faculty members, especially committee chairs Dr. Hansen and Dr. Teng for their various inputs and directions in developing this thesis. Also, a big thank you goes to Dr. Lundrigan from Electrical and Computer Engineering Department for his mentorship and guidance throughout this journey. The NET (Network Enhanced Technologies) Lab team was a great experience and I’m grateful to be part of it and the chance to work together in various tasks. Dennis Eggett from the Center for Collaborative Research and Statistical Consulting from Brigham Young University for his statistics expertise contribution in this thesis. Finally, I would like to express my gratitude to my wife, Liesel, for her patience and support in all my education goals throughout the years I have been in school.
# TABLE OF CONTENTS

LIST OF TABLES ............................................................................................................................... vi
LIST OF FIGURES ............................................................................................................................. vii

1 Introduction ........................................................................................................................................ 1
  1.1 Key Terms ......................................................................................................................................... 3
  1.2 Research Objectives and Questions ............................................................................................... 4
  1.3 Outline of Thesis .............................................................................................................................. 4

2 Literature Review ............................................................................................................................... 6
  2.1 Air Quality Health Effects and Measurement ............................................................................... 6
  2.2 History and Efforts to Regulate PM$_{2.5}$ Levels ........................................................................... 7
  2.3 Air Quality Sensor Deployment ..................................................................................................... 12
  2.4 Deploying Sensors in Mongolia .................................................................................................... 13
    2.4.1 Infrastructure Challenges ........................................................................................................ 14
    2.4.2 Skilled Professionals & Cost .................................................................................................. 15
    2.4.3 Security & Reliability ............................................................................................................. 15
    2.4.4 Mongolia Geography & Deployment Challenges ................................................................. 15
  2.5 Mongolian Modified Ger Project ................................................................................................. 16
  2.6 Summary ........................................................................................................................................ 16

3 Methodology ....................................................................................................................................... 18
  3.1 Field Deployment Methodology .................................................................................................... 18
  3.2 Data Analysis Methodology ........................................................................................................... 20
  3.3 Air Quality Sensors Deployment and Analysis Replicability ....................................................... 25
    3.3.1 Accuracy and Precision of Air Quality Sensor Sensirion SPS30 ............................................. 25
      3.3.1.1 Laboratory Evaluation of Sensirion SPS30 .................................................................. 26
      3.3.1.1.1 Accuracy (PM$_{2.5}$) ............................................................................................. 26
      3.3.1.1.2 Precision (PM$_{2.5}$) ............................................................................................. 27
      3.3.1.1.3 Coefficient of Determination ................................................................................. 28
      3.3.1.2 Field Evaluation of Sensirion SPS30 .............................................................................. 29
  4 Mongolian Field Deployment & Architecture .................................................................................. 31
    4.1 Q1A: Design Tradeoffs of Using Different Wireless Architectures .......................................... 31
    4.1.1 Mongolia Air Quality Deployment Architecture Ratings Explanation ................................ 34
    4.2 Q1B: Sensor-Based Network in Mongolian Gers Challenges ............................................... 38
      4.2.1 Challenges & Lessons Learned from Deployment ............................................................... 40
        4.2.1.1 Unplugged Sensors ....................................................................................................... 40
4.2.1.2 WiFi vs Cellular vs LoRa ................................................................. 40
4.2.1.3 Management Tools............................................................................. 41
5 Air Quality Intervention Analysis ............................................................... 42
  5.1 Q2A: Analyzing Effective of Modified Gers in Reducing PM$_{2.5}$ ................ 43
  5.2 Q2B: Temperature Affect PM$_{2.5}$ Gains on Modified Gers...................... 43
  5.3 Q2C: Day of Week and Time of Day Affect PM$_{2.5}$ Gains on Modified Gers 45
    5.3.1 Hourly Median PM$_{2.5}$ for Unmodified Gers .................................... 45
    5.3.2 Hourly Median PM$_{2.5}$ for Modified Gers ....................................... 46
6 Discussion ................................................................................................. 50
  6.1 Summary .............................................................................................. 50
  6.2 Limitations .......................................................................................... 52
  6.3 Future Work ......................................................................................... 53
References .................................................................................................... 54
Appendix A: Supplemental Materials .......................................................... 61
LIST OF TABLES

Table 2-1: WHO Interim PM$_{2.5}$ Levels and Countries That Follow It........................................... 9
Table 2-2: WHO Short Term (24 hr) PM$_{2.5}$ Levels and Countries That Follow It .................... 10
Table 2-3: Consideration of Deploying Sensors Network in Developing Countries...................... 14
Table 4-1: Developing Countries Air Quality Deployment Architecture Ratings ....................... 33
Table 4-2: Circle’s Legend .............................................................................................................. 33
Table 5-1: All Temperature Ranges PM$_{2.5}$ Estimates for Unmodified/Modified Gers ............ 43
Table 5-2: PM$_{2.5}$ Estimates for Modified/Unmodified Gers by Temperature Range ............... 45
Table 5-3: Grand Total Difference for PM$_{2.5}$ Average Hourly Median................................. 47
Table 5-4: PM$_{2.5}$ Estimates for Modified/Unmodified Gers by Hour ....................................... 48
Table 5-5: PM$_{2.5}$ Estimates for Unmodified/Modified Gers by Time Buckets ......................... 49
LIST OF FIGURES

Figure 2-1: PM$_{2.5}$ Trend in the USA ........................................................................................................ 9
Figure 3-1: All Sensors Graph by Hourly Median ......................................................................................... 23
Figure 3-2: Grafana Heatmap with Hourly Median for All Sensors ............................................................. 23
Figure 3-3: Sensirion SPS30 ......................................................................................................................... 26
Figure 3-4: PM$_{2.5}$ Accuracy Between Sensirion ............................................................................................ 26
Figure 3-5: PM$_{2.5}$ Precision Between Sensirion SPS30 Vs FEM GRIMM ................................................... 27
Figure 3-6: Coefficient of Determination Between Sensirion SPS30 Vs FEM GRIMM ............................... 28
Figure 3-7: PM$_{2.5}$ 1 Hour Mean, µ/m$^3$ Coefficient of Determination ....................................................... 29
Figure 3-8: PM$_{2.5}$ 1 Hour Mean, µ/m$^3$ ..................................................................................................... 30
Figure 4-1: Air Quality Sensor Installation Location ..................................................................................... 39
Figure 4-2: Cover Off of an Air Quality Sensor Deployed in a Ger ................................................................. 39
Figure 4-3: A Mongolian Ger Where the Air Quality Sensors Were Deployed .............................................. 39
Figure 4-4: Ulaanbaatar Ger District ............................................................................................................. 39
Figure 5-1: Average PM$_{2.5}$ Levels Compared with Average Daily Temperature ......................................... 44
Figure 5-2: Hourly Median PM$_{2.5}$ for Unmodified Gers ............................................................................. 45
Figure 5-3: Hourly Median PM$_{2.5}$ for Modified Gers .................................................................................. 46
Poor air quality is a problem that knows no boundaries. It affects populations all over the world (Carrington & Taylor, 2018; Sengupta, 2018). Epidemiological studies show that ambient air pollutants like PM, O3, SO2, and NO2 are contributors to several respiratory problems including bronchitis, emphysema and asthma (Mabahwi et al., 2014). One area of the world that is hit particularly hard by poor air quality is Mongolia.

Traditionally Mongolian families that live in gers use coal to both cook meals, as well as heat up the gers during the Winter months where temperature can drop as low as -40 °C. The burning of coal inside the ger causes hazardous amounts of particles that are harmful. During the Winter months, 80% of Ulaanbaatar’s air pollution is caused by burning raw coal in stoves in the ger districts. Mongolia has a population of 3 million people, and in 2016 an estimated 1800 people died from diseases due to gers air pollution, and more than 1500 people died from diseases attributable to outdoor air pollution (Alcantara et al., 2020; Bulletin of the World Health Organization, n.d.).

Deseret International Charities and the Gerhub (Gerhub, n.d.) are looking for ways of improving the ger structure to be more energy-efficient, thus needing less coal or no coal at all to be burned with the intent to improve air quality inside the ger. Some ideas include building new structures instead of using traditional gers or modifying current gers to be more energy-efficient (A Ger with Less Brr, 2019). In making the gers more energy-efficient, an electric heater was
used as opposed to using a coal stove. Although these solutions apparently provide an abundant benefit to the people living in gers and the community all together, no indoor air quality data has been collected in Mongolia to measure the effects of these modifications to gers and quantify the benefits (Alcantara et al., 2020).

Many commercial air quality sensors exist (United States Environmental Protection Agency, n.d.); however, deploying air quality sensors in this context requires special considerations. First, the cost of the device needs to be as low as possible. Since these projects are typically humanitarian based, there is not much money to spend on instrumentation. Second, it is vital to be able to monitor the data in real-time. Real-time data allows maintainers to know when the sensors are unplugged or malfunctioning and fix the problem. Without real-time data capability, data collection becomes challenging (you must send someone to every sensor to download the data off of the device), and any sensors problems would not have been known until after all the data is collected. Third, most commercial sensors that are real-time use WiFi to upload the data this is a severe infrastructure limitation because in the ger districts of Mongolia, there is no WiFi connectivity as the residents use cellular data through their smartphones instead; Fourth, solutions that do not require significant human effort to maintain are desirable. A sensor that fulfills these requirements (an inexpensive real-time sensor that uploads data through cellular) is not commercially available (Alcantara et al., 2020).

Although some initial attempts have been made to deploy air quality sensors in developing countries, as discussed in the literature review, there is a need to research on various aspects in regard to deploy air quality sensors in developing countries such as the following:

- Deployment architecture best practices, pros and cons of each architecture.
- Deployment challenges faced while deploying sensors in Mongolian gers.
• Analyze organizations efforts in reducing air pollution inside the gers.
• Check modified gers participants usage patterns while using electrical heaters and how it correlates to PM$_{2.5}$ levels. This will help anyone looking into deploying sensors in remote areas and how to analyze the data.

1.1 Key Terms

A few terms to be aware of in this research paper are the following:

**Ger**: It is a portable round tent covered with skin or felt used as a form of dwelling by many nomadic groups in Central Asia (*Yurt*, n.d.).

**PM$_{2.5}$**: Particulate matter are microscopic particles or liquid matter suspended in the air. The PM$_{2.5}$ refers to the size of the particulate matter, in this case meaning a diameter of 2.5 micrometers or less. PM$_{2.5}$ also known as fine particulate matter, tends to penetrate into the gas exchange regions of the lung (*Particulates*, n.d.).

**WiFi**: It is a family of wireless network protocols based on IEEE 802.11 standards, it is commonly used for local area networking between devices and internet access (*Wi-Fi*, n.d.).

**Cellular**: This is a communication network using wireless as the last link. This network is then distributed over land areas named “cells” (*Cellular Network*, n.d.).

**IoT**: Internet of Things – a network of physical objects embedded with sensors with a goal of connecting and exchanging data with other devices and systems using the internet (*Internet of Things*, n.d.).

**Modified Ger**: Mongolian gers that were retrofitted by Deseret International Charities in conjunction with Brigham Young University Capstone engineering students by replacing the traditional Mongolian stove with an electrical heater, added a radiant barrier to reflect/keep heat inside the gers during the Winter months, and the floor was insulated under the wood slats.
**Unmodified Ger**: Traditional Mongolian ger that uses a coal stove to heat the ger during the Winter months.

**ISP**: Internet service provider.

### 1.2 Research Objectives and Questions

**(Q1)** What are the architecture and design constraints when collecting remote sensor data in Mongolian gers?

**(Q1A)** What are the design tradeoffs of using different architectures (WiFi/Cellular/LoRa) for collecting sensor-based data? What are the pros and cons of each architecture?

**(Q1B)** What challenges are faced when implementing a field deployment sensor-based network in Mongolian gers?

**(Q2)** How effective are the modified gers in reducing PM$_{2.5}$ levels in Mongolian gers?

**(Q2A)** How effective are the modified gers versus the unmodified gers in reducing PM$_{2.5}$?

**(Q2B)** How does temperature affect the PM$_{2.5}$ gains from the modified gers?

**(Q2C)** How does day of week and time of day affect the PM$_{2.5}$ gains from the modified gers?

### 1.3 Outline of Thesis

- Chapter 2: Literature Review
  - An overview of what has been done in air quality sensing deployments focused on air quality health effects and measurements, deploying sensors in developing countries, Mongolia improved heater project and how this research informs this
thesis. In depth discussion of challenges faced while deploying sensors in developing countries.

- Chapter 3: Methodology
  - Methods employed to answer the research questions and objectives and accuracy and precision of used air quality sensor are discussed in detail.

- Chapter 4: Mongolian Field Deployment & Deployment Architecture
  - Architecture and design constraints when collecting remote sensor data in Mongolian gers. This chapter also includes lessons learned from the Mongolian field deployment.

- Chapter 5: Air Quality Intervention Analysis
  - In depth analysis of PM$_{2.5}$ data and how it correlates to the Mongolia improved heater project.

- Chapter 6: Discussion
  - Summary and discussion of findings and limitations. Includes a brief discussion of possible future work.

- Appendix A: Supplementary Materials
2 LITERATURE REVIEW

2.1 Air Quality Health Effects and Measurement

Several studies indicate that air pollutants contribute to increased mortality and hospital admissions. The human effects range from nausea and trouble breathing to skin irritation to cancer (Brunekreef & Holgate, 2002; Kampa & Castanas, 2008). It can also lead to birth defects, serious developmental delays in kids, and reduced activity of the immune system, leading to many diseases (Brunekreef & Holgate, 2002).

Many studies show that all kinds of air pollution at a high pollution concentration affects the airway (Balmes et al., 1987; Kagawa, 1985; Kampa & Castanas, 2008). Similarly, long-term exposure to lower pollutant concentrations also affects the airway. Symptoms like nose and throat irritation, dyspnea and bronchoconstriction, in particular those who have asthma, are normally experienced after exposure to high levels of sulfur dioxide, nitrogen oxides and heavy metals like arsenic, nickel or vanadium. Air pollution in the lungs can cause angina or even myocardial infarction. Memory disturbances, hand tremors, anger, sleep disorders, blurred vision, fatigue, and slurred speech, neurological cancer were all observed after arsenic, lead and mercury exposure. In the digestive system dioxins induce liver cell damage as well as gastrointestinal and liver cancer (Kampa & Castanas, 2008).

The size of pollution particles varies. Typical measures include PM$_{2.5}$ and PM$_{10}$ where the number indicates that the aerodynamic diameter is smaller than 2.5 mm and 10 mm respectively.
The size of the particles determines the location in the respiratory tract that they will deposit. PM$_{10}$ particles deposit mainly in the upper respiratory tract while PM$_{2.5}$ particles are able to reach lung alveoli, which is more damaging to the lungs (Kampa & Castanas, 2008).

In a Winter season in Ulaanbaatar, Mongolia a central government site reported PM$_{2.5}$ of 250 µg/m$^3$ inside gers (Allen et al., 2013). The average PM$_{2.5}$ in gers during the daytime in Winter ranged from 178.4 ± 152.7 to 208 ± 173 µg/m$^3$ (Ban et al., 2017; Lee et al., 2018). Prolonged exposure to such levels can be very detrimental to your health causing an estimate of 40,000 attributable disability-adjusted life years and 1400 attributable deaths annually (Hill et al., 2017). Children in Mongolia with respiratory diseases and neurodevelopmental disorders have been linked to exposure to toxic chemicals from both indoor and outdoor air pollution (Jadambaa et al., 2015). PM$_{2.5}$ can get into the lungs and into the bloodstream causing hypertension, damage to blood vessels, spread to other organs such as the heart and cause damage to its functions and cell structure (Martins & Carrilho da Graça, 2018). In this thesis, I will focus on PM$_{2.5}$ levels for reasons explained in the following sections.

2.2 History and Efforts to Regulate PM$_{2.5}$ Levels

Since the 1950s the United States of America suffered severe air pollution because of increased energy consumption and motor vehicles. In order to improve air quality, the Air Pollution Control Act was established in 1955, it was the first federal legislation on air pollution. The Clean Air Act was funded from it in 1963. CAA authorized the development of comprehensive federal and state regulations to limit emissions from both stationary and mobile sources (Pui et al., 2014).

Many other organizations were brought up with the purpose of setting standards and regulations in order to reduce pollutant particulate matter. PM$_{2.5}$ became the standard
measurement because of the significant adverse health effects outlined in the prior section (Pui et al., 2014).

According to (Pui et al., 2014) the U.S. achieved substantial reduction of PM$_{2.5}$ through the years which may be attributed to the following efforts:

1. Effectively conducted countrywide PM$_{2.5}$ measurements and sampling
2. Accurately measured PM$_{2.5}$ mass and chemical composition concentrations
3. Accurately derived the source allocation of PM$_{2.5}$
4. Effectively implemented both voluntary emission control and mandatory
5. Established partnerships between several organizations such as: academia, industrial groups, and environmental organizations, federal, state, local, and tribal governments (Pui et al., 2014)

As shown in Figure 2-1, these efforts have led to significant decreases in PM$_{2.5}$ over time. The U.S. has been a leader in reducing PM$_{2.5}$ levels. Unfortunately, not all countries have seen reductions. It’s also not clear that the same strategies that worked in the U.S. will also work in other countries with vastly different infrastructures and sources of pollution as described in the following sections.

According to (Martins & Carrilho da Graça, 2018) many Countries adopt the WHO (World Health Organization) guidelines which has different categories such as annual and short term (24hour) limits on human exposure to PM$_{2.5}$. WHO organization advises countries to set and enforce their own national standards. WHO defined an average annual air quality guideline and 3 interims PM$_{2.5}$ exposure levels to help countries to set goals to lower PM$_{2.5}$ and also recommends that annual average should take more importance than 24 hours average as it can be influenced
by a sporadic high PM$_{2.5}$ event. The Tables 2-1 & Table 2-2 describe the different guidelines and a few countries and certain areas in those countries that follow these guidelines.

**Figure 2-1: PM$_{2.5}$ Trend in the USA (https://www.epa.gov/air-trends/particulate-matter-pm25-trends)**

<table>
<thead>
<tr>
<th>WHO AQG/Interim Level</th>
<th>PM$_{2.5}$</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHO AQG</td>
<td>10 µg/m$^3$</td>
<td>Canada, Bolivia, Guatemala</td>
</tr>
<tr>
<td>First Interim Level</td>
<td>35 µg/m$^3$</td>
<td>China, Hong Kong, Malaysia, Rwanda</td>
</tr>
<tr>
<td>Second Interim Level</td>
<td>25 µg/m$^3$</td>
<td>Montenegro, Norway, the Republic of Kosovo, Russia, Colombia, Mongolia, the Philippines, the Republic of Korea, Thailand</td>
</tr>
<tr>
<td>Third Interim Level</td>
<td>15 µg/m$^3$</td>
<td>The Dominican Republic, El Salvador, Honduras, Puerto Rico, Ecuador, Paraguay, Peru Albania, Bangladesh, Japan, Jordan, Saudi Arabia and Taiwan</td>
</tr>
</tbody>
</table>
Table 2-2: WHO Short Term (24 hr) PM$_{2.5}$ Levels and Countries That Follow It

<table>
<thead>
<tr>
<th>Short term (24 hr) PM$_{2.5}$ limit</th>
<th>PM$_{2.5}$</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25 μg/m$^3$ (WHO AQG)</td>
<td>Australia, Bolivia, Guatemala and British Columbia (Canada)</td>
</tr>
<tr>
<td></td>
<td>28-35 μg/m$^3$</td>
<td>Canada, Paraguay, Japan, Russia, Saudi Arabia, Taiwan, the United States, Puerto Rico</td>
</tr>
<tr>
<td></td>
<td>37.5-40 μg/m$^3$</td>
<td>Singapore, Bulgaria, South Africa</td>
</tr>
<tr>
<td></td>
<td>45-60 μg/m$^3$</td>
<td>Mexico, Chile, Colombia, Ecuador, Mongolia, Philippines, The Republic of Korea, Thailand, Brazil, India</td>
</tr>
<tr>
<td></td>
<td>65-80 μg/m$^3$</td>
<td>Bangladesh, The Dominican Republic, El Salvador, Honduras, Jamaica, Jordan, Peru, Trinidad, Tobago and Buenos Aires, Albania, China, Malaysia, Hong Kong, Egypt</td>
</tr>
</tbody>
</table>

Accurately measuring PM$_{2.5}$ is a critical step to improve public health (Mahajan et al., 2018). Also according to (Giannadaki et al., 2016) if an effort world wide was indeed followed by adopting what is currently implemented in the USA which is as of 2012 of 12 μg/m$^3$ for air quality standard for PM$_{2.5}$, millions of premature deaths per year can be prevented. The following sections will discuss strategies for deploying air quality sensors that measure PM$_{2.5}$.

Other countries have developed strategies to reduce PM$_{2.5}$, for instance China, a developing country that has adopted actions to improve PM$_{2.5}$ by setting restrictions to both outdoors and indoors PM$_{2.5}$. Outdoors regulations such as: reducing dust from land and buildings
in urban areas where the government ensures urban greening to reduce uncovered land; vehicles and industrial emissions restrictions. Indoors regulations such as: controlling indoors smoking in public areas by designating smoking and non-smoking zones; using clean fuels like natural gas, methane and electricity as opposed to straws, coal and wood; optimized and well-ventilated stoves should be promoted. Lastly clean the house regularly, avoid burning incense and increase indoor greening helps to lower PM$_{2.5}$ (Li et al., 2017); using portable air cleaners in building as to reduce PM$_{2.5}$ is also being implemented and used (Martins & Carrilho da Graça, 2018).

The focus of this thesis is on air quality sensors and impacts in Mongolia. Mongolia is unique because high levels of PM$_{2.5}$ largely come from residents trying to survive the brutal cold in the Winter season by utilizing coal stoves, the most affordable way to cook and heat up the ger. The coal burning inside the ger exposes its residents to PM$_{2.5}$ at high and prolonged levels that are not healthy (Lim et al., 2018). The focus of this thesis is on measuring reductions in PM$_{2.5}$ due to the BYU capstone modified gers project described later in this chapter. Specifically, I observed how effective a modified ger is compared to a traditional unmodified ger in reducing PM$_{2.5}$ and identified usage patterns in modified gers and how they affect PM$_{2.5}$ levels.

According to (Lim et al., 2018), the U.S. Millennium Challenge Account funded a program in Mongolia designed to develop energy efficient stoves to improve PM$_{2.5}$ air quality inside of gers. The project targeted the ger district in Ulaanbaatar where they recorded the PM$_{2.5}$ levels for a full 24 hours. 60 gers were selected to participate in this study, 31 with improved stove and the other 29 with traditional Mongolian stove and it was projected that they would reduce 50-80% PM$_{2.5}$. However, in this particulate effort they found that the indoor PM in the ger with the improved stove were not significantly lower when compared to the gers with traditional stove. As a matter of fact, the study found that the indoor PM$_{2.5}$ in gers with the more
energy efficient stove were significantly higher than the gers that had the traditional stove. This suggests that there is room for improvement in heating interventions, such as the one evaluated in this thesis.

2.3 Air Quality Sensor Deployment

There have been several initiatives monitoring PM$_{2.5}$ using various types of sensors in previous years. Most of the conventional approaches rely on having air quality monitoring stations deployed strategically and operated by environmental agencies or other environment monitoring organizations. Monitoring stations are costly and big which creates issues during large-scale deployment (Mahajan et al., 2018). The cost of a Continuous Ambient Air Quality Monitoring Station (CAAQMS) can run between $100-130K and around $20-25K for maintenance (Gulia et al., 2020). More recently, deploying large numbers of inexpensive air quality sensors has been possible as described below.

Many commercially available low-cost air quality sensors already exist (United States Environmental Protection Agency, n.d.). Of particular interest are the Purple Air (PurpleAir: Real Time Air Quality Monitoring, n.d.) and AirU (Kelly & Gaillardon, 2017) sensors because they are low-cost (around $260 USD) and reasonably accurate. These sensors have been used in a large number of studies and have been deployed in many places around the world (Hegde et al., 2017). However, these two sensors use WiFi to report data in real-time. Since WiFi is not a viable option in Mongolia and other remote locations, these sensors are not well suited for those contexts. The Mongolia deployment requires a low-cost stationary sensor that must upload data in real-time.
There are many platforms and architectures designed to handle air quality data, for instance EpiFi (Lundrigan et al., 2018; Moore et al., 2018). EpiFi is a system designed for epidemiological research/study purposes. It is designed to deploy sensors in participant’s homes. Because EpiFi relies on WiFi, it is also not feasible for deployment in Mongolia. There is a need for deployments that solve the same problem as EpiFi, but that work in more remote locations that aren’t dependent on WiFi (Alcantara et al., 2020).

Other work has been done on sharing experiences and lessons learned related to deploying less expensive air quality sensors in home environments. Hnat et al. shared their experience of residential sensing deployments in the U.S. (Hnat et al., 2011). They shared deployment issues including running out of wall sockets, wireless connectivity issues, environmental hazards such as children, pets, and robotic vacuums, participants dropping from the study, and maintenance challenges due to combinations of commercial and custom designed devices. Similarly, Barrenetxea et al. shared their experience with wireless sensor network deployments ranging from Swiss Federal Institute of Technology in Lausanne university campus to up high in the Swiss Alps. They recommend not making the system more complex than it needs to be, not immediately trusting the data coming in from the sensors, looking for data outliers and detecting software bugs, and determining when sensors were unplugged (Barrenetxea et al., 2008).

2.4 Deploying Sensors in Mongolia

Several studies have examined air quality sensor deployment in developing countries (Alcantara et al., 2020; Gulia et al., 2020; Miazi et al., 2016). Deploying technology in developing countries, including Mongolia, can be challenging for a number of reasons outlined in this section.
Table 2-3: Consideration of Deploying Sensors Network in Developing Countries

<table>
<thead>
<tr>
<th>Infrastructure</th>
<th>Designers should tailor interventions that fit with the current infrastructure including power, internet connectivity, data centers.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skilled Professionals</td>
<td>Need to be able to hire skilled professionals to implement and support the sensors network.</td>
</tr>
<tr>
<td>Cost</td>
<td>Need to have money to be able to design, architect, implement and support the sensors network.</td>
</tr>
<tr>
<td>Security</td>
<td>Needs to be secured so the money invested does not cause losses.</td>
</tr>
<tr>
<td>Reliability</td>
<td>Sensors need to be reliable, so support is easily maintained for both sensors and its network.</td>
</tr>
</tbody>
</table>

2.4.1 Infrastructure Challenges

In order to successfully deploy sensors, each sensor needs an adequate and reliable Internet connection, scalable infrastructure, continuous power and data backup system. Such an infrastructure does not exist and presents a problem in most developing countries due to prohibitive costs (Miazi et al., 2016).

Electrical power industries are among the costliest in an economy, yet also critical to most IoT sensor deployments. Many homes in developing countries do not have access to electrical power. Data centers are also essential, and they consume so much energy and rely upon efficient and reliable energy sources, such that they are not feasible to build in many developing countries (Miazi et al., 2016).
2.4.2 Skilled Professionals & Cost

Knowledgeable professionals to work with IoT such as technicians, engineers, scientists are scarce in developing countries which can pose a problem to deploy IoT sensors. The acquisition of sensors can be quite costly, also the need to keep up with the latest innovations and hiring scientists from developed worlds is again costly (Miazi et al., 2016).

Once again, just to emphasize one main point, a deployment in Delhi city, India for instance states that one of the many cons about deploying air quality sensors in developing countries is the cost aspect of it. Not only there is a cost to purchase air quality sensors, but there is also a cost to hire skilled workers to deploy and maintain sensors, purchase instruments to calibrate sensors, and pay for electricity and security. As the size of a sensors network increases, there is also a need to analyze the data coming out of it (Gulia et al., 2020).

2.4.3 Security & Reliability

For a successful IoT deployment, the sensors need to be robust, and use batteries that last months and are recharged by solar power. They need to last a long time and be easily replaceable. Whenever a sensor is disconnected or re-connected from the network it is important to know when these events happen, therefore managing each sensor is crucial to correct problems when they arise in a network of sensors. As security of IoT sensors are still in the beginning stages, security is an important aspect that needs to be figured out before it is widely deployed (Miazi et al., 2016).

2.4.4 Mongolia Geography & Deployment Challenges

Ulaanbaatar is the capital of Mongolia; the country is located in between Russia and China. The elevation is 1,528 m, terrain is semidesert, desert plains, grassy steppe, mountains in the
west and southwest; Gobi Desert in south-central. Mongolia has harsh Winter conditions. There is limited natural fresh water in some areas; the burning coal of soft coal in power plants and the lack of enforcement of environmental laws leads to air pollution in Ulaanbaatar (Mongolia - The World Factbook, n.d.).

Mongolia is a uniquely challenging context for deploying technology as many residents live in gers where there are no set power outlets to plug into. Each home is wired depending on the residents’ abilities and preferences, which makes connecting sensors to power plugs more transient and sparse compared to a traditional American home (Alcantara et al., 2020).

2.5 Mongolian Modified Ger Project

Deseret International Charities in conjunction with Brigham Young University Capstone engineering students introduced the idea to replace the traditional Mongolian stove with an electrical heater, add a radiant barrier to reflect/keep heat inside the gers during the Winter months, and the floor was insulated under the wood slats.

After retrofitting 150 gers Deseret International Charities then asked the NET (Network Enhanced Technologies) Lab led by Dr. Lundrigan from Brigham Young University to track the PM$_{2.5}$ in the modified gers as to determine whether the retrofit of gers makes an impact in relation to making it less polluted and therefore improve the quality of life of the residents. The NET Lab team used this practical project to learn more generalizable insights about how to effectively deploy air quality sensors in Mongolia.

2.6 Summary

In summary, although significant research has explored the deployment and accuracy of air quality sensors in public health contexts, little work has taken on these issues in a developing
country context, and none in Mongolia. I hope to add this experience of deploying and analyzing air quality sensors in Mongolia to anyone thinking of deploying and analyzing a sensors-based network in remote locations.
This chapter outlines the mixed method approach used to address the research questions. The first question used a field deployment of air quality sensors in Mongolian gers from September 20, 2019 through April 30, 2020. The second question used quantitative methods to analyze the data captured from the field deployment. Finally, I shared details about the accuracy of the air quality monitors that were used in the study, since the findings are dependent on them. Each of these are discussed below.

3.1 Field Deployment Methodology

Our first research question identifies the architecture and design constraints that exist when collecting remote sensor data in Mongolian gers. This section describes the methodology used to address each of the sub-questions.

(Q1A) What are the design tradeoffs of using different architectures (WiFi/Cellular/LoRa) for collecting sensor-based data? What are the pros and cons of each architecture?

To address our first sub-question, I used a field deployment approach with a team from BYU that visited Mongolia in September 20-24th. I started by conducting a literature review on known challenges of implementing sensor-based networks in developing countries including Mongolia. Before traveling to Mongolia, we created prototypes using both LoRaWAN and
cellular platforms, ultimately deciding to go with the cellular platform as described in (Alcantara et al., 2020).

(Q1B) What challenges are faced when implementing a field deployment sensor-based network in Mongolian gers?

To address this second sub-question, I performed a literature review about known issues when implementing networks of sensors in developing countries and Mongolian gers. The team also identified several challenges through our own field deployment. To identify these challenges, we discussed as a team design tradeoff of using different architectures (WiFi/Cellular/LoRa) for collecting sensor-based data.

During the deployment, the team installed air quality sensors in 35 gers, troubleshooting devices and setup as needed. We met as a team regularly and discussed the needs of specific families and issues that came up during deployment.

When on-site in Mongolia we ran into problems related electrical outlet availability. Some participants were short in power outlets and therefore by adding a plugged-in air quality sensor we were occupying at times an outlet that was important to the participants at one time or another. We ended up in a follow-up trip providing power outlets to make the plugging of the air quality sensor less intrusive.

Another on-site problem was related to sensors having issues sending in air quality sensors data, at times it would send data just fine and then it stopped, after further troubleshooting we have improved the software that writes data points to the microSD card and keeps track of what has been acknowledged. Fortunately, the software update was done over the
air by using Particle Boron microprocessor software update feature delivered via internet through the cell towers.

After returning to the U.S., the team met weekly to review status of the sensors, identify any problems, assign a contact in Mongolia to troubleshoot and fix the sensors, and receive feedback from the contact about what he learned. Lessons learned from a field deployment were discussed and synthesized by the team. Details of our implementation can be found in (Alcantara et al., 2020) and Chapter 4.

### 3.2 Data Analysis Methodology

Our second research question focuses on analyzing data from the sensors to gain insights about the effectiveness of modified gers in reducing PM$_{2.5}$. This section describes the methodology used to address each of the sub-question.

To address this question and the three sub-questions related to it, I captured PM$_{2.5}$ data, performed some data cleanup as described below, aggregated the data, visualized the data, and performed statistical analyses. Because the air quality sensor being used was the Sensirion SPS30 (See pages 33-35 for the reason), although a reliable and accurate air quality sensor, it was not able to remove the humidity out of the PM$_{2.5}$ data collection, which makes the PM$_{2.5}$ data higher than it should be. Therefore, the removal of humidity is imperative in order to get to the true PM$_{2.5}$ value. According to (Di Antonio et al., 2018) in order to remove humidity from particulate matter the following formula can be used, see Formulas 3-1 and 3-2.

\[
C = 1 + \left( \frac{k}{1 + \frac{1.65}{aw}} \right) \tag{3-1}
\]

\[
PM \ (Corrected) = \frac{PM \ (Raw)}{C} \tag{3-2}
\]
Where $k = 0.4$ and water activity is the same as Ammonium Sulphate $(RH = 35\%)$ which is amongst the most common in urban areas (Di Antonio et al., 2018). Tim Dye, an expert in air quality data, suggested that because we were using a low-cost air quality sensor that did not remove humidity when it was collecting PM$_{2.5}$, we should eliminate any data above 80% humidity. The Sensirion SPS30 not only senses PM$_{2.5}$ but also humidity which is important because it is used to remove water particles due to humidity.

Aggregating gers by categories such as modified and unmodified helped in analyzing the data, although many participants that had modified gers were still at times using coal stove during the day and electrical heaters at night because electricity was cheaper. All PM$_{2.5}$ data that had anomaly data like negative or above 1000 PM$_{2.5}$ were modified to 0 and 1000 PM$_{2.5}$ respectively. In the Sensirion SPS30 specifications documentation it mentions that the mass concentration range it can measure PM is between 1 to 1000, thus the reason to only keep PM$_{2.5}$ measurement value up to 1000 (Particulate Matter Sensor SPS30 | Sensirion, n.d.). Timestamps were converted from UTC to Mongolian time, so the correct date and time when a data was collected were represented.

The tools used in this thesis were mainly comprised of Jupyter Notebook for running certain commands once and not having to re-run for a subsequent command as it saves the previous computations in memory, Grafana and also Python along with some libraries including the following:

1. Influxdb – Important to perform a data dump of the database.
2. Pandas – Crucial to read the database dump and store it into a DataFrame, manipulate data as needed and perform data clean up.

4. Numpy – Library used to manipulate arrays.

5. Scipy – Allows to perform mathematics on data.

6. Tqdm – This is quite helpful for the developer as it is a smarter progress meter helpful when downloading large data and display how much data it has downloaded and how much is left to be downloaded.

7. Os – Interacts with the operating system. This was useful to check whether a database dump file already exists and load the data from it. If no file exists that contained the data, then query the database, download data and save it to a file.

These were all helpful to download the data from the influxDB database and after the data was downloaded to plot it, analyze it, calculate statistics, and visualize the data. Microsoft Excel was also used to perform some data analysis. All of these were necessary to make sense of the raw air quality data.

After the data was cleaned up as discussed previously, I was then able to join both modified and unmodified gers in Jupyter Notebook and export the data to a CSV file. Once the CSV file was created, it could be imported into other tools. I created VMware Ubuntu virtual machine and installed InfluxDB so as to run it locally. I then created a script to read each line of the exported CSV file, transform it into a format that can be inserted into InfluxDB and loaded all data to a newly created InfluxDB database.

In order to make sense of the 5,238,205 data points and to remove extreme outliers, I summarized the data into hourly medians. I then used Grafana to help visualize it. See Figures 3-1 and 3-2 below.
Figure 3-1: All Sensors Graph by Hourly Median

Figure 3-2: Grafana Heatmap with Hourly Median for All Sensors
Although this helped make initial sense of the data, I needed to perform additional data statistics and create more helpful visual data aggregations to make more sense of the data. To do so, I exported the hourly median to a different CSV dataset for each ger. I then joined all CSV files into one Excel sheet in preparation for more statistical data analysis and used Microsoft Excel to create a heatmap for both aggregated categories (modified and unmodified gers) and for each ger as well to analyze and see patterns. This allowed us to identify outliers, including 4 gers whose data was consistently so high, I believe there was a problem with the sensor. It also helped us view missing data and create summary graphs that are shown in Chapter 5. I also added average temperature data for Ulaanbaatar for each day in our dataset. This was needed for the statistical analysis used to answer our research questions. Our statistical analysis was performed using SAS with the aid of Dennis Eggett, from the Center for Collaborative Research and Statistical Consulting from Brigham Young University.

(Q2A) How effective are the modified gers versus the unmodified gers in reducing PM$_{2.5}$? And (Q2B) How does temperature affect the PM$_{2.5}$ gains from the modified gers? We addressed our first (Q2A) and second (Q2B) research questions in a single statistical model, since it was important to account for temperature when considering the differences in PM$_{2.5}$ from the modified and unmodified gers. We used a mixed method analysis of covariates, blocking on individual ger and using modified/unmodified ger type as the treatment effect. We also ran the model for four different temperature ranges, since temperature is highly correlated with PM$_{2.5}$ due to the need to run the heaters more in the colder weather (See Figure 5-1). We applied a Bonferrroni correction to account for the fact that were making multiple comparisons. Results are shown in Chapter 5.
(Q2C) How does day of week and time of day affect the PM$_{2.5}$ gains from the modified gers? I created heat maps showing the median PM$_{2.5}$ values for the day of week and time of day on the X and Y axis for each ger, aggregated modified gers, aggregated unmodified gers, and each individual ger. These were visually analyzed to identify common patterns and are shown in Chapter 5 to demonstrate the common patterns of use across days and time of day.

We also performed the same statistical analyses described above, but separating it based on day of the week, hour, or four time blocks (night, morning, midday, and afternoon) for all temperature ranges. Results are shown in Chapter 5.

### 3.3 Air Quality Sensors Deployment and Analysis Replicability

In this section I go over in details about how Sensirion SPS30 accuracy and precision were performed as to give confidence in this research thesis so that anyone would be able to replicate this study using the same air quality sensor with a level of confidence. By reading this thesis along with this research paper (Alcantara et al., 2020) anyone should be able to replicate the air quality sensors deployment, its logic and be able to analyze air quality sensors data.

#### 3.3.1 Accuracy and Precision of Air Quality Sensor Sensirion SPS30

The Sensirion SPS30 (Figure 3-3) is an optical particulate matter sensor. Its measurement is based on laser scattering and contamination-resistance technology. It was selected for this deployment as it fits the low-cost sensor requirement, and it has good accuracy and precision as it is demonstrated in this section (Particulate Matter Sensor SPS30 | Sensirion, n.d.).
3.3.1.1 Laboratory Evaluation of Sensirion SPS30

3.3.1.1.1 Accuracy (PM$_{2.5}$)

To demonstrate the accuracy for the sensor Sensirion SPS30, 3 sensors are matched up against FEM GRIMM which is the gold standard used to compare and contrast with the Sensirion SPS30. PM$_{2.5}$ is collected and aggregated for comparisons. At 20 °C and 40% Relative humidity, see Figure 3-4.

<table>
<thead>
<tr>
<th>Steady state #</th>
<th>Sensor Mean (µg/m$^3$)</th>
<th>FEM GRIMM (µg/m$^3$)</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.7</td>
<td>6.5</td>
<td>96.2</td>
</tr>
<tr>
<td>2</td>
<td>13.3</td>
<td>11.4</td>
<td>83.4</td>
</tr>
<tr>
<td>3</td>
<td>41.1</td>
<td>34.8</td>
<td>81.8</td>
</tr>
<tr>
<td>4</td>
<td>120.8</td>
<td>108.8</td>
<td>89.0</td>
</tr>
<tr>
<td>5</td>
<td>218.8</td>
<td>193.5</td>
<td>86.9</td>
</tr>
<tr>
<td>6</td>
<td>359.4</td>
<td>302.7</td>
<td>81.3</td>
</tr>
</tbody>
</table>
For a detailed discussion of the accuracy measures in this analysis see (Sensirion-Sps30---Summary-Report.Pdf, n.d.). As shown in the Figure 6 above Sensirion SPS30 overestimates measurements from FEM GRIMM on mass concentration at 20 °C and 40% Relative Humidity. Accuracy for the Sensirion SPS30 was fairly constant (81% to 96%) for the PM$_{2.5}$ mass concentration range tested (Sensirion-Sps30---Summary-Report.Pdf, n.d.).

3.3.1.1.2 Precision (PM$_{2.5}$)

A combination of nine evaluations dealing with temperature and relative humidity were performed to measure the precision of PM$_{2.5}$ concentration at either low, medium or high pollutant levels. See Figure 3-5 for further findings where 100% signifies high precision. Details of the definition of precision are provided in (Sensirion-Sps30---Summary-Report.Pdf, n.d.).

![Figure 3-5: PM$_{2.5}$ Precision Between Sensirion SPS30 Vs FEM GRIMM (Sensirion-Sps30---Summary-Report.Pdf, n.d.)](image-url)
3.3.1.1.3 Coefficient of Determination

$R^2$ is the coefficient of determination; it quantifies how the 3 sensors followed the PM$_{2.5}$ Concentration change by FEM GRIMM instrument. When the values of $R^2$ are close to 1 it means near perfect agreement and 0 means no correlation.

From Figure 3-6 it is demonstrated that at 20 °C/40% relative Humidity the Sensirion SPS30 shows strong correlation against FEM PM$_{2.5}$ data with $R^2 > 0.99$ in the lab setting. It also shows that relative humidity and temperature had only a small effect on the Sensirion SPS30 sensor precision. It showed significant variation on colder temperature and high humidity such as 5 °C/65% relative humidity.

Figure 3-6: Coefficient of Determination Between Sensirion SPS30 Vs FEM GRIMM (Sensirion-Sps30---Summary-Report.Pdf, n.d.)
Spikes were present at relative humidity changes in concentration. See Figure 3-6 where it shows the correlation between time and PM$_{2.5}$ mass concentration changes between three Sensirion SPS30 (7CE8, D038 and 5455) against the gold standard FEM GRIMM (Sensirion-Sps30---Summary-Report.Pdf, n.d.).

### 3.3.1.2 Field Evaluation of Sensirion SPS30

Three Sensirion SPS30 were deployed between 03/07/2019 and 05/14/2019 showing moderate to strong correlation for PM$_{2.5}$ compared to the lab experiment. In all 3 sensors, it showed low intra-model variability and data recovery approximately 100%. See details in (Sensirion-Sps30---Summary-Report.Pdf, n.d.).

The Coefficient of determination PM$_{2.5}$ for all sensors (7CE8, D038 and 5455) and FEM GRIMM instrument is: $0.64 < R^2 < 0.85$ indicating moderate to strong correlation which can be seen on Figures 3-7 & 3-8.

![Figure 3-7: PM$_{2.5}$ 1 Hour Mean, μ/m³ Coefficient of Determination (Sensirion-Sps30---Summary-Report.Pdf, n.d.)](image-url)
Figure 3-8: PM$_{2.5}$ 1 Hour Mean, µg/m$^3$ (Sensirion-Sps30---Summary-Report.Pdf, n.d.)
4 MONGOLIAN FIELD DEPLOYMENT & ARCHITECTURE

The purpose of this chapter is to answer our two research questions related to the deployment of PM$_{2.5}$ sensors in Mongolian gers. Specifically, I addressed the high-level question: What are the architecture and design constraints when collecting remote sensor data in Mongolian gers? I did this by addressing two sub-questions, each of which is addressed in a separate subsection:

Research Question 1A (Q1A): What are the design tradeoffs of using different architectures (WiFi/Cellular/LoRa) for collecting sensor-based data? What are the pros and cons of each wireless architecture?

Research Question 1B (Q1B): What challenges are faced when implementing a sensor-based network in Mongolian gers?

4.1 Q1A: Design Tradeoffs of Using Different Wireless Architectures

Choosing to work with wireless technologies while deploying air quality sensors is important in Mongolia because gers are temporary living circumstances where even though electricity is easy to find in the ger district, WiFi being a more permanent and wired solution is not viable. There are three major options for wireless protocol system: WiFi, Cellular, and LoRaWAN. These are each discussed separately below.
1. WiFi Protocol

According to (Wi-Fi, n.d.) WiFi is a wireless network protocol commonly used for local area networking of devices and internet access. WiFi has high absorption, and it works best for line-of-sight usage. Obstructions such as walls, home appliances substantially reduce its range. A WiFi access point has often an indoor range of around 20 feet, while outdoor is around 150 feet. WiFi most common radio bands are 2.4 and 5 gigahertz (Wi-Fi, n.d.).

WiFi access points usually are wired with an internet service provider. Because of its rather short range and the air quality sensors deployment network covered a big area meaning we would need many WiFi access points which would potentially be many single points of failure, WiFi is not the solution for this deployment.

2. Cellular Protocol

Cellular network consists of a communication network in which the last link is wireless, and the distribution is over land areas called cells. To avoid interference a cell uses a different set of frequencies from neighboring cells (Cellular Network, n.d.). In suburban areas the cell towers are spaced between 1-2 miles apart, in dense metropolitan cities they are spaced between 0.25-0.5 miles apart, and in rural/flat terrain they may be spaced between 30-45 miles apart (Kanchwala, 2018).

Different cell signal encodings were developed to distinguish signals from different transmitters such as frequency-division multiple access (FDMA), time-division multiple access (TDMA) and code-division multiple access (CDMA) (Cellular Network, n.d.).
3. LoRa Protocol - Non-Cellular LPWAN Wireless Communication

LoRa is a fairly new wireless protocol designed for long-range while using low power usage. LoRa clients upload data to LoRa gateways that are already connected to the Internet. LoRa is designed to get comparable ranges as cellular, but with no extra cost.

Table 4-1: Developing Countries Air Quality Deployment Architecture Ratings

<table>
<thead>
<tr>
<th></th>
<th>WiFi</th>
<th>Cellular</th>
<th>LoRa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
<td><img src="image1.png" alt="Circle" /></td>
<td><img src="image2.png" alt="Circle" /></td>
<td><img src="image3.png" alt="Circle" /></td>
</tr>
<tr>
<td>Existing Infrastructure</td>
<td><img src="image4.png" alt="Circle" /></td>
<td><img src="image5.png" alt="Circle" /></td>
<td><img src="image6.png" alt="Circle" /></td>
</tr>
<tr>
<td>Cost</td>
<td><img src="image7.png" alt="Circle" /></td>
<td><img src="image8.png" alt="Circle" /></td>
<td><img src="image9.png" alt="Circle" /></td>
</tr>
<tr>
<td>Range</td>
<td><img src="image10.png" alt="Circle" /></td>
<td><img src="image11.png" alt="Circle" /></td>
<td><img src="image12.png" alt="Circle" /></td>
</tr>
<tr>
<td>Hardware/Software availability</td>
<td><img src="image13.png" alt="Circle" /></td>
<td><img src="image14.png" alt="Circle" /></td>
<td><img src="image15.png" alt="Circle" /></td>
</tr>
<tr>
<td>Power Consumption</td>
<td><img src="image16.png" alt="Circle" /></td>
<td><img src="image17.png" alt="Circle" /></td>
<td><img src="image18.png" alt="Circle" /></td>
</tr>
</tbody>
</table>

Table 4-2: Circle’s Legend

<table>
<thead>
<tr>
<th></th>
<th>Best</th>
<th>Better</th>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image19.png" alt="Circle" /></td>
<td><img src="image20.png" alt="Circle" /></td>
<td><img src="image21.png" alt="Circle" /></td>
<td></td>
</tr>
</tbody>
</table>
For our project in Mongolia, and likely other developing country projects, certain dimensions are more important than others. Reliability, Existing Infrastructure are the most important because the remote deployment needed to be reliable due to the distance as we were not close enough to deal with issues. Existing Infrastructure was also just as important because once again due to this deployment being a remote deployment, we were not close enough to debug or maintain the network. Cost for the cellular network was inexpensive enough by bundling measurements and sending to cellular network once every ten minutes as opposed to the cost of building an entire network infrastructure and maintenance related to a LoRa deployment and ISP cost related to WiFi deployment including monthly ISP fees and hardware cost and maintenance.

4.1.1 Mongolia Air Quality Deployment Architecture Ratings Explanation

The definition of reliability used in this context is “the extent to which an experiment, test, or measuring procedure yields the same results on repeated trials” (Definition of RELIABILITY, n.d.). All wireless technologies have one common problem in terms of range which is they are all affected by obstructions.

1. WiFi

   a. Reliability: The conventional method while deploying air quality sensors is to use WiFi architecture, as used by many existing products air quality sensors such as Purple Air and AirU (Kelly & Gaillardon, 2017; PurpleAir: Real Time Air Quality Monitoring, n.d.). This is a highly reliable approach that is well tested. It is extremely reliable and there are several air quality sensors off the shelf using WiFi.
b. Existing Infrastructure: WiFi is not a viable solution because it is not widely available in all parts of the world in particular Mongolia and in the gers.

c. Cost: Although the hardware is inexpensive as it is a popular technology which brings cost down. Due to its short range, many WiFi devices would need to be purchased and maintained driving the cost up. Moreover, there is a monthly ISP monthly fee for each study participant.

d. Range: Because of its small range, about 50 m (LoRa — LoRa Documentation, n.d.), we would need several access points and in turn it would mean several potential single points of failure and also cost would be higher because of cost of several access points and support as well.

e. Hardware/Software Availability: Because of its popularity hardware and software is highly available.

f. Power Consumption: In Ulaanbaatar, Mongolian power was available in gers, however, in some regions, power is not available inside of individual homes. In addition to powering the air quality sensor device and receiving network acknowledgements, to send small data packages through WiFi in a range of 50 m, the power usage is 80 mW (LoRa — LoRa Documentation, n.d.).

2. Cellular

a. Reliability: Cellular networks are highly reliable when devices are within the range of the cell towers, though it can be problematic for those outside of range. Cellular companies are able to support the system if issues arise (Alcantara et al., 2020).
b. Existing Infrastructure: Using an existing cellular network is another option for uploading sensor data. Cellular networks are more widely available than WiFi in developing countries including Mongolia.

c. Cost: Cellular hardware is costlier, and there are service charges (e.g., a monthly cellular data plan) for using the cellular network regularly. The infrastructure and maintenance cost are covered by the communications companies which makes this option less expensive.

d. Range: Widely available in Mongolia to cover the gers district with range of 5 km (LoRa — LoRa Documentation, n.d.).

e. Hardware/Software Availability: Some hardware/software microprocessors are available but need to integrate with air quality sensors as many sensors are not natively equipped to work with cellular.

f. Power Consumption: In addition to powering the air quality sensor device and receiving network acknowledgements, to send small data packages through 3G/4G cellular network in a range of 5 km, the power usage is 5000 mW (LoRa — LoRa Documentation, n.d.).

3. LoRa

a. Reliability: Due to it being a new technology it is not well tested/used as compared to WiFi and cellular. Performance degrades dramatically when there is signal interference and the distance between LoRa devices and LoRa gateways is beyond 2 km (Qadir, 2021).
b. Existing Infrastructure: LoRa is not readily available in Mongolia, therefore a new LoRa infrastructure would have to be deployed.

c. Cost: In designing an air quality sensor to deploy in Mongolia, LoRa was a good option because it could be less cost as compared to using cellular without the monthly data expenses. To cover a study area multiple gateways and maintenance for them would need to be provided driving the cost up.

d. Range: In an urban setting it has long range between 2-5 km, but infrastructure would have to be built including deploying a few gateways to cover the study area and because of the remote deploying aspect, this was not a viable solution due to distance and potential many single points of failures. Because of the distance, this would make for unsustainable deployment (Alcantara et al., 2020).

e. Hardware/Software Availability: Fairly recent technology, so not widely available.

f. Power Consumption: Lowest power consumption of the 3. In addition to powering the air quality sensor device and receiving network acknowledgements, to send small data packages through LoRaWAN in an urban range between 2-5 km, the power usage is 20 mW (LoRa — LoRa Documentation, n.d.).

In order to have a successful air quality sensor deployment in Mongolia, it is necessary to pick an architecture that is reliable, allows for remote support and no need to send someone to each ger to collect data from micro-SD card for example; an infrastructure needs to already exist in Mongolia, so no extra infrastructure/support is needed; inexpensive cost; and hardware to connect to the internet exists off the shelf so no need to spend time building the sensor.
connectivity to the internet. Because of all of these points previously discussed, cellular architecture fits well for a remote deployment in Mongolia.

4.2 Q1B: Sensor-Based Network in Mongolian Gers Challenges

The Mongolia field deployment consisted of four team members from the NET (Network Enhanced Technologies) Lab from Brigham Young University. The team consisted of students from different majors including Electrical and Computer Engineering and Information Technology & Cybersecurity.

The goal was to deploy 35 air quality sensors in Ulaanbaatar, Mongolia in the gers district in the gers that have been modified with alternative heating by adding efficient insulation and using electrical heaters to heat up the gers and also to deploy some air quality sensors in the gers that were not modified.

The air quality sensors deployment took 2 days to deploy starting on 09/20/2019; some adjustments/debugging were made on the next 2 days to make sure we were getting good data from the sensors while we were still there. The deployment consisted of 35 air quality sensors of which it was distributed 10 sensors went to unmodified gers and 25 sensors went to modified gers; This was by designed with the intent to compare and contrast each set of gers (modified and unmodified gers).

The air quality sensors were deployed in each ger in the middle where there are two central pillars. Attaching the air quality sensor on the left pillar, same location in all the gers in order to get similar measurements between all the different gers, see Figures 4-1 & 4-2 below.
The Mongolian ger air quality sensors deployment was located in the ger district which is located in an outskirt of Ulaanbaatar city. A ger district is a form of residential district in Mongolia consisting of parcels with detached houses or gers (Ger District, n.d.). See Figures 4-3 & 4-4 below.
4.2.1 Challenges & Lessons Learned from Deployment

There were many challenges faced in this deployment that we did not anticipate beforehand. In this section I will highlight the main challenges and lessons learned that would be beneficial to anyone deploying air quality sensors in Mongolian gers.

(Q1B) What challenges are faced when implementing a sensor-based network in Mongolian gers?

4.2.1.1 Unplugged Sensors

There were many instances where sensors were sending data and suddenly stopped. As we troubleshooted these incidents it turned out that the study participants were unplugging the sensors often because they needed at one time or another to use the outlet that the sensor was plugged into. The lesson learned here was that dependency on study participants should be kept to a minimum and to accomplish that we could have either used a small battery to have the sensor still be able to send data until the sensor was plugged in again or designed a battery powered enough to not need study participants power plug at all (Alcantara et al., 2020).

4.2.1.2 WiFi vs Cellular vs LoRa

WiFi is the traditional way when deploying air quality sensors as used by AirU and Purple Air, but because WiFi is not readily available in gers it is then not a viable option. Cellular is a good option, but usually the hardware is more expensive and there is a monthly cost associated with its service. LoRa is a relatively new wireless protocol designed for long range and low power consumption, but with no readily available infrastructure and the difficulties of a remote deployment support, multiple single points of failure, and limited access to certain locations to deploy the gateways that will cover the area in which the gers are located makes
LoRa also not viable. Cellular was then chosen because infrastructure was readily available and it is well supported by local communication companies (Alcantara et al., 2020).

4.2.1.3 Management Tools

As a lot of air quality data were coming in from the sensors, it became unmanageable to know exactly if all the data was correct/missing. It is important for the sensors network to have a way to manage them when they are up or down or come back up, or if they are sending data within a certain spectrum, etc. Although there were some tools that were helpful, there were many that yet didn’t exist (Alcantara et al., 2020).

A tool that helped with metadata management as to keep track about sensor location, family information, swapping sensors locations, etc. For this deployment we then created surveys for families, gathered data from them such as location (GPS and address) and family information (Alcantara et al., 2020).

A visualization and data monitoring tool, Grafana helped with graphs and with alerting when sensors went down and back up again. We also needed a way to know if there were missing data or bad data. Data clean-up was needed in order to normalize the data (Alcantara et al., 2020).

It is crucial to be able to send updates over the air in order to fix software bugs and there are tools that help in this regard such as using a Particle Boron microprocessor. For any successful deployment there needs to be data collection and tools to help manage the sensor data; if any tools are lacking, then you need to figure out a way to solve the lack of them, even if it means to perform some manual work due to lack of tools or automation tools to make the deployment easier (Alcantara et al., 2020).
This chapter presents an analysis of the PM$_{2.5}$ data collected from the sensors. This includes statistical analyses and visualizations that help answer the research questions related to (Q2): How effective are the modified gers in reducing PM$_{2.5}$ levels in Mongolian gers? I did this by addressing three sub-questions, each of which is addressed in a separate subsection:

Research Question 2A (Q2A): How effective are the modified gers versus the unmodified gers in reducing PM$_{2.5}$?

Research Question 2B (Q2B): How does temperature affect the PM$_{2.5}$ gains from the modified gers?

Research Question 3B (Q3B): How does day of week and time of day affect the PM$_{2.5}$ gains from the modified gers?

Before analyzing the gers, I removed data from 4 outlier gers (3 modified and 1 unmodified). They were an order of magnitude higher than the other gers, likely due to air quality monitoring calibrations issues. In the end, we had data from 22 modified gers and 9 unmodified gers. The timeframe that we had data from each ger varied, since some sensors in the gers came on and off-line at different times. Our statistical analysis approach helped deal with missing data.
5.1 **Q2A: Analyzing Effective of Modified Gers in Reducing PM$_{2.5}$**

Our analysis comparing all modified and unmodified gers did not show a statistically significant difference at the 5% confidence level, though it was statistically significant at the 10% confidence level. The average PM$_{2.5}$ for unmodified gers was 99 with standard error of 16 compared to 67 with standard error of 10 for modified gers, suggesting a drop of approximately one third (See Table 5-1).

The lack of statistical significance is largely due to the high variation among the gers and the low number of unmodified gers in comparison to the modified gers. If there was more data where we have many missing data points and there were more unmodified gers to make a more equal data comparison with the modified gers then it would have helped to demonstrate its statically significance as shown in Figure 5-1.

| Table 5-1: All Temperature Ranges PM$_{2.5}$ Estimates for Unmodified/Modified Gers |
|----------------------------------------|----------------------------------|--------------------------|-----------------|
|                                       | unmodified | modified               | PM$_{2.5}$ reduction | p-value         |
| All temperature ranges                | 99 (16 S.E.) | 67 (10 S.E.)           | 32                  | 0.0966          |

5.2 **Q2B: Temperature Affect PM$_{2.5}$ Gains on Modified Gers**

A scatterplot based on temperature and average PM$_{2.5}$ across all gers shows that there is a high correlation of -0.70 between the two variables. Specifically, it shows that as the temperature drops, the PM$_{2.5}$ increases due to the need to use the coal stoves or electric heaters more often to heat the gers. See Figure 5-1.
Table 5-2 shows the PM$_{2.5}$ values for modified and unmodified gers for the four different temperature ranges, as well as a p-value that indicates the statistical significance of the comparison of the modified and unmodified gers. Temperature ranges are based on the quartile values of the data points shown on Figure 5-1. For example, 1/4th of the PM$_{2.5}$ daily readings were below -16 °C. Those with a p-value of under .05 are indicated with an asterisk (*). As is shown in Table 5-2, the modified gers did show a statistically significant reduction in PM$_{2.5}$ for the colder temperature ranges. This makes sense, since stoves are needed to run more often during the colder time periods and the electric heaters put off less PM$_{2.5}$. The improved insulation may have also lowered the need to run the heaters as often. Notice the reduction amounts are quite substantial (e.g., a 32-35% reduction for the colder temperatures).
Table 5-2: PM$_{2.5}$ Estimates for Modified/Unmodified Gers by Temperature Range

<table>
<thead>
<tr>
<th>Temperature Range</th>
<th>unmodified</th>
<th>modified</th>
<th>PM$_{2.5}$ reduction</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than -22 °C</td>
<td>130 (17 S.E.)</td>
<td>88 (11 S.E.)</td>
<td>42</td>
<td>*0.0406</td>
</tr>
<tr>
<td>From -16 to -22 °C</td>
<td>119 (17 S.E.)</td>
<td>77 (11 S.E.)</td>
<td>42</td>
<td>*0.0372</td>
</tr>
<tr>
<td>From -4 to -16 °C</td>
<td>89 (17 S.E.)</td>
<td>59 (11 S.E.)</td>
<td>30</td>
<td>0.1365</td>
</tr>
<tr>
<td>Greater than -4 °C</td>
<td>58 (17 S.E.)</td>
<td>42 (11 S.E.)</td>
<td>16</td>
<td>0.4323</td>
</tr>
</tbody>
</table>

5.3 Q2C: Day of Week and Time of Day Affect PM$_{2.5}$ Gains on Modified Gers

5.3.1 Hourly Median PM$_{2.5}$ for Unmodified Gers

Figure 5-2 shows the average of the hourly PM$_{2.5}$ median values for all unmodified gers across each day of week and time of day. Monday It also shows the grand total average as well for the hourly PM$_{2.5}$ median values.

<table>
<thead>
<tr>
<th>Average of Unmodified</th>
<th>Column Labels</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row Labels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>midnight</td>
<td></td>
<td>97</td>
<td>88</td>
<td>82</td>
<td>94</td>
<td>105</td>
<td>103</td>
<td>108</td>
</tr>
<tr>
<td>1am</td>
<td></td>
<td>88</td>
<td>88</td>
<td>74</td>
<td>75</td>
<td>79</td>
<td>85</td>
<td>100</td>
</tr>
<tr>
<td>2am</td>
<td></td>
<td>65</td>
<td>70</td>
<td>51</td>
<td>64</td>
<td>66</td>
<td>77</td>
<td>84</td>
</tr>
<tr>
<td>3am</td>
<td></td>
<td>53</td>
<td>58</td>
<td>39</td>
<td>47</td>
<td>49</td>
<td>53</td>
<td>68</td>
</tr>
<tr>
<td>4am</td>
<td></td>
<td>44</td>
<td>39</td>
<td>35</td>
<td>35</td>
<td>34</td>
<td>42</td>
<td>53</td>
</tr>
<tr>
<td>5am</td>
<td></td>
<td>35</td>
<td>33</td>
<td>28</td>
<td>30</td>
<td>28</td>
<td>31</td>
<td>41</td>
</tr>
<tr>
<td>6am</td>
<td></td>
<td>34</td>
<td>43</td>
<td>27</td>
<td>30</td>
<td>32</td>
<td>31</td>
<td>35</td>
</tr>
<tr>
<td>7am</td>
<td></td>
<td>44</td>
<td>45</td>
<td>28</td>
<td>34</td>
<td>44</td>
<td>32</td>
<td>40</td>
</tr>
<tr>
<td>8am</td>
<td></td>
<td>75</td>
<td>74</td>
<td>56</td>
<td>70</td>
<td>68</td>
<td>50</td>
<td>54</td>
</tr>
<tr>
<td>9am</td>
<td></td>
<td>88</td>
<td>110</td>
<td>122</td>
<td>105</td>
<td>110</td>
<td>101</td>
<td>128</td>
</tr>
<tr>
<td>10am</td>
<td></td>
<td>96</td>
<td>108</td>
<td>114</td>
<td>104</td>
<td>122</td>
<td>112</td>
<td>110</td>
</tr>
<tr>
<td>11am</td>
<td></td>
<td>92</td>
<td>95</td>
<td>92</td>
<td>101</td>
<td>93</td>
<td>83</td>
<td>105</td>
</tr>
<tr>
<td>12pm</td>
<td></td>
<td>70</td>
<td>78</td>
<td>71</td>
<td>73</td>
<td>69</td>
<td>66</td>
<td>78</td>
</tr>
<tr>
<td>1pm</td>
<td></td>
<td>48</td>
<td>64</td>
<td>48</td>
<td>50</td>
<td>57</td>
<td>62</td>
<td>69</td>
</tr>
<tr>
<td>2pm</td>
<td></td>
<td>39</td>
<td>54</td>
<td>44</td>
<td>44</td>
<td>37</td>
<td>57</td>
<td>76</td>
</tr>
<tr>
<td>3pm</td>
<td></td>
<td>45</td>
<td>42</td>
<td>40</td>
<td>53</td>
<td>41</td>
<td>35</td>
<td>51</td>
</tr>
<tr>
<td>4pm</td>
<td></td>
<td>52</td>
<td>50</td>
<td>37</td>
<td>50</td>
<td>48</td>
<td>56</td>
<td>43</td>
</tr>
<tr>
<td>5pm</td>
<td></td>
<td>60</td>
<td>65</td>
<td>68</td>
<td>63</td>
<td>58</td>
<td>56</td>
<td>48</td>
</tr>
<tr>
<td>6pm</td>
<td></td>
<td>60</td>
<td>103</td>
<td>76</td>
<td>74</td>
<td>70</td>
<td>67</td>
<td>60</td>
</tr>
<tr>
<td>7pm</td>
<td></td>
<td>73</td>
<td>100</td>
<td>98</td>
<td>88</td>
<td>82</td>
<td>73</td>
<td>89</td>
</tr>
<tr>
<td>8pm</td>
<td></td>
<td>92</td>
<td>98</td>
<td>101</td>
<td>97</td>
<td>110</td>
<td>98</td>
<td>96</td>
</tr>
<tr>
<td>9pm</td>
<td></td>
<td>119</td>
<td>114</td>
<td>110</td>
<td>101</td>
<td>107</td>
<td>106</td>
<td>102</td>
</tr>
<tr>
<td>10pm</td>
<td></td>
<td>110</td>
<td>118</td>
<td>112</td>
<td>114</td>
<td>128</td>
<td>110</td>
<td>104</td>
</tr>
<tr>
<td>11pm</td>
<td></td>
<td>98</td>
<td>107</td>
<td>104</td>
<td>120</td>
<td>104</td>
<td>105</td>
<td>104</td>
</tr>
</tbody>
</table>

Figure 5-2: Hourly Median PM$_{2.5}$ for Unmodified Gers
The heatmap shows a consistent daily pattern. Starting at 8 am there is a considerable spike in PM$_{2.5}$ most likely correlating with cooking breakfast. Then it stays high until after noon where PM$_{2.5}$ is reduced for a few hours. Another spike starts at 6 pm, likely coinciding with cooking dinner. The PM$_{2.5}$ values stays elevated most of the nighttime, getting lower starting at 2 am, which is likely due to the colder outdoor temperatures, which require more heating inside to stay at a reasonable indoor temperature. These daily patterns are consistent each day of the week, although the average PM$_{2.5}$ values for Sunday are a bit more spread out during the day.

5.3.2 Hourly Median PM$_{2.5}$ for Modified Gers

The daily and weekday pattern of PM$_{2.5}$ for the modified and unmodified gers as shown in Figures 5-2 and 5-3 are quite similar. The absolute numbers are quite different, with the modified gers showing consistently lower values, but same weekly and daily patterns emerge.

<table>
<thead>
<tr>
<th>Average of Modified Row Labels</th>
<th>Column Labels</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>midnight</td>
<td></td>
<td>64</td>
<td>63</td>
<td>62</td>
<td>58</td>
<td>67</td>
<td>72</td>
<td>72</td>
</tr>
<tr>
<td>1am</td>
<td></td>
<td>56</td>
<td>61</td>
<td>54</td>
<td>47</td>
<td>56</td>
<td>59</td>
<td>61</td>
</tr>
<tr>
<td>2am</td>
<td></td>
<td>46</td>
<td>45</td>
<td>42</td>
<td>39</td>
<td>43</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>3am</td>
<td></td>
<td>37</td>
<td>33</td>
<td>33</td>
<td>34</td>
<td>34</td>
<td>41</td>
<td>39</td>
</tr>
<tr>
<td>4am</td>
<td></td>
<td>28</td>
<td>26</td>
<td>29</td>
<td>33</td>
<td>29</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>5am</td>
<td></td>
<td>23</td>
<td>27</td>
<td>26</td>
<td>25</td>
<td>26</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td>6am</td>
<td></td>
<td>23</td>
<td>26</td>
<td>28</td>
<td>25</td>
<td>27</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td>7am</td>
<td></td>
<td>38</td>
<td>35</td>
<td>32</td>
<td>35</td>
<td>38</td>
<td>26</td>
<td>31</td>
</tr>
<tr>
<td>8am</td>
<td></td>
<td>50</td>
<td>49</td>
<td>48</td>
<td>52</td>
<td>54</td>
<td>46</td>
<td>47</td>
</tr>
<tr>
<td>9am</td>
<td></td>
<td>56</td>
<td>72</td>
<td>70</td>
<td>65</td>
<td>71</td>
<td>73</td>
<td>84</td>
</tr>
<tr>
<td>10am</td>
<td></td>
<td>64</td>
<td>79</td>
<td>69</td>
<td>78</td>
<td>79</td>
<td>71</td>
<td>82</td>
</tr>
<tr>
<td>11am</td>
<td></td>
<td>66</td>
<td>72</td>
<td>68</td>
<td>67</td>
<td>67</td>
<td>65</td>
<td>79</td>
</tr>
<tr>
<td>12pm</td>
<td></td>
<td>49</td>
<td>57</td>
<td>57</td>
<td>54</td>
<td>64</td>
<td>48</td>
<td>55</td>
</tr>
<tr>
<td>1pm</td>
<td></td>
<td>41</td>
<td>38</td>
<td>41</td>
<td>46</td>
<td>50</td>
<td>40</td>
<td>48</td>
</tr>
<tr>
<td>2pm</td>
<td></td>
<td>40</td>
<td>33</td>
<td>40</td>
<td>48</td>
<td>36</td>
<td>37</td>
<td>40</td>
</tr>
<tr>
<td>3pm</td>
<td></td>
<td>32</td>
<td>29</td>
<td>33</td>
<td>38</td>
<td>36</td>
<td>32</td>
<td>35</td>
</tr>
<tr>
<td>4pm</td>
<td></td>
<td>39</td>
<td>36</td>
<td>35</td>
<td>38</td>
<td>37</td>
<td>35</td>
<td>42</td>
</tr>
<tr>
<td>5pm</td>
<td></td>
<td>47</td>
<td>51</td>
<td>44</td>
<td>51</td>
<td>50</td>
<td>50</td>
<td>45</td>
</tr>
<tr>
<td>6pm</td>
<td></td>
<td>61</td>
<td>61</td>
<td>66</td>
<td>58</td>
<td>68</td>
<td>60</td>
<td>57</td>
</tr>
<tr>
<td>7pm</td>
<td></td>
<td>64</td>
<td>72</td>
<td>74</td>
<td>64</td>
<td>73</td>
<td>61</td>
<td>69</td>
</tr>
<tr>
<td>8pm</td>
<td></td>
<td>81</td>
<td>81</td>
<td>73</td>
<td>67</td>
<td>74</td>
<td>67</td>
<td>89</td>
</tr>
<tr>
<td>9pm</td>
<td></td>
<td>75</td>
<td>80</td>
<td>65</td>
<td>82</td>
<td>75</td>
<td>78</td>
<td>75</td>
</tr>
<tr>
<td>10pm</td>
<td></td>
<td>73</td>
<td>74</td>
<td>66</td>
<td>78</td>
<td>80</td>
<td>81</td>
<td>81</td>
</tr>
<tr>
<td>11pm</td>
<td></td>
<td>70</td>
<td>70</td>
<td>67</td>
<td>72</td>
<td>78</td>
<td>80</td>
<td>73</td>
</tr>
</tbody>
</table>

Figure 5-3: Hourly Median PM$_{2.5}$ for Modified Gers
Table 5-3 summarizes the different daily averages for each day of the week. The data clearly shows that the modified gers consistently have less PM$_{2.5}$ in comparison to the unmodified gers.

<table>
<thead>
<tr>
<th></th>
<th>Sunday</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unmodified Gers</strong></td>
<td>70</td>
<td>77</td>
<td>69</td>
<td>72</td>
<td>72</td>
<td>70</td>
<td>77</td>
</tr>
<tr>
<td><strong>Modified Gers</strong></td>
<td>51</td>
<td>53</td>
<td>51</td>
<td>52</td>
<td>55</td>
<td>52</td>
<td>56</td>
</tr>
<tr>
<td><strong>Grand Total Difference</strong></td>
<td>19</td>
<td>24</td>
<td>18</td>
<td>20</td>
<td>17</td>
<td>18</td>
<td>21</td>
</tr>
</tbody>
</table>

A more detailed analysis of heatmaps for each ger was performed. The pattern for each ger was very similar to the overall patterns shown above. This indicates that households typically follow a similar schedule of cooking and sleeping. See Appendix A for more details.

Although the overall weekday and daily pattern was similar, we wanted to see if there were any statistically significant differences between various time periods. For this analysis, we used the same mixed model as above, except that instead of separating out daily temperature, we separated out hour of the day and compared modified and unmodified gers.

There were certain hours that had significant results (See Table 5-4). These occurred during the nighttime between 9pm and 1am, and the heavy cooking hour of 10am; all of which are among the highest PM$_{2.5}$ hours per day. Those with a p-value of under 0.05 are indicated with an asterisk (*).
I also grouped the hours into four time buckets (morning, midday, afternoon, and night).

Only the night bucket was significant (See Table 5-5), indicating that the ger modifications are particular useful when it is colder outside. For modified gers at night, there was is a 45% reduction in PM$_{2.5}$, a large amount in comparison with other time buckets, which are: Morning = 30%, Midday = 30%, Afternoon = 36%.

No statistically significant differences for each day of the week were observed.

<table>
<thead>
<tr>
<th>Time</th>
<th>PM$_{2.5}$ reduction</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>midnight</td>
<td>49.3</td>
<td>*0.0101</td>
</tr>
<tr>
<td>1am</td>
<td>43.1</td>
<td>*0.0245</td>
</tr>
<tr>
<td>2am</td>
<td>31.7</td>
<td>0.0976</td>
</tr>
<tr>
<td>3am</td>
<td>19.1</td>
<td>0.3200</td>
</tr>
<tr>
<td>4am</td>
<td>7.8</td>
<td>0.6859</td>
</tr>
<tr>
<td>5am</td>
<td>5.1</td>
<td>0.7925</td>
</tr>
<tr>
<td>6am</td>
<td>8.7</td>
<td>0.6512</td>
</tr>
<tr>
<td>7am</td>
<td>6.8</td>
<td>0.7250</td>
</tr>
<tr>
<td>8am</td>
<td>8.6</td>
<td>0.6548</td>
</tr>
<tr>
<td>9am</td>
<td>25.1</td>
<td>0.1927</td>
</tr>
<tr>
<td>10am</td>
<td>37.7</td>
<td>*0.0498</td>
</tr>
<tr>
<td>11am</td>
<td>25.5</td>
<td>0.1836</td>
</tr>
<tr>
<td>noon</td>
<td>22.2</td>
<td>0.2459</td>
</tr>
<tr>
<td>1pm</td>
<td>22.2</td>
<td>0.2457</td>
</tr>
<tr>
<td>2pm</td>
<td>18.2</td>
<td>0.3397</td>
</tr>
<tr>
<td>3pm</td>
<td>15.0</td>
<td>0.4339</td>
</tr>
<tr>
<td>4pm</td>
<td>16.0</td>
<td>0.4026</td>
</tr>
<tr>
<td>5pm</td>
<td>15.0</td>
<td>0.4329</td>
</tr>
<tr>
<td>6pm</td>
<td>18.4</td>
<td>0.3360</td>
</tr>
<tr>
<td>7pm</td>
<td>23.4</td>
<td>0.2209</td>
</tr>
<tr>
<td>8pm</td>
<td>28.7</td>
<td>0.1337</td>
</tr>
<tr>
<td>9pm</td>
<td>45.0</td>
<td>*0.0190</td>
</tr>
<tr>
<td>10pm</td>
<td>57.9</td>
<td>*0.0026</td>
</tr>
<tr>
<td>11pm</td>
<td>48.6</td>
<td>*0.0114</td>
</tr>
</tbody>
</table>
## Table 5-5: PM$_{2.5}$ Estimates for Unmodified/Modified Gers by Time Buckets

<table>
<thead>
<tr>
<th>Time Bucket</th>
<th>unmodified (S.E.)</th>
<th>modified (S.E.)</th>
<th>PM$_{2.5}$ reduction</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morning (3am-9am)</td>
<td>21 (15 S.E.)</td>
<td>15 (10 S.E.)</td>
<td>6</td>
<td>0.7240</td>
</tr>
<tr>
<td>Midday (9am-1pm)</td>
<td>82 (15 S.E.)</td>
<td>57 (10 S.E.)</td>
<td>25</td>
<td>0.1774</td>
</tr>
<tr>
<td>Afternoon (1pm-7pm)</td>
<td>47 (15 S.E.)</td>
<td>30 (10 S.E.)</td>
<td>17</td>
<td>0.3509</td>
</tr>
<tr>
<td>Night (7pm-3am)</td>
<td>96 (15 S.E.)</td>
<td>53 (10 S.E.)</td>
<td>43</td>
<td>*0.0191</td>
</tr>
</tbody>
</table>
6 DISCUSSION

6.1 Summary

In summary, the first part of this thesis described architecture and design constraints when collecting remote sensor data in developing countries like Mongolia. This will help anyone planning to deploy a sensors network in a developing country to learn ahead of time the many challenges and apply lessons learned from a previously deployment. A detailed comparison of the pros and cons of the key wireless technologies (WiFi, LoRaWAN, and cellular) for remote deployments was provided.

Deployment and data collection of PM$_{2.5}$ was from September 20, 2019 through April 30, 2020. It consisted of 35 air quality sensors; 10 sensors were in unmodified gers and 25 sensors were in modified gers. I discussed the challenges and lessons learned while deploying air quality sensors in remote areas like Mongolia and how to still be able to support the deployment mostly remotely with the usage of over the air software updates.

There were several lessons learned about this deployment such as the following:

1. Keep dependency on study participants to a minimum. Relying on the study participants to always provide electrical power to the sensors caused many data losses.

2. There are many wireless technologies choices to choose from some worth mentioning are WiFi/LoRaWAN, and cellular. Cellular being the most suited for remote locations
due to infrastructure was readily available and it is well supported by local communication companies.

3. For a successful deployment there needs to be data collection and tools to help manage the sensors data, it may be necessary to perform some manual work due to lack of tools or automation tools to make the deployment easier.

Cellular was a good choice for these remote locations, when it’s available. According to (FactsFigures2020.Pdf, n.d.) in developing countries the cellular population coverage using 4G is 82.2% and 3G is 10% with a total combined of 92.2%. Testing in a “real” (uncontrolled) environment leads to challenges that need to be recognized, understanding the users and their needs is important for a successful air quality sensors deployment.

Using low-cost sensors was partially successful because there was a need to throw out data from some of them as there were many missing data points or outlier data. Similar to this air quality sensor deployment according to (Castell et al., 2017) there are still challenges in using low-cost air quality sensors in regards to sensor robustness and measurement repeatability going as far as needing to calibrate each deployed sensor in the field, potentially multiple times, due to sensors no being able to correct for real world conditions over time. The development of high accuracy and low-cost sensors is still needed.

The second part of this thesis reported on the PM$_{2.5}$ levels in Mongolian gers. Although there was not a statistically significant difference between the modified and unmodified gers for the entire duration of the study (Oct – April) at a 5% confidence level, the results were promising, showing an average drop of $1/3$rd. However, statistically significant differences did emerge when looking at colder days and colder times of day (e.g., night-time). As seen in Figure 5-1, the average PM$_{2.5}$ levels increase dramatically when average daily temperature decreases
(correlation = -0.70). The modifications done to the modified gers, the inclusion of more insulation and usage of an electrical heater, helped the most at the coldest times.

The PM$_{2.5}$ level patterns between both modified and unmodified gers are surprisingly similar, even though the PM$_{2.5}$ levels are different, the spikes, increasing and decreasing of PM$_{2.5}$ levels are mostly consistent throughout the days.

The results in this study are important because we were able to set up an air quality sensor network in the modified and unmodified gers that captured the data we wanted overtime. The analyzed collected data showed that during the coldest times the modifications done on the gers worked well.

6.2 Limitations

This study was originally intended to completely separate two distinct categories. One group was unmodified gers, this is the group in which, as the name implies, are gers that were not modified by the retrofit gers project and therefore the data collected was for the sole purpose to compare and contrast. Due to technical difficulties some sensors failed and therefore the numbers of deployed sensors for unmodified gers were significantly reduced.

Unexpectedly, another limitation with this study was that the modified ger, of which as the name implies, were modified gers with more insulation and they were equipped with electrical heaters for the sole purpose to be able to heat up the ger during the Winter season. As the study progressed, we have found that most modified gers participants were actually using the coal stove during the day; at nighttime when electricity was cheaper, they used the electrical heaters to warm up the ger.
Some data went missing whenever the study participants either unplugged the air quality sensors for extensive periods of time or power outages occurred in the area.

6.3 Future Work

More research can be done to better understand and effectively conduct a study, like this one, along with participants of modified gers and understand what obstacles we can collectively work on so that in the modified gers only electrical heaters are used to heat up the gers during the Winter months. The solution will help create better conclusions and be more accurate in terms of PM$_{2.5}$ values between the two different categories (modified and unmodified gers).

Future work also needs to deploy more sensors on unmodified gers to give a better and more statistical significance when comparing with modified gers, thus improving the results of the project.

More work can be done on battery management for air quality sensors so that there is no dependency on study participants in terms of powering the sensors; this would be also helpful so no missing data would occur due to lack of power.

Future research can also work on making the study participants see the importance of modified gers clearer by showing not only improved PM$_{2.5}$ but also showing participants that it may be possible to save money by using a modified ger and electrical heater in a certain way as opposed to the purchase of coal for the stoves. The project needs to include both PM$_{2.5}$ values as well specific cost of electrical heater on a monthly basis.
REFERENCES


https://airu.coe.utah.edu/about/


APPENDIX A: SUPPLEMENTAL MATERIALS

All the Jupyter Notebooks containing earlier and later code including cleaning up data, gers data exports to csv, time series graphs, modified/unmodified gers box plots and modified/unmodified gers for weeks, weekends, day and nights boxplots are included in the compressed folder found at the link at the bottom of this page.

Scripts to load csv data to InfluxDB and load temperature files data to csv are included in the compressed folder found at the link at the bottom of this page.

Earlier UI project prototype showing written in VUE.js showing PM$_{2.5}$ trend graphs with filters and categories can also be found here: https://ylehilds.github.io/analyzing-mongolia/pm_2_5

Spreadsheets files containing all gers hourly PM$_{2.5}$ median, heatmaps for all gers, modified and unmodified gers are included in the compressed folder found at the link at the bottom of this page.

All resource files, code, spreadsheets, graphs, statistical files from The Center for Collaborative Research and Statistical Consulting from Brigham Young University can be found here: https://scholarsarchive.byu.edu/data/22/