TRIBHUWAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

THESIS NO: 078/MSCCD/004

Indoor air quality from cook-stoves in rural Nepal: comparing $PM_{2.5}$ exposures, predicting Carbon Monoxide with machine learning, and characterizing ventilation rates.

by

Binamra Bhusal

A THESIS

SUBMITTED TO THE DEPARTMENT OF APPLIED SCIENCE AND CHEMICAL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS IN CLIMATE CHANGE AND DEVELOPMENT

DEPARTMENT OF APPLIED SCIENCE AND CHEMICAL ENGINEERING

LALITPUR, NEPAL

MAY, 2025

COPYRIGHT

The author has agreed that the library, Department of Applied Sciences, Pulchowk Campus, Institute of Engineering may make this thesis freely available for inspection. Moreover, the author has agreed that permission for extensive copying of this thesis for scholarly purpose may be granted by the professor(s) who supervised the work recorded herein or, in their absence, by the Head of the Department wherein the thesis was done. It is understood that the recognition will be given to the author of this thesis and to the Department of Applied Sciences, Pulchowk Campus, Institute of Engineering in any use of the material of this thesis. Copying or publication or the other use of this thesis for financial gain without approval of the Department of Applied Sciences, Pulchowk Campus, Institute of Engineering and author's written permission is prohibited. Request for permission to copy or to make any other use of the material in this thesis in whole or in part should be addressed to:

____

Dr. Prof. Sahira Joshi
Head
Department of Applied Sciences
Pulchowk Campus, Institute of Engineering
Lalitpur, Kathmandu
Nepal

TRIBHUWAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

DEPARTMENT OF APPLIED SCIENCES

The thesis titled **"INDOOR AIR QUALITY FROM COOK-STOVES IN RURAL NEPAL: COMPARING $PM_{2.5}$ EXPOSURES, PREDICTING CARBON MONOXIDE WITH MACHINE LEARNING, AND CHARACTERIZING VENTILATION RATES"** prepared and submitted by **Binamra Bhusal (078MSCCD004)** in partial fulfillment of the requirements for the degree of Master of Science (M. Sc.) in Climate Change and Development has been examined by us and is accepted for the award of M. Sc. in Climate Change and Development by Tribhuvan University.

The undersigned certify that they have read, and recommended to the Institute of Engineering for acceptance, a thesis report entitled **"Indoor Air Quality From Cook-Stoves In Rural Nepal: Comparing $PM_{2.5}$ Exposures, Predicting Carbon Monoxide With Machine Learning, And Characterizing Ventilation Rates"** submitted by **Binamra Bhusal (078MSCCD004)** in partial fulfillment of the requirements for the degree of Master in Climate Change and Development.

Supervisor: Assoc. Prof. Dr. Nawraj Bhattarai, Dept. of Mechanincal and Aerospace Engineering

External Examiner: Dr. Prof. Khem N Poudyal, Institute of Engineering, Tribhuwan University

Head of Department: Dr. Prof. Sahira Joshi

Department of Applied Sciences and Chemical Engineering

Program Coordinator: Dr. Prof. Rinita Rajbhandari

M.Sc. in Climate Change and Development

Department of Applied Sciences and Chemical Engineering

Date: 2025-05-11

DECLARATION

I hereby declare that this study titled "Indoor Air Quality From Cook-Stoves In Rural Nepal: Comparing $PM_{2.5}$ Exposures, Predicting Carbon Monoxide With Machine Learning, And Characterizing Ventilation Rates" is based on my original research work. Related works on the topic by other researchers have been duly acknowledged. I owe all the liabilities relating to the accuracy and authenticity of the data and any other information included hereunder.

____

Name of student: Binamra Bhusal

Roll number of the student 078MSCCD004

MSc in Climate Change and Development

Date: 2025-05-11

ACKNOWLEDGEMENTS

I am grateful to the Department of Applied Science and Chemical Engineering, Central Campus Pulchowk for providing me the opportunity to work in this thesis study. I would like to extend my sincere gratitude to Dr. Nawraj Bhattarai for his supervision, guidance and support without which this research would not have been completed.

I want to extend my sincere thanks to our MSCCD coordinators, Dr. Rinita Rajbhandari and Dr. Khem Narayan Poudel for guiding me with their valuable inputs, suggestions and supportive push for my thesis work at different phase of research work. I am extremely grateful towards ICIMOD, for providing with the necessary equipments and overall funding during the period of research. My sincere thanks goes to Dr. Binaya KC for providing me the opportunity to work alongside him as a research assistant. I would like to specially mention Er. Amshu Chitrakar for his valuable support and suggestions in research period.

I reserve a special note of thanks to my family and friends, especially all the classmates of MSCCD 2078 for their immense support. Finally I feel indebted to all the people for having direct or indirect support towards me in giving this thesis a tangible form.

ABSTRACT

In the rural household of Nepal, cooking with open fire has an adverse impact on the health of the occupants, especially women and children. The most prevalent stoves using wood as a primary source of fuel are Traditional Cook-Stoves (TCS) and Improved Cook-Stoves (ICS). TCS is an open fire cook-stove where cooking pot is placed on a cast iron cooking stand, whereas, ICS is a two pot mud type stove having a 1.2 m chimney. Indoor Air Quality measurement devices were used to measure pollutant concentration like $PM_{2.5}$ and Carbon Monoxide (CO). Air Exchange Rates (AER), calculated using the CO decay curve method, were characterized for the sampled kitchens and this value is compared with respect to American Society of Heating, Refrigeration and Air Conditioning Engineers (ASHRAE) standards. Furthermore, a data-driven Random Forest machine learning model was developed to predict minute-by-minute CO concentrations.

The real time indoor mean concentration and standard deviation of $PM_{2.5}$ during cooking period was 109.7 $\mu g/m^3$ (σ=29.3 $\mu g/m^3$) for ICS and 314.5 $\mu g/m^3$ (σ=120.4 $\mu g/m^3$) for TCS. Similarly the mean and standard deviation for personal measurement condition was 109.3 $\mu g/m^3$ (σ=26.0 $\mu g/m^3$) for ICS and 248.9 $\mu g/m^3$ (σ=121.1 $\mu g/m^3$) for TCS. The result showed a 65% ($p<0.0001$) and 56.1% ($p<0.001$) reduction in $PM_{2.5}$ for indoor and personal conditions while comparing ICS over TCS. The AER for the sampled households were found to range from 0.2 to 11.6 ACH, with less than 50% meeting the ASHRAE guideline of 5 ACH. The relationship between these calculated AERs and CO decay time was observed to be weak (R^2=0.21), highlighting potential limitations in the simple decay analysis for robust ventilation assessment in this context.

Initial attempts to predict CO concentrations using a standard box model showed significant discrepancies with observed values, likely due to model oversimplifications. Therefore, a data-driven Random Forest machine learning model was developed and evaluated. On a held-out test set of households, the Random Forest model achieved significantly better predictive performance (R^2 = 0.915, RMSE = 16.69 ppm). Although emissions from ICS were observed to be less than TCS, both of them failed to meet the WHO standard permissible value for safe health. Furthermore, factors such as potential smoke back-flow and variations in stove operation, not explicitly included in the models, likely contributed to the observed concentrations and the challenges in simple modeling approaches.

Keywords: ICS, TCS, AER, $PM_{2.5}$, Machine Learning, Random Forest, CO Prediction

List of Abbreviations

ACH: Air Changes per Hour

AEPC: Alternative Energy Promotion Center

AER: Air Exchange Rate

ALRI: Acute Lower Respiratory Infection

ASHRAE: American Society of Heating Refrigeration and Air Conditioning Engineers

CFD: Computational Fluid Dynamics

CO: Carbon monoxide

COPD: Chronic Obstructive Pulmonary Disease

ESAP: Energy Sector Assistance Program

ICS: Improved Cook Stove

IAP: Indoor Air Pollutants

IAQ: Indoor Air Quality

ML: Machine Learning

MAE: Mean Absolute Error

NGO: Non-Governmental Organization

NHRC: Nepal Health Research Council

NLSS: Nepal Living Standard Survey

$PM_{2.5}$: Particulate Matter having aerodynamic diameter of $_{2.5}$ microns

RF: Random Forest

RMSE: Root Mean Square Error

R^2: Coefficient of determination (or R-squared)

TCS: Traditional Cook Stove

VOC: Volatile Organic Compound

VDC: Village Development Committee

WHO: World Health Organization

List of Figures

List of Tables

Table of Contents

CHAPTER 1

INTRODUCTION

1.1 Background

Biomass is the main source of fuel in most rural areas throughout the world and cooking in open flames has been one of the major reasons for poor health of a major portion of the world population (Parajuli et al., 2016). About 3 billion people worldwide rely on biomass (wood, animal dung and crop waste) as their primary source of fuel and the count for premature death due to poor indoor air quality is around 4 million people (Household Air Pollution and Health, 2018). In the context of Nepal around 76% of the population use biomass for cooking and heating purposes (Nepal Energy Situation, 2018). Some of the notable effects of poor indoor air quality due to biomass burning are, Chronic Obstructive Pulmonary Disease (COPD), Acute Lower Respiratory Infections (ALRI), low birth weight and still birth. Women and children are more prone to these effects as developing countries have a much traditional role of women tending over household works like cooking and taking care of children (World Health Organization, 2005).

Natural ventilation is an effective passive technique to reduce some of the effects of indoor air pollution and ventilation models helps understand and analyze the indoor air quality (Zhai, Mankibi et. al, 2015). Natural ventilation is cost effective and leads to better indoor air quality and creates a productive environment among occupants and is an efficient tool for sustainable, urban development (Mozaffarian, 2009).

One effective way to quantify ventilation is calculating the Air Exchange Rate as external air plays a vital role in the ventilation of the building (Chu, 2015). Air Exchange Rate (AER) or Air Change per Hour (ACH) is defined as the air added of removed from a space divided by the volume of the space. Decay curve technique can be used to calculate the carbon monoxide decay time and ultimately the AER (Parajuli, 2016). The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) standards help us compare and understand the condition of ventilation.

Box models are basic and effective tools in analyzing natural processes. Mass balance of carbon monoxide in a box of specified volume which is assumed to be completely mixed can be assessed to predict the concentration of the gas (Ramaswami, 2005). However, inherent assumptions like complete mixing and simplified emission profiles often limit their predictive accuracy in complex, real-world environments like naturally ventilated kitchens.

Furthermore, adding the nuisance of indoor air pollution is the use of Traditional Cooking Stove (TCS), negligence in usage of Improved Cook Stoves (ICS) and poor ventilation. Among others the people who suffer most due to indoor air pollution are women and children because of their cooking roles in rural households. (Parajuli et al., 2016) Women are exposed to biomass smoke for an hour long period for at least three times a day. This makes indoor air pollution a major threat worldwide. More than half of the world's population still relies on biomass combustion to meet basic domestic energy needs (Rehfuess et al., 2006).

Cooking in many developing countries usually consists of burning solid fuels over an open fire or in a poorly functioning traditional stove (Rehfuess et al., 2006). Improved stoves have been designed to burn fuel more efficiently and usually incorporate a chimney or flue. These designs can substantially reduce pollutant emissions and IAP (Albalak et al., 2001; Bruce et al., 2004; Smith, 2002); however, evaluations of improved stoves are limited (Smith, 2002; Albalak et al., 2001).

Similarly, there have been assumptions that the size distribution of traditional cook stove (TCS) and improved cook stove (ICS) are not significantly different and that the relationship between indoor concentration and personal exposure don't vary much between the stoves. We are all aware that firewood is the main energy source for Nepal and that around 78% households are still solely reliant on solid biomass fuels for cooking and space heating purposes (NLSS, 2011/2012). Hence the problems arising from indoor air pollution has been a major concern in Nepali households. High levels of indoor smoke have been reported in kitchens using traditional cook-stoves (TCS) in the rural parts of Nepal (Davidson et al., 1986; Nepal Health Research Council, 2004; Reid et al., 1986). In order to address this problem, the Energy Sector Assistance Program (ESAP) of the Alternative Energy Promotion Centre (AEPC) (www.aepc.gov.np) has been promoting improved cook stoves (ICS) in Nepal since 1999. Over 250,000 stoves have been installed throughout the country to date (Alternative Energy Promotion Centre, 2006). The two pot stove as shown in figure (1.1) is the most widely disseminated model. It has two distinguishing features: improved efficiency in fuel wood consumption and the presence of a chimney to vent smoke outside the kitchen. Monitoring these cook stoves can provide an evidence to support the possibility of the health risks associated with smoke emission during use of TCS and would strongly suggest for the use of ICS in rural populations.

The plancha mejorada cook stove in Guatemala (Albalak, 2001), the pastari stove in Mexico (Armendáriz et al., 2010), the justa stove in Honduras (Clark, 2010), three pot metallic cook stove and 2 pot mud cook stove (Thapa & Shrestha, 2013) in Nepal are some other significant developments made around the world in design and dissemination of improved cook stoves for emission control. These studies have been very useful in the acknowledgement of the pollutant problems being faced day to day by the rural communities of developing countries.

1.2 Rationale

The research focuses on studying the effect of ventilation on indoor air quality. It is established that ventilation plays a vital role in IAQ and buildings can be designed to make the best use of natural ventilation for cost effectiveness, reduced risk of IAP related diseases and thermal comfort. Models and simulations are effective tools to predict and visualize prevalent conditions and help us better understand the dynamics of natural processes. While physics-based box models offer theoretical understanding, their predictive power can be limited by simplifying assumptions, especially concerning variable emission rates and complex airflow. Data-driven modeling, such as machine learning, provides a complementary approach to potentially capture these real-world dynamics more effectively based on observed data patterns. The research can act as a stepping stone for more efficient designs of buildings and houses that focus on indoor air quality in rural settings.

The use of traditional stoves are still prevalent in rural areas of Nepal, while very few households have been utilizing the improved cook stove for cooking and heating purposes nowadays. When burned in simple traditional stoves, the incomplete combustion of firewood leads to an increase in indoor air pollutant concentration usually much higher than the international WHO air quality standards (Armendáriz et al., 2010). Improved stoves on the other hand, may have comparatively lesser pollutant concentration emitted, considering the fact that the stove design is well-engineered than the traditional one. Particle size distribution, along with factors such as age, height, ventilation rate and breathing patterns play a major role in deposition of pollutants in various areas of the lung (Reid et al., 2000), leading to adverse health impacts in developing countries like Nepal.

Hence, through this project, we look forward to quantifying the indoor pollutant concentration, analyzing the difference in particle size distribution in indoor air by the result of traditional cook stove (TCS) and improved cook stove (ICS) and also evaluating the implications of the measured effectiveness of improved cook stove. From this project we hope to achieve a clear difference between the pollutant concentration emitted from both traditional and improved cook stoves and hence provide a strong evidence to support the effectiveness of ICS over TCS. Furthermore, we hope to observe for any relationship between the AER and the CO decay time to evaluate how air flow can affect the emission rate, while also developing a robust model for an accurate prediction of CO emissions from these stove types.

1.3 Objectives

The primary objectives of the research are:

- To quantify the real-time exposure of indoor pollutant concentrations ($PM_{2.5}$, CO values)
- To develop and evaluate a data-driven machine learning (Random Forest) model to predict minute-by-minute indoor CO concentrations, utilizing relevant temporal, environmental, and operational features.
- To measure the AER of houses using biomass as the primary source for fuel and comparing it with ASHRAE standards to better understand the condition on Indoor Air Quality in rural Nepal
- To calculate the difference in particle size distribution in indoor air as a result of TCS and ICS
- To analyze the measured effectiveness of ICS over TCS.

1.4 Scope and Limitations

Some of the limitations of the study are:

- The relationship between calculated Air Exchange Rates (AER) and CO decay time was characterized to provide context on ventilation conditions, although this analysis highlighted the limitations of simple decay methods for robust AER quantification in this setting.

- Some observations like ‘backflow’ which drastically affects the IAP concentration are not considered for our research.
- The chimney orientation and height may tend to hinder in the backflow of smoke. Hence, study about the best location and orientation of chimney should be done in order to maintain IAQ for the safe health of dwellers.
- In a more practical approach, efficient stoves may often be unclean and the clean stoves may often be inefficient as they consume more fuel wood than the traditional stove without chimney. Hence, the pollutant concentration may not be reduced as per the permissible values given by WHO. Thus, there may not be a proper healthy environment for the individuals
- Initial explorations using simple box models confirmed their limitations for accurate CO prediction due to simplifying assumptions (e.g., perfect mixing, constant emissions), motivating the adoption of a data-driven modeling approach.
- The developed Random Forest model's performance is contingent upon the specific characteristics and size of the training dataset; its generalization to different regions or conditions requires further validation.
- The selection of input features (e.g., lag times, time variables) for the machine learning model influences its predictive capability.
- While machine learning models can achieve high predictive accuracy, interpreting the specific physical reasons behind individual predictions can be less direct than with mechanistic models, although feature importance analysis provides valuable insights.

Figure 1.1: Improved Cook Stove

Figure 1.2: Traditional Cook Stove

CHAPTER 2

Literature Review

1. Review of Natural Ventilation

Zhai et al. (2015) discuss the importance of natural ventilation as an efficient passive technique to reduce building cooling energy and improve the indoor air quality of the building. They discuss about the varying complexities of various models and their prediction accuracy. They review different models and simulations tools and compare their prediction capabilities. They also discuss about the geometries of openings, pressure co-efficient, discharge coefficient and other various parameters for accurate prediction using ventilation models. Two of the prominent model types of this research are, Analytical and Empirical models made by applying mass and energy conservation equation in particular conditions and Network Models which are models constructed to represent the airflow in the building using a series of matrix equations (Zhai, Mankibi et al., 2015)

2. Wind-Driven Natural Ventilation for Buildings with Two Openings on the Same External Wall

Chu et al. (2015) conducted experiments in a wind tunnel using tracer gases under different wind speed, direction and opening size. A semi-empirical prediction model is also used to calculate the ACH of the setup at various simulated conditions. Furthermore, they conclude that ACH is linearly proportional to wind-speed and opening area, this may help us in developing and designing control measures for improved air quality (Chu, 2015)

3. **Indoor-to-outdoor relationship of aerosol particles inside a naturally ventilated apartment- A comparison between single parameter analysis and indoor aerosol model simulation**

   Hussein et al. (2017) uses simulation tools to study about the indoor-to-outdoor relationship of aerosol particles. Two different techniques were used which were, single parameter analysis and indoor aerosol simulation model. The paper also discusses the underlying factors like ventilation, penetration, dry deposition rates and emission sources. The aerosol simulation approach is capable of describing the effect of these factors based on high time resolution calculations. It also discusses about the I/O value in different times of the year and particle concentration during various activities like vacuuming, cooking, children playing, spraying insecticides and pesticides and tobacco smoking.

4. **Indoor air pollution in Developing Countries**

   Chen et al. (1997) shed light on the condition of indoor air pollution in developing countries. The research spans over many countries in South East Asia and study on time of exposure of the occupant and adverse effects of exposure to health (Chen et al., 1997).

5. **Modeling Indoor Air Pollution from Cook Stove Emissions in Developing Countries using a Monte Carlo Single-Box Model**

   Johnson et al. (2011) presents a first approach to indoor air pollution modeling using Monte-Carlo single box model. The model relates the emission of pollutants from fuel/cook-stove combinations and the resulting indoor air pollution concentrations. The model combines stove emission rates with expected distributions of kitchen volumes with air exchange rates to produce a distribution of IAP concentration estimates. Then the resulting distribution can be used to predict the likelihood that IAP will meet air quality guidelines set by World Health Organization for fine particulate matter ($PM_{2.5}$) and carbon monoxide. It concludes that only 4% of homes using fuel-wood in a rocket-style cook-stove under idealized condition meet the IAP concentration of 35µg m^{-3} set by WHO. The model was then said to have over-estimated the IAP values due to the fact that the stove emissions in the room was assumed to be completely mixed.

6. **Natural Ventilation in Buildings and the Tools for Analysis**

Mozaffarian (2009) discusses natural ventilation in buildings and the tools required for analysis. Natural ventilation practice has been disregarded in recent times due to developing electrical cooling and heating devices. The research different techniques and designs for natural ventilation in buildings and discusses its economic benefits due reduced energy and operating costs, better indoor air quality and healthier environment. It focuses most on natural ventilation as a result of wind. It also analyses the usage of wind tunnel (badgir) for natural ventilation. It proceeds to discuss about the different tools used to design and analyze natural ventilation like LoopDA, AIOLOS Software, Autodesk Ecotect, Green Building Studio and Computational Fluid Dynamics-CFD using FLUENT package (Mozaffarian, 2009).

7. **Indoor Carbon Monoxide and $PM_{2.5}$ Concentrations by Cooking Fuels in Pakistan**

Siddiqui et al. (2008) conducted measurements of Carbon Monoxide and particulate matter ($PM_{2.5}$) in kitchens using wood or Natural Gas in semi-rural conditions in Pakistan. In the daytime eight hours of continuous measurement for CO and $PM_{2.5}$ was taken, among which 51 houses used wood and 44 houses used natural gas. $PM_{2.5}$ was measured using laser photometer which was calibrated for field conditions and reduced be a factor of 2.77 and for CO, electrochemical monitor was used. The arithmetic means for CO and $PM_{2.5}$ were found to be 29.4 ppm, 2.74 mg m^3 and $_{2.5}$ ppm, 0.38 mg m^3 for wood and natural gas respectively. Time spent in the kitchen using wood was found to be significantly higher than in the kitchens using natural gas, combined with higher peaks in CO and $PM_{2.5}$ which lead them to conclude that wood fuel may lead to hazardous exposure to IAPs (Siddiqui et al., 2008).

8. Indoor Air Quality and Ventilation Assessment of Rural Mountainous Households of Nepal

Parajuli et al. (2016) conducted experiments at Khasauli and Bhairabstan VDCs of Palpa district to find the mean CO and $PM_{2.5}$ concentrations due to biomass burning in Traditional Cook-Stoves (TCS) and Improved Cook-Stoves (ICS) and Air Exchange Rates were calculated using Decay Curve technique used to calculate the Carbon Monoxide decay time. The AER was calculated using a box model of specified volume which was assumed to be completely mixed.

The mean CO and $PM_{2.5}$ concentration for ICS and TCS were observed to be 27.11 ppm and 825 $\mu g/m^3$ ($p<0.0001$) and 36.03 ppm and 1336 $\mu g/m^3$ ($p<0.0001$) respectively. The ventilation analysis showed that more than 80% household lacked the minimum AER requirements set by ASHRAE of 5 hr^{-1} (Parajuli, 2016).

9. Effect of enhanced stove on exposure to $PM_{2.5}$ and CO

The use of improved stove instead of traditional one has nowadays been encouraged from various sources like Nepal Government, NGOs, academic institutions, etc. and also has been implemented by the people living in the rural settings of Nepal. One notable effort was by KC et al. (2015) from Kathmandu University, in close collaboration with Dhulikhel hospital in designing and disseminating an improved cook stove (2 pot mud cook stove with chimney) in almost every household in Salambu, a rural village present in Kavre district. Replacement of traditional stove with an efficient improved one has been found to drastically decrease the indoor pollutant concentration levels and has been an effective method for pollutant reduction.

The study conducted on Salambu found the real time exposures of $PM_{2.5}$ and CO to be reduced by 65% and 50% respectively in the households when ICS was compared over TCS (KC et al., 2015). The concentration of $PM_{2.5}$ and CO around kitchen premises was recorded at an interval of 10 seconds for actual cooking periods and one-minute intervals for 24 hours period under real time conditions.

The real-time mean (μ) and standard deviations (σ) of $PM_{2.5}$ concentration for TCS (13 households) while cooking were $\mu=943.8\ \mu g/m^3$ and $\sigma=426.5\ \mu g/m^3$ and concentration of CO were measured to be $\mu= 13.5$ ppm and $\sigma= 5.2$ ppm. Similarly the concentration of

$PM_{2.5}$ and CO for ICS (13 households) while cooking were μ=334.6 μg/m3 and σ=228.6 μg/m3 and μ= 6.5 ppm and σ= 4.8 ppm respectively. This shows a significant reduction in real-time PM and CO concentration when comparing ICS over TCS. However the average $PM_{2.5}$ concentrations still exceeded the WHO indoor air quality thresholds for $PM_{2.5}$ and national air quality guidelines.

Therefore, a large proportion of local people are still prone to developing diseases related to increased levels of air pollutants regardless of their ownership of ICS, although the use of ICS resulted in an overall reduction of pollutant concentration and hence the users of ICS had a comparatively lower exposure to concentrations than the users of TCS.

The real time personal exposures to $PM_{2.5}$ and CO and 24 hour average $PM_{2.5}$ and CO in the household were monitored using the IAP meter 5000 series (Aprovecho Research center, USA). Real-time pollutant concentrations were measured at an interval of 10 seconds over each cooking period from the moment the fire started till the fire extinguished. The wearer had the meter on her back with the adjustable tube attached to the meter box throughout the cooking period from the start. Data were analyzed using GraphPad Prism software (Version 6, GraphPad Software Inc.). The mean of the arithmetic means of indoor emissions of $PM_{2.5}$ and CO from ICS and TCS were obtained and compared for any significant difference using the t-test.

10. Indoor Air Quality and ventilation assessment of mountainous regions of Nepal

Parajuli et al. (2016) conducted a study in the mid hills of western Nepal which monitored the general IAQ of houses having Improved Cook Stove (ICS) and Traditional Cook Stove (ICS) as their source of cooking. The study found mean and standard deviation (σ) of CO and $PM_{2.5}$ concentrations for ICS and TCS to be 27.11 ppm and 825.4 $\mu g/m^3$ (σ=14.24 ppm and 730$\mu g/m^3$) with significant correlation ($p < 0.0001$) and 36.03 ppm and 1336 $\mu g/m^3$ (σ=19.06 ppm and 952.8 $\mu g/m^3$) with significant correlation ($p < 0.0481$), respectively. The concentrations were measured in 16 households having TCS (n=6) and ICS (n=10) as their source of cooking. The result concluded that the households having ICS had comparatively lower pollutant concentrations than in the households with TCS. However, both types of stove exceeded the threshold value recommended by various organizations for safe health.

The IAQ was assessed by continuous monitoring for 24 hours in a sampled household using a sampling protocol that was developed by Center for Entrepreneurship in International Health and Development (CEIHD), University of California, Berkeley. CO was measured using CO T82 data logger that records the CO data with the interval of every one-minute throughout measurement period. The $PM_{2.5}$ was measured using PM buck pump set (Model LP-5) with buck calibrator having pumping rate 2200 ml/min. This study recommends for the replacement of TCS with ICS focusing greatly on chimney height to reduce the indoor pollution.

11. Effectiveness of improved cook stove for reducing Indoor air pollution

This research monitoring conducted by Singh et al. (2012) assessed the impact of mud ICS, disseminated by the Nepali national cook stove program, in reducing indoor air pollution (IAP). A longitudinal, "before and after" research design previously conducted by Edwards et al. (2007) was employed. Mean 24 h $PM_{2.5}$ (particulate matter less than 2.5 micrometers in aerodynamic diameter) and CO (carbon monoxide) concentrations were measured in the kitchen. Household pollution monitoring was conducted in two phases - 3 months and 12 months post installation of the ICS.

After 1 year of ICS use, the mean values of $PM_{2.5}$ and CO were reduced 63.2% and 60.0% respectively. $PM_{2.5}$ concentration was significantly lowered from 2.07 mg/m^3 (95 % CI: 1.42–2.71) during traditional cook stove (TCS) use to 0.76 mg/m^3 (95 % CI 0.521–1.00) during ICS use. The mean CO concentration was reduced significantly from 21.5 ppm (95 % CI: 14.5–28.6) to 8.62 ppm (95% CI: 6.18–11.1). The mean concentrations of $PM_{2.5}$ and CO measured during 3-month and 12-month phases were not statistically significant in houses operating ICS. The study established the use of mud ICS as an appropriate intervention to reduce $PM_{2.5}$ and CO concentrations in rural kitchens. This study recommends greater focus on proper operation and maintenance of ICS and ventilation to lower the smoke level even further.

12. Indoor particle size distributions in homes with open fires and improved Patsari cook stoves:

The paper conducted by Armendáriz-Arnez et al. (2010) presents the difference in particle size distributions in indoor air as a result of biomass burning in traditional (open fire) and improved Pastari stoves conducted in rural homes in Mexico. They also evaluated the measured effectiveness of improved stoves over the traditional one. 10 homes that used improved *Patsari* stoves and 11 homes that used open fire stoves were randomly selected and personal exposures and indoor air concentrations of particulate matter were assessed over a 24 h period during the dry season. PM size fractions of >2.5, 2.5-1.0, 1.0-0.5, and 0.5-0.25 µm were deposited on 25 mm, 0.5 µm pore size Teflon filters (SKC Inc., USA) and particles <0.25 µm on 37 mm, 2.0 µm pore size Teflon filters (SKC Inc., USA).

The largest differences in the size distribution in indoor air between homes with open fires and *Patsari* improved stoves was in the smallest size fraction (< $_{2.5}$ µm) where the mass concentration in homes with improved *Patsari* homes was 72% smaller than that in homes with open fires. Significant differences were seen between indoor air concentrations for the open fire and the improved *Patsari* stove for the smallest size fraction <0.25 µm, which contributed 67% of indoor air $PM_{2.5}$ concentrations for the open fire and 48% for the *Patsari*, showing a reduced relative contribution of this size fraction of 19% for homes with *Patsari* stoves relative to open fires. On average, the *Patsari* improved stove showed 72% ($p < 0.05$) reduction in concentrations of particles less than 0.25 mm.

13. Indoor $PM_{2.5}$ and CO concentrations estimation in Southern Nepal

Chen et al. (2016) analyzed a study to characterize the indoor concentrations of particulate matter ($PM_{2.5}$) and carbon monoxide (CO), and to understand their impact on health in rural Southern Nepal. This study analyzed daily monitoring data collected with DataRAM pDR-1000 and LASCAR CO data logger in 2980 households. Daily average $PM_{2.5}$ and CO concentrations near stove were collected to be 1,376 (95% CI, 1,331–1,423) µg/m3 and 10.9 (10.5–11.3) parts per million (ppm) among households with traditional cook stoves. Average stove-influenced concentrations were 3,469 (3,350–3,588) $\mu g/m^3$ for $PM_{2.5}$ and 21.8 (21.1–22.6) ppm for CO. Dry season was also a factor in increasing the $PM_{2.5}$ concentration. Adding dung into the fuel also affected the concentrations of both pollutants by increasing it. This study helped reveal some knowledge on household air pollution (HAP) in rural Nepal using traditional cook stoves and accounted for very high concentrations in these households.

14. Indoor Respirable Particulate Matter from Open fire, Improved cook stove and LPG/ open fire combination

This article, written by Albalak et al. (2001) presents a comparative study of 24 h $PM_{3.5}$ concentrations for three cook stove/ fuel conditions: a traditional open fire cook stove, an improved cook stove called the *plancha mejorada* and a LPG stove/ open fire combination.

Twenty-four hour geometric mean PM3.5 concentrations were 1560 $\mu g/m^3$, 280 $\mu g/m^3$, and 850 $\mu g/m^3$ for the open fire, *plancha*, and LPG/open fire combination, respectively. The test showed a 45% reduction in $PM_{3.5}$ while comparing LPG/ open fire combination with the open fire alone ($p< 0.0737$). The *plancha* showed an 85% reduction in $PM_{3.5}$ concentrations as compared to the open fire ($p < 0.0001$). While studying the interaction of time with stove types, they found no significant difference in the temporal trend in pollution among the three stove types. The reduced $PM_{3.5}$ concentrations were maintained over time. The *Plancha* stove offered the best results for achieving substantial reductions in indoor air pollutants. Issues of cost and stove maintenance remain to be addressed.

CHAPTER 3

METHODOLOGY

3.1 Instruments

3.1.1 AEROCET 831

The AEROCET 831 as shown in figure 3.1 is a mass profiler capable of measuring real-time concentrations of particulate matters of sizes 1, 2.5, 4, 10 and TSP. It first counts and sizes the particles and assigns them a density. The density is assigned by K-factor setting which must be empirically derived by comparing it with a reference unit. A 780nm laser diode light source is used and a single particle in allowed into the detection chamber and the intensity of the light scattered by the particle is analyzed to determine the size of the particle. Using their size data and the assigned density, the AEROCET831 is capable of converting these values to concentration unit (μg/m^3). It is able to measure and record pollutant concentration in indoor environments every minute. The flow rate of the device is maintained at 2.83 liters/minute. It measures particles of sizes ranging from 0 to 1000 μg/m^3 and has a resolution of 0.1μg/m^3.

Two AEROCET831 were used per sampling kitchen one for personal air sampling and one for ambient air sampling. The AEROCET831 used for personal air sampling was mounted on a vest on the person subjected to the kitchen smoke, as close to the breathing area as possible. The AEROCET831 used for indoor air sampling was placed near the stove. The measurements given by the AEROCET831 were downloaded into Excel after each day of sampling and stored into Microsoft Excel.

3.1.2 IAQ probe

The IAQ probe as shown in figure 3.2 was used to measure real-time concentrations of carbon monoxide (CO). The CO was detected and measured using an electrochemical sensor present on the probe with and accuracy of ±2ppm<50ppm. The IAQ probe is directly connected to WolfSense desktop software for data review, graph generation etc. Due to this, IAQ probe does not require any post measurement data handling. During measurement, all the information required for data analysis can be entered and saved to the Windows device. Hence the IAQ probe is a powerful tool for Indoor Air Quality measurement. Besides CO it is able to measure, CO_2

concentration, VOC concentration, temperature and relative humidity within the interval of one minute.

The IAQ probe was placed around a distance of one foot from the stove. The code for each house was entered into the software and the measurement was started. The information could be viewed as soon as the measurement ended to observe the trend lines for any major event during the cooking period.

3.1.3 Grimm

The Grimm EDM 164 is and environmental dust monitor that combines the optical detection as a means for counting and categorizing dust particles. This device is suitable for monitoring dust particles continuously and aids for the real time examination of dust measurement data i.e., PM_{1}, $PM_{2.5}$ and PM_{10} values. The device is capable of measuring particles ranging from sizes 0.25 to 32µm and concentration ranging from 0.1µg/m^3 to 100 mg/m^3. The sample flow rate is 1.2 liters/minute. Particles present in the sample air are classified by size and number using scattering light measurements. Each particle detected in the optical measurement cell is classified into defined particle sizes based on the intensity with which the particle scatters the light signal. When the particle diameter and density are determined, based on the assumption that the particles are spherical, the particle mass is derived. The EDM 164 uses the light scattering at 655 nm to determine the mass concentration. For precision in measurement and protection against contamination of measuring cell, portion of the sample flow volume of 1.2 liters/minute is filtered and returned to measurement device as purge air.

Grimm was used as a standard for particulate measurement and all the devices measuring PM were corrected by collocating them with Grimm.

Figure 3.1: IAQ probe with interface

Figure 3.2: AeroCet 831

3.2 Site Description

3.2.1 Salambu

Salambu of Majhi Feda VDC (Ward no. 3) is in Kavrepalanchok district. Majhi Feda VDC has 627 households with a total population of 2669 out of which 617 households use firewood as main fuel source (Central Bureau of Statistics, 2011). It is at an altitude of 1694 meters and has the coordinates 27° 35' 28.55” N, 85° 49' 59.76” E. the male and female population is 1225 and 1444 respectively. The average household size is 4.26 and the sex ratio is 84.83. The major occupation of this area is agriculture and cattle rearing. Dhulikhel Hospital Outreach Center is located in this area. Local people have access to firewood from the nearby forest. *Pinus roxburghii* (Salla) and cornstalk along with agricultural residues are the main sources of fuel in this area. Dhulikhel Hospital collaborated with Kathmandu University to design and distribute improved cook stove in almost every household of this village.

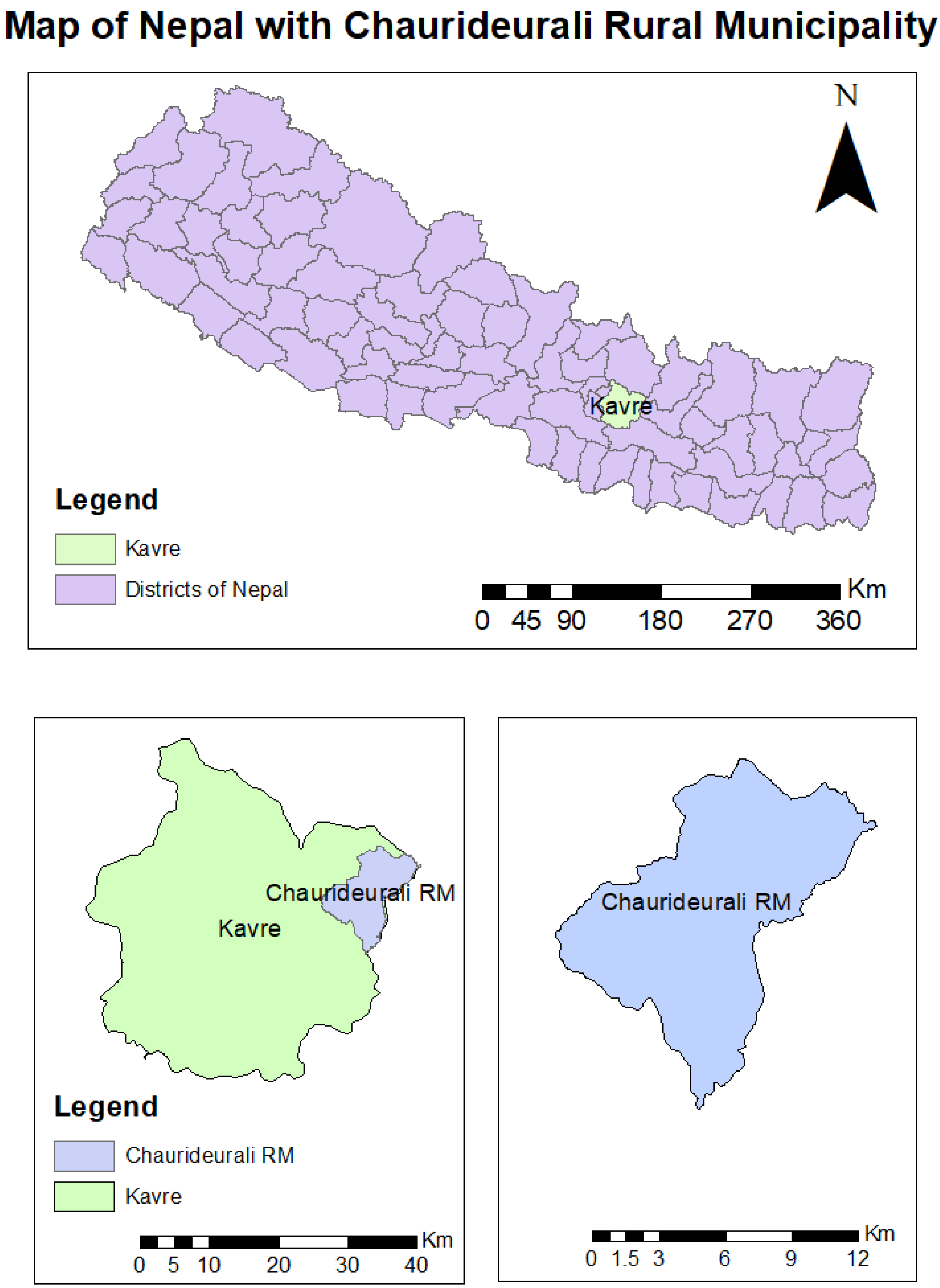


Figure 3.3: Site for Salambu VDC

Figure 3.4: a) TCS (open fire); b) ICS (two-pot mud type)

3.3 Stoves studied

In the course of this project we compared the real time indoor and personal emissions to $PM_{2.5}$ during morning and evening cooking periods between two stove designs, the Traditional Cook Stove (TCS) and the Improved Cook Stove (ICS) as shown in figure 3.5: a) and b).

TCS is a simple open fire cook stove where the cooking pot is placed in a cast iron cooking stand. Firewood is ignited under the stand and can be placed from all directions. Although this might increase the cooking speed, it can have a disadvantage of excess fuel consumption.

ICS on the other hand, is a two pot mud type stove having better fuel efficiency and attached with a chimney. The houses using TCS did not have any particular ventilation for the extraction of pollutants from the enclosure, while ICS had proper installment of chimney to remove most of the smoke coming from combustion out of the kitchen premises. This gave the houses having ICS a better advantage over TCS in terms of ventilation and exposure reduction. However, we did face problems regarding damaged chimneys having low maintenance issues which reduced its ventilation efficiency than the actual capability.

3.4 Data Analysis

The mean and standard deviation of real time data of $PM_{2.5}$ emission of both the stoves were calculated and compared for any significant differences using independent two-sample t-test. The data was first divided into three groups: before cooking, during cooking and after cooking. The test was applied to the data falling in the "during cooking" period as there was maximum $PM_{2.5}$ emission from both stoves in this interval. The analysis was done for both indoor as well as personal conditions. For indoor condition, the measurement devices were kept inside the kitchen premises close to the cook stove while for personal condition, the participant had to wear a jacket which encased the measurement devices, during the entire cooking period. The nozzle of the instrument was adjusted close to the nasal region of the participant such that the $PM_{2.5}$ emission that could enter into the participant's body can be measured.

The independent two sample t-test was chosen as a comparative basis due to small and unequal sample size of the cook stoves, considering the assumption that the data obtained was normally distributed. We had 15 ICS samples and 17 TCS samples available for the calculation of our test ($n > 15$). A graphical representation of $PM_{2.5}$ emission before, during and after cooking was also depicted. Furthermore, the normal distributions and boxplots of ICS and TCS obtained through the two data were also represented and compared graphically. A simple linear trend line of $PM_{2.5}$ was also plotted with respect to the time of cooking. All the statistical calculations were conducted through the use of Python programming language and MS Excel.

3.4.1 Collocation of Data

Collocation is the act of arranging two instruments such that the data obtained from these instruments are in unison. The AeroCet 831 used for indoor and personal measurement of $PM_{2.5}$ was collocated with Grimm EDM 164 as the standard reference. The $PM_{2.5}$ concentration measured by the Grimm EDM 164 and AeroCet 831 were compared to find the correction factor. Both the devices were initiated at the same time and the data was taken for a minimum of 24 hours before the use of the instruments in the actual field. MS Excel was used for data analysis, where the trend lines between $PM_{2.5}$ measured by Grimm and AeroCet were selected for calibration. The linear relationships between the two devices were established using linear regression with coefficient of determination (R^2) close to 0.9 (for minimum deviation of data) as shown in figure 3.6 (b). The trends and timing of data from the devices were adjusted so that the peaks observed in the data followed the same trend as shown in figure 3.6 (a). Slope intercept equation was determined. Finally the measured values observed in the field were adjusted using this equation.

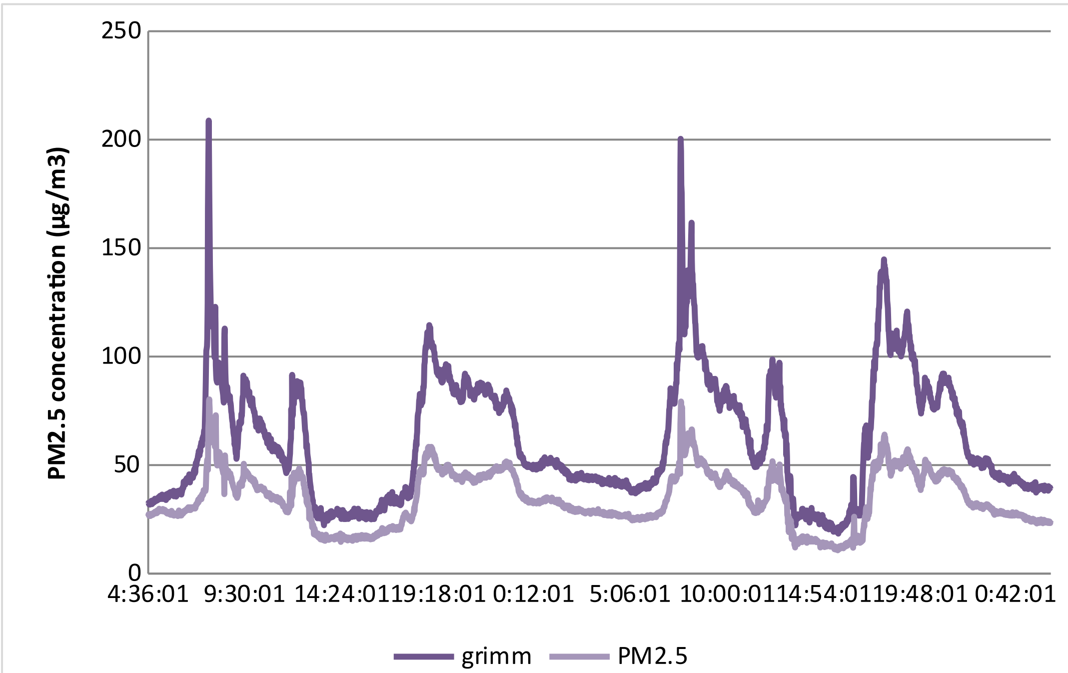


Figure 3.5 a): Calibration of data by collocating with Grimm

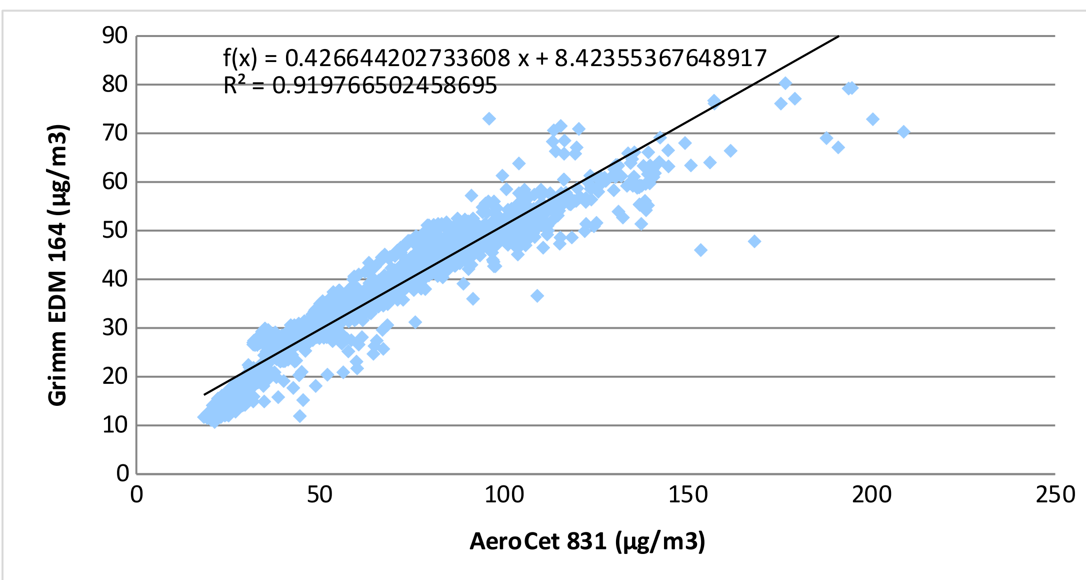


Figure 3.5 b): Linear relationship between Grimm and one of the AEROCETs

3.5 Ventilation Characterization and CO Prediction Modeling

3.5.1 Air Exchange Rate

The room was assumed to be a box of specified volume and the indoor pollutants were completely mixed. After the cooking was complete, the time to reach considerably low CO concentration was noted to measure the Indoor-to-Outdoor (I/O) Air Exchange Rate (AER). AER is a measure of air volume added or removed from the space per unit volume of the space and the unit is per hour (hr^{-1}). The AER is calculated as:

$$V\frac{dc}{dt}=Q.Co-Q.C(t)+S-K.C(t) \ldots (1)$$

Where,

V=volume of the room (m^3)

$\frac{dc}{dt}$=change in concentration with time

Q=air flow rate (m^3/min)

Co=Ambient pollutant concentration (ppm)

$C(t)$=Concentration at time t (min)

S=indoor emission rate ($mg.min^{-1}$)

K=first order degradation constant (min^{-1})

In a conservative equation K=0 and after the source is no longer burning, S=0 hence, equation (1) becomes,

$$V\frac{dc}{dt}=Q.Co-Q.C(t) \ldots (2)$$

Integrating equation (2), we get,

$$C(t)=Co-[C(0)-Co].e^{Q/V} \ldots (3)$$

Where,

$C(0)$=Concentration at t=0 (ppm)

Q/V=Air Exchange Rate (min^{-1})

Hence, equation (3) can be re-written as,

$$AER = \frac{1}{t} . \ln \left[\frac{C(0) - Co}{C(t) - Co} \right] \ldots \textbf{(4)}$$

This equation provides a method based on two concentration points during the decay phase. While used for initial characterization (Section 4.2.1), this approach can be highly sensitive to measurement noise and the specific points chosen, potentially limiting the reliability of AER estimates derived solely from it in field conditions.

Equation (4) was used to calculate the AER using Python.

3.5.2 Machine Learning Model for CO Prediction

a) Rationale for Employing Machine Learning:

Preliminary analysis using a standard single-box model to predict Carbon Monoxide (CO) concentrations revealed significant limitations in capturing the observed dynamics within the study kitchens. Assumptions inherent in simple box models, such as complete air mixing, constant emission rates during cooking, instantaneous emission cessation, and simplified ventilation parameters, often fail to represent the complex, transient conditions present during real-world biomass combustion events in naturally ventilated spaces. Observed factors like inconsistent fuel addition, variable stove operation, potential smoke back-flow, and complex airflow patterns contribute to CO concentration profiles that deviate substantially from idealized exponential decay or simple build-up curves.

To overcome these limitations and develop a more accurate predictive tool based on the available field measurements, a data-driven machine learning approach was adopted. The objective shifted from validating a simple mechanistic model to building a predictive model capable of learning the complex, non-linear relationships between observed CO concentrations and various influencing factors directly from the measurement data.

b) Model Selection: Random Forest Regressor:

Among various machine learning algorithms suitable for supervised regression tasks, the Random Forest Regressor was selected for this study. Key advantages favoring this choice include:

1. **Non-Linear Handling:** Random Forests inherently capture complex, non-linear relationships between predictors and the target variable without requiring explicit definition of these relationships (unlike linear models). This is crucial given the expected non-linearities in emission and dispersion processes.
2. **Feature Interaction:** The tree-based structure naturally handles interactions between different input features (e.g., how ventilation might affect CO differently depending on whether the stove is actively cooking).
3. **Robustness to Overfitting:** By aggregating predictions from multiple de-correlated decision trees (ensemble learning), Random Forests are generally less prone to over-fitting compared to individual decision trees, especially with appropriate hyper-parameter tuning.
4. **Feature Importance:** The algorithm provides a readily interpretable measure of feature importance, allowing identification of the key factors driving CO concentration predictions within the model.
5. **Data Type Handling:** It effectively handles a mix of numerical (e.g., volume, lagged concentrations) and categorical (e.g., stove type, although handled via encoding here) features.
6. **Scalability & Efficiency:** It generally performs well on tabular datasets of moderate size, such as the one compiled in this study, without the extensive data requirements often associated with deep learning models (like LSTMs) for time-series tasks.

c) Random Forest Algorithm Conceptual Overview:

The Random Forest algorithm operates by constructing a multitude of individual decision trees during training and outputting the average of the predictions from these individual trees. Its effectiveness stems from introducing randomness in two key ways to ensure the trees are diverse (de-correlated):

1. **Bagging (Bootstrap Aggregating):** Each individual decision tree in the forest is trained on a different random subset of the original training data, sampled *with*

replacement (a bootstrap sample). This means some data points may appear multiple times in a sample for one tree, while others might be omitted entirely.

2. **Random Feature Sub-space:** When splitting a node during the construction of an individual tree, the algorithm does not consider all available input features. Instead, it considers only a random subset of the features to determine the best split. This prevents strong predictors from dominating all trees and increases the diversity of the forest.

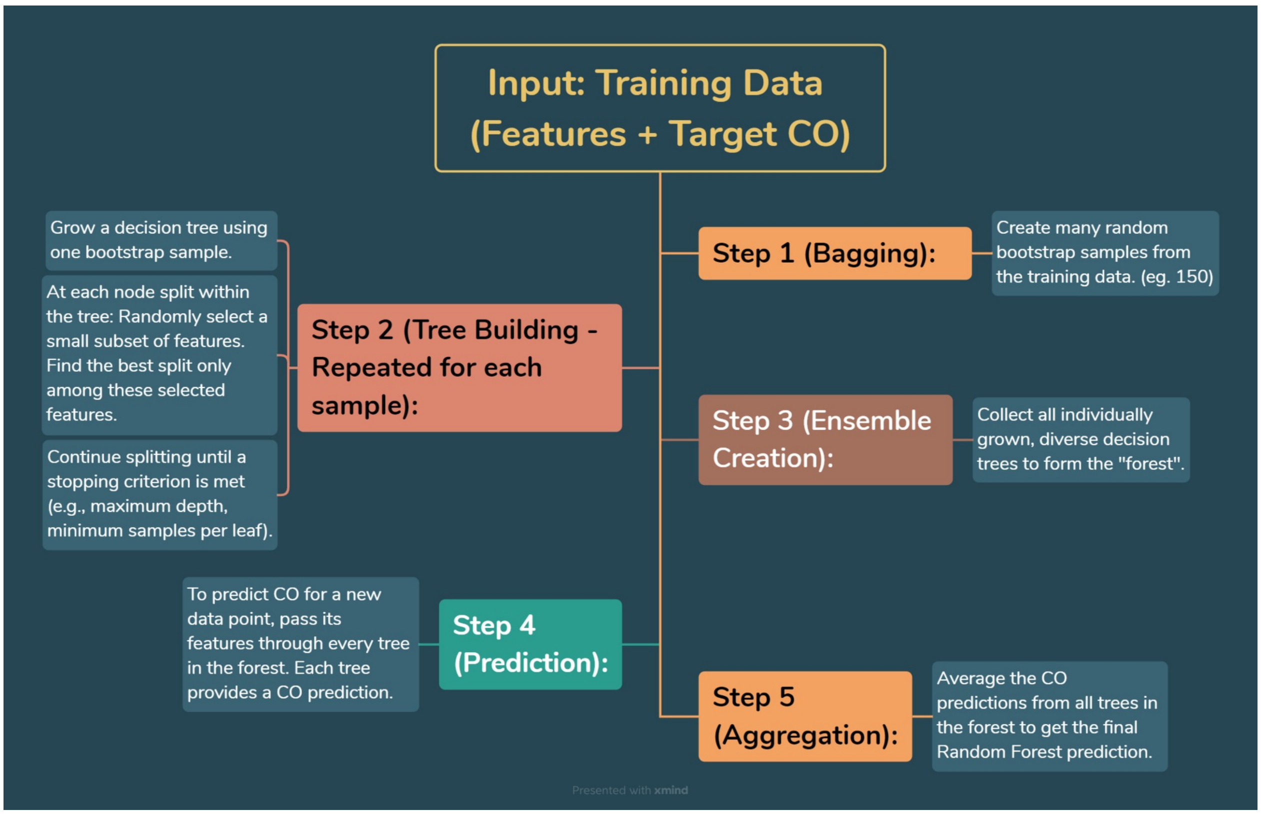


Figure 3.6: Conceptual overview of Random Forest Algorithm

d) Data Preparation and Feature Engineering:

The machine learning model was trained using a structured dataset compiled from the raw field measurements across all participating households. The preparation involved several steps, executed via Python scripts using the pandas library (see Annexes for script details):

1. **Consolidation:** Data from all household measurement sessions were combined into a single dataset.

2. **Timestamp Creation:** A complete timestamp (date and time) was created for each measurement record.

3. **Data Cleaning:** Basic cleaning included handling potential non-numeric CO values and ensuring concentrations were non-negative.

4. **Merging Auxiliary Data:** Static information specific to each household (*Volume_m3*, *StoveType*) and the session's estimated ambient CO level (*AmbientCO_ppm_Est*) were merged from the *household_info.csv* file based on *HouseholdID*. The ambient CO estimate used the measured pre-cooking value where available, otherwise resorting to the minimum CO observed during the pre-cooking phase (Phase 1) for that session.

5. **Target Variable:** The target variable (y) for prediction was the measured Carbon Monoxide concentration at each minute, *CO_ppm*.

6. **Input Features (X):** The following features were selected or engineered to predict the target:

- **Static Household Features:**
  - *Volume_m3*: The measured kitchen volume (m^3).
  - *StoveType_TCS, StoveType_ICS* indicators derived from one-hot encoding the StoveType column (1 indicating presence, 0 otherwise).
- **Session Context Feature:**
  - *AmbientCO_ppm_est*: The estimated background CO concentration (ppm) for the specific session.
- **Temporal Features:**
  - *HourOfDay:* The hour (0-23) extracted from the Timestamp.
  - *MinuteOfHour:* The minute (0-59) extracted from the Timestamp.
- **Event-Based Features:**
  - *IsCooking*: A binary flag (1 if the timestamp falls within the identified cooking period [Phase 2], 0 otherwise). The cooking period start/end times were determined programmatically based on Phase 2 data for each session.
  - *TimeSinceCookStart_min*: Time elapsed in minutes since the start of the cooking phase for the current session (set to 0 before cooking begins).

- **Autoregressive Features (Lagged CO):**
    - *CO_ppm_Lag1, CO_ppm_Lag2, CO_ppm_Lag5, CO_ppm_Lag10*: The measured CO concentration from 1, 2, 5, and 10 minutes prior to the current timestamp, respectively. These were calculated using the shift() function grouped by household session to capture the temporal dependence of CO levels.

Table 3.1: Input features for our model

Features	**Remarks**
Volume_m3	Kitchen volume.
AmbientCO_ppm_Est	Estimated ambient CO concentration for the session.
HourOfDay, MinuteOfHour	Temporal indicators derived from the timestamp.
IsCooking	Binary indicator (1 if during the main cooking phase [Phase 2], 0 otherwise)
TimeSinceCookStart_min	Minutes elapsed since the start of the cooking phase.
StoveType_ICS, StoveType_TCS	One-hot encoded representation of the stove type.
CO_ppm_Lag1, CO_ppm_Lag2, CO_ppm_Lag5, CO_ppm_Lag10	Lagged CO concentrations from 1, 2, 5, and 10 minutes prior.

7. Handling Missing Values: Rows containing NaN values, primarily occurring at the beginning of each household session due to the creation of lagged features, were removed from the dataset before training and testing using *dropna()*.

The final dataset used for modeling (*ML_Ready_Data.csv*) contained 2307 rows after cleaning and lag creation.

e) Model Training, Validation, and Hyperparameters:

To ensure an unbiased evaluation of the model's ability to generalize to unseen households, the dataset was split using *sklearn.model_selection.GroupShuffleSplit* This method partitions the data based on the *HouseholdID*, allocating approximately 75% of the unique households to the training set and the remaining 25% to the test set. This guarantees that all data from a specific household belongs exclusively to either the training or the testing partition, preventing overly optimistic results that might arise from randomly splitting time series data within the same household.

The *RandomForestRegressor* model was then instantiated and trained using the *fit()* method solely on the training data (X_train, y_train). The following key hyperparameters were used for (X_test, y_test) the final reported model:

- n_estimators:150 (Number of trees in the forest)
- max_depth: 15 (Maximum depth allowed for each tree)
- min_samples_leaf: 5 (Minimum number of samples required to be at a leaf node)
- random_state: 42 (Ensures reproducibility)
- n_jobs: -1 (Utilizes all available CPU cores for faster training)

These hyperparameters were chosen as reasonable defaults based on common practice and preliminary runs.

f) Model Evaluation:

The performance of the trained Random Forest model was assessed exclusively on the held-out test set (X_test, y_test). Predictions (y_pred) were generated for the test set using the predict() method. The following metrics were calculated to quantify predictive accuracy:

1. **Root Mean Squared Error (RMSE):** Measures the standard deviation of the prediction errors (residuals) in the original units (ppm). Lower values indicate better fit.

$$RMSE = \sqrt{\frac{1}{n} * \Sigma \left(y_{test} - y_{pred} \right)^2} \; \textbf{... (5)}$$

2. **Mean Absolute Error (MAE):** Measures the average absolute difference between observed and predicted values, also in the original units (ppm). It is less sensitive to large

outliers than RMSE.

$$MAE = \frac{\Sigma \left| y_{test} - y_{pred} \right|}{n} \quad \textbf{… (6)}$$

3. **Coefficient of Determination (R^2):** Represents the proportion of the variance in the observed target variable (y_test) that is predictable from the input features using the model. Values range from negative infinity to 1, with 1 indicating a perfect fit and 0 indicating the model performs no better than predicting the mean.

$$R^2 = 1 - \frac{\Sigma \left(y_{test} - \hat{y}_{pred} \right)^2}{\Sigma \left(y_{test} - \overline{\left(y_{test} \right)} \right)^2} \quad \textbf{… (7)}$$

In addition to these quantitative metrics, visual comparisons of the observed vs. predicted time series plots for representative test households were performed. Feature importances were also extracted from the trained model to understand the relative contribution of different input variables to the predictions.

g) Software:

Data processing and analysis were performed using Python (Version 3.11). Key libraries utilized include latest versions of *pandas* for data manipulation, *numpy* for numerical operations, *scikit-learn* for machine learning model implementation and evaluation, *scipy* for statistical functions, and *matplotlib* & *seaborn* for data visualization.

3.6 General Overview

The following chart (fig 3.8) provides a general overview of how we conducted the project to achieve the results.

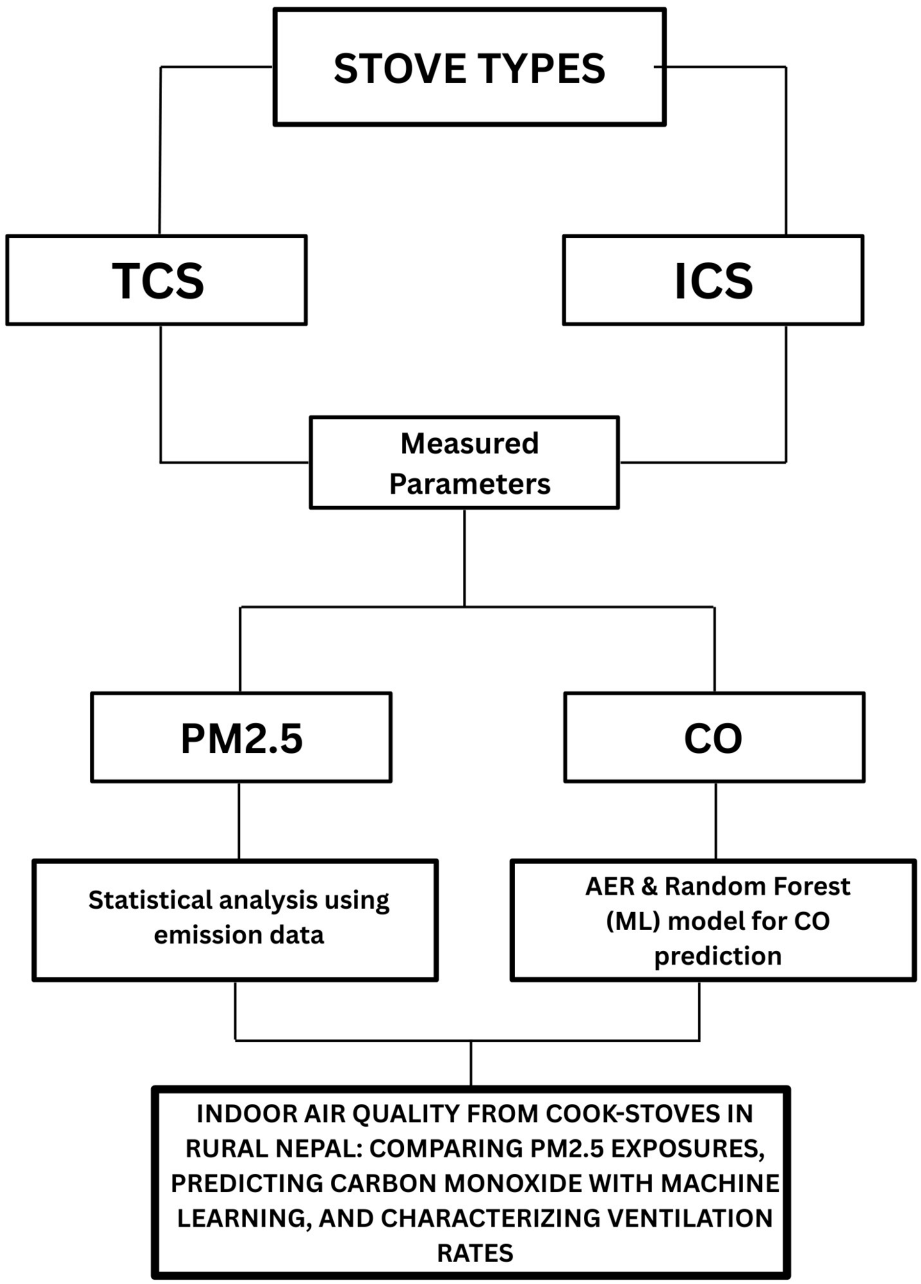


Figure 3.7: Schematic representation of the project methodology

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Real- time exposure measurement

The indoor and personal concentration of $PM_{2.5}$ was measured in 38 households of Salambu and Pokhara which had either ICS or TCS as their main source of cooking. However, some of the measured data didn't give favorable results due to device defects and accidental errors. So these unwanted data were removed and the sample size was narrowed down to 32 households for indoor concentration and 29 households for personal concentration. Under indoor concentration, there were 15 ICS and 17 TCS samples while under personal concentration, there were 13 ICS and 16 TCS samples.

The $PM_{2.5}$ concentration increased for both conditions throughout the cooking period particularly at the start and end of cooking. The average, range and peak of pollutant concentration for the two stove designs differed significantly. The real-time indoor mean concentration and standard deviation of $PM_{2.5}$ during cooking period was 109.7 $\mu g/m^3$ (σ = 29.3 $\mu g/m^3$) for ICS and 314.5 $\mu g/m^3$ (σ = 120.4 $\mu g/m^3$) for TCS. Similarly, the real-time personal mean and standard deviation was 109.4 $\mu g/m^3$ (σ = 26.0 $\mu g/m^3$) for ICS and 248.9 $\mu g/m^3$ (σ = 121.1 $\mu g/m^3$) for TCS. The average indoor concentration of $PM_{2.5}$ emission using ICS was ranged from 60.9 $\mu g/m^3$ to 161.93 $\mu g/m^3$ and using TCS was ranged from 177.9 $\mu g/m^3$ to 601.8 $\mu g/m^3$. Similarly for personal conditions, the emission using ICS was ranged from 70.9 $\mu g/m^3$ to 156.5 $\mu g/m^3$ and using TCS was ranged from 136.2 $\mu g/m^3$ to 596.2 $\mu g/m^3$ (refer to Table 4.1 and 4.2).

Figure 4.1 and 4.2 gives a graphical representation of real-time indoor and personal $PM_{2.5}$ concentration during the entire cooking period of households having ICS and TCS. Both the indoor and personal measurements were conducted at the same time and so from the graphs, we can compare the difference in $PM_{2.5}$ emission between the two conditions. Between the two scatter plots, TCS seems to have much higher $PM_{2.5}$ emitted than ICS during cooking. Also, the $PM_{2.5}$ emission in personal condition has a significant decrease in exposure when compared with indoor condition due to continuous movement of the participant in and around the kitchen. We can clearly observe that the personal $PM_{2.5}$ concentration measured is of a much lesser value in some cases when the indoor concentration measured is high at that time. This tells us that other family members especially children are also heavily affected by the pollutant exposure due to smoke retained indoor.

Table 4.1: Comparison between ICS and TCS for indoor conditions

Stove type	N	Mean (μ) (μg/m³)			SD (σ) (μg/m³)			Min. value (during)	Max. value (during)	Range (during)
		Before	During	After	Before	During	After			
Improved	15	40.92	**109.6**	45.5	19.90	**29.26**	13.60	60.957	161.93	100.97
Traditional	17	40.32	**314.5**	63.5	17.30	**120.4**	29.45	177.875	601.793	423.91
Change in reduction		65%								
T statistics		-6.79								
P value		2.2e-06								

Table 4.2: Comparison between ICS and TCS for personal conditions

Stove type	N	Mean (μ) (μg/m³)			SD (σ) (μg/m³)			Min. value (during)	Max. value (during)	Range (during)
		Before	During	After	Before	During	After			
Improved	13	40.66	**109.3**	35.4	12.30	**26.04**	9.524	70.894	156.47	85.576
Traditional	16	40.35	**248.9**	41.6	20.69	**121.1**	12.727	136.24	596.22	459.98
Change in reduction			56.1%							
T- statistics			-4.484							
P value			0.0003							

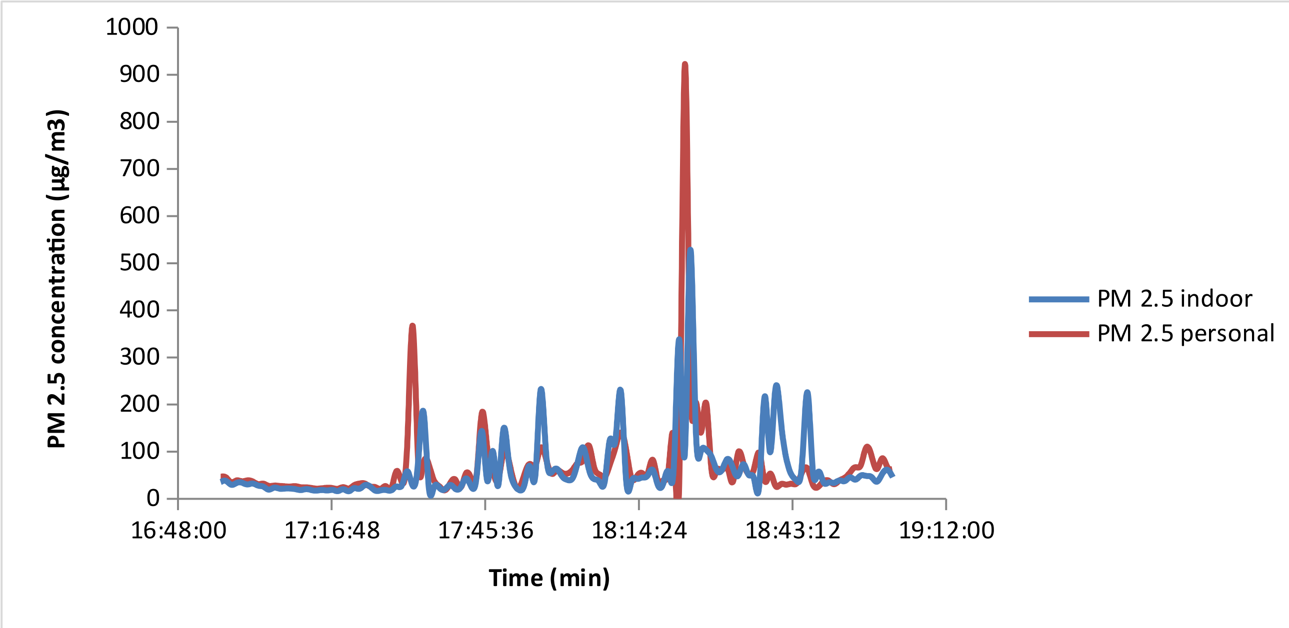


Figure 4.1: Real time indoor and personal $PM_{2.5}$ concentration for ICS during cooking period

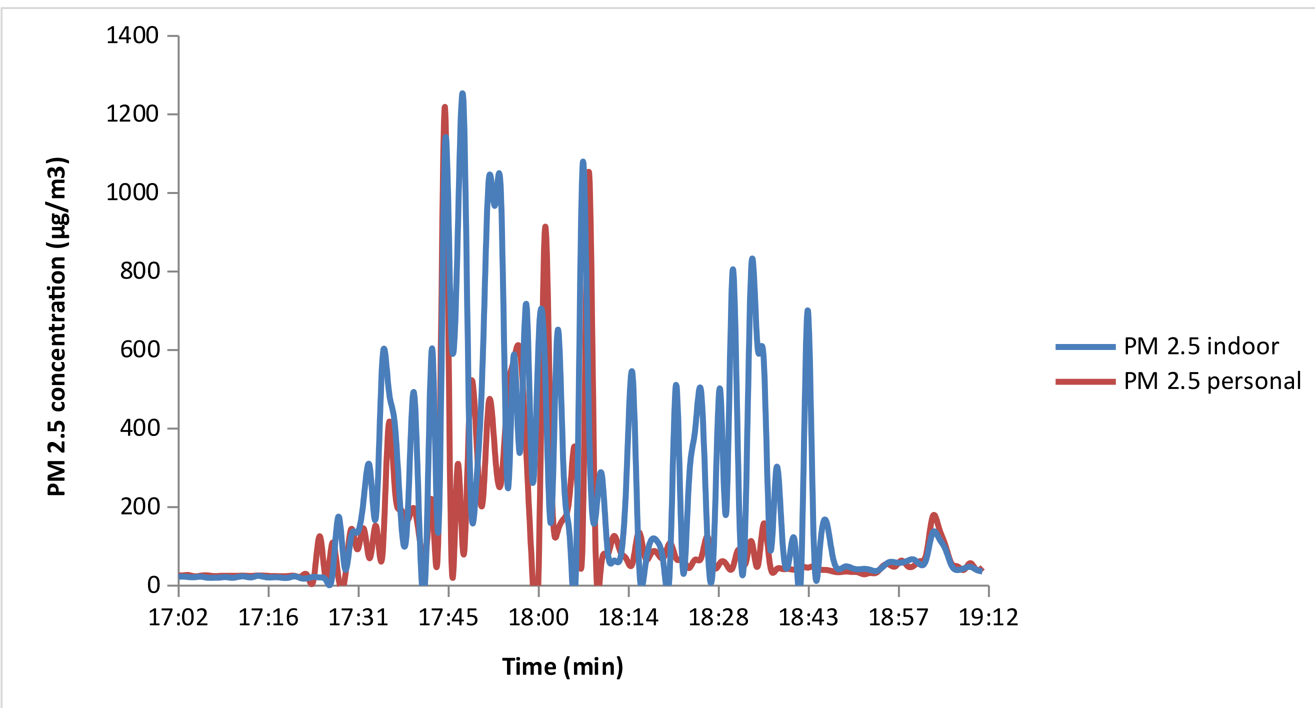


Figure 4.2: Real time indoor and personal $PM_{2.5}$ concentration for TCS during cooking period

4.2 Significance tests and box-plots

There is a 65% and 56.1% reduction in concentration for $PM_{2.5}$ of indoor and personal condition respectively between TCS and ICS as shown in Table 4.1 and 4.2. A two sample t test between the two stove designs suggests that there is an extremely significant reduction of pollutant concentration in both indoor ($p < 0.0001$) and personal ($p < 0.001$) conditions. Figure 4.3 and 4.4 shows a boxplot comparing $PM_{2.5}$ concentration emitted by the two stove types in the indoor and personal conditions respectively.

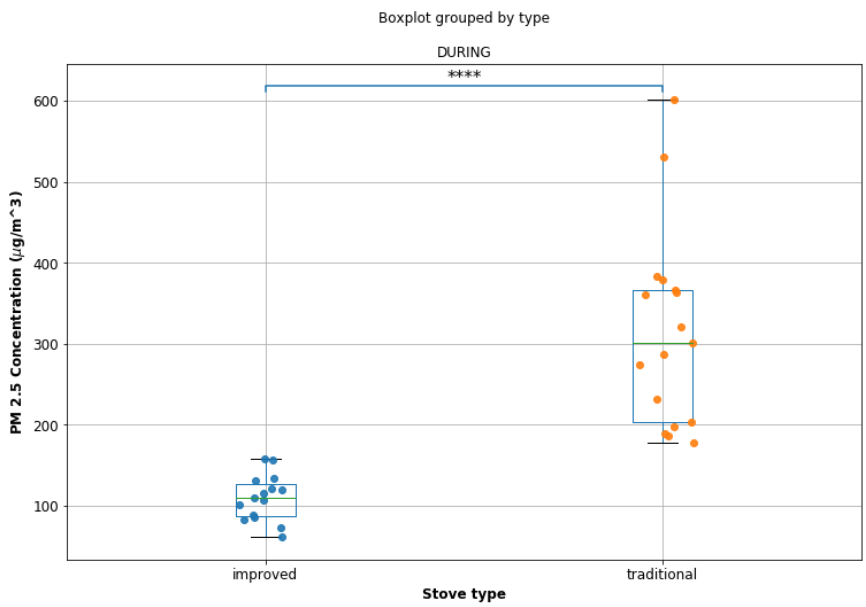


Figure 4.3: Box plot grouped by stove type showing $PM_{2.5}$ distribution for Indoor conditions (**** = p<0.0001)

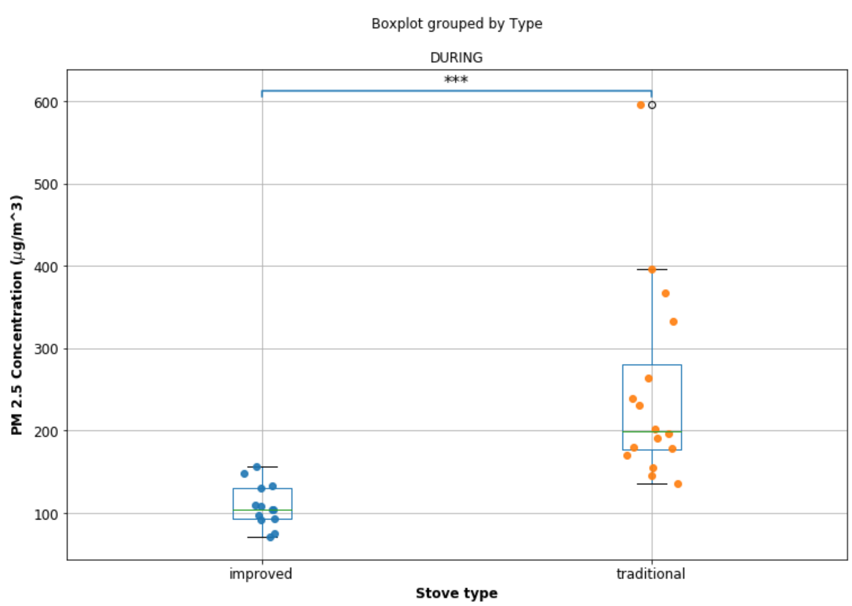


Figure 4.4: Box plot grouped by stove type showing $PM_{2.5}$ distribution for personal conditions (*** = p<0.001)

Furthermore, graphs representing $PM_{2.5}$ emitted before, during and after cooking period for indoor and personal condition was plotted respectively and analyzed for comparing the differences between the stove types. The graph is as shown in figure 4.5 and 4.6. We can clearly see from the graph that the ICS and TCS samples in both the conditions have massive difference in $PM_{2.5}$ emission and that TCS has a much higher emission range than ICS.

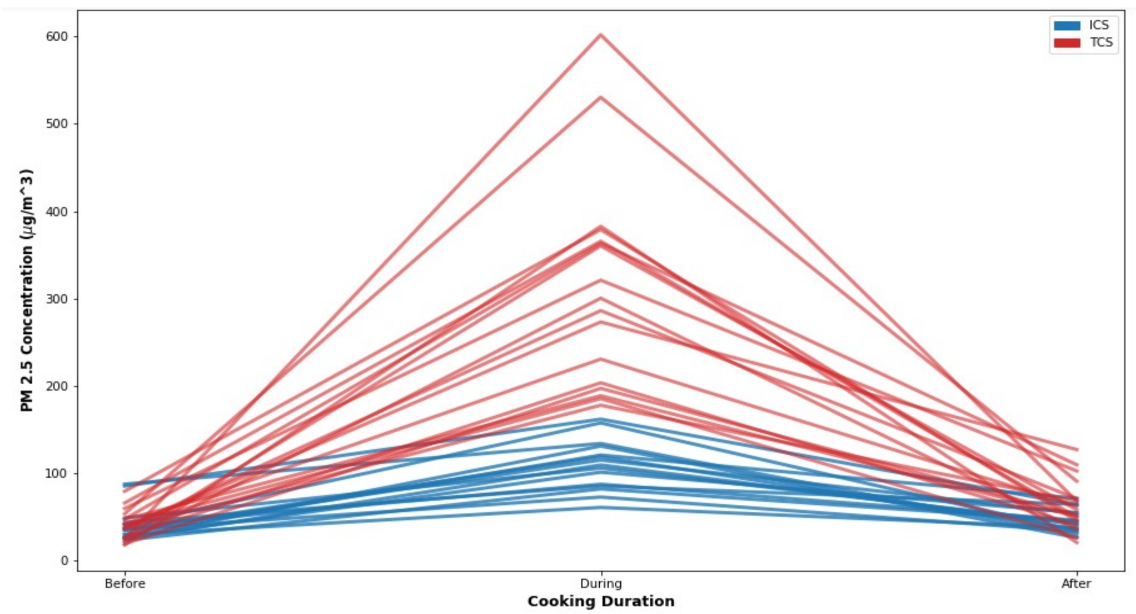


Figure 4.5: Variation in $PM_{2.5}$ before, during and after cooking for ICS and TCS in indoor condition.

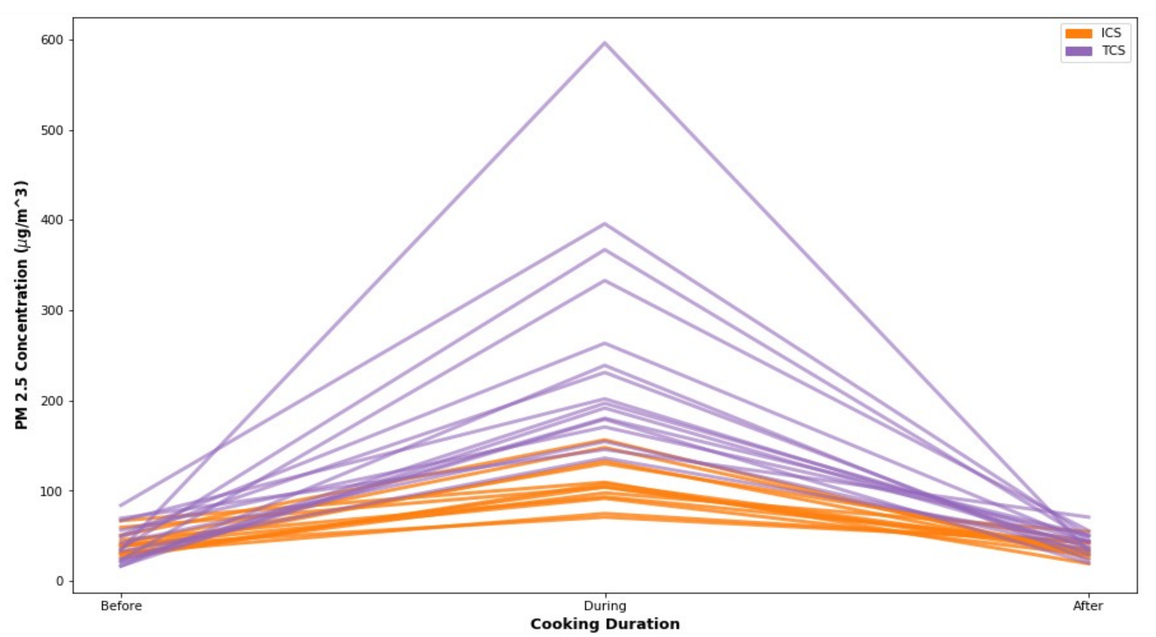


Figure 4.6: Variation in $PM_{2.5}$ before, during and after cooking for ICS and TCS in personal condition.

4.3 Comparison through distribution

We also graphed and compared the normal distribution of ICS and TCS for indoor conditions in figure 4.7. Normal distribution shows how often the emission measured will converge in distribution to produce a particular result that is close to the normal i.e. mean. From our result we can analyze the distribution of ICS to be much taller and skinnier than TCS due to a smaller mean and standard deviation. TCS on the other hand, has a wider and shorter distribution due to a larger mean and standard deviation. This clearly shows a heavy difference in the distribution and deviation of $PM_{2.5}$ emission from the two stove designs.

TCS on the other hand, has a wider and shorter distribution due to a larger mean and standard deviation

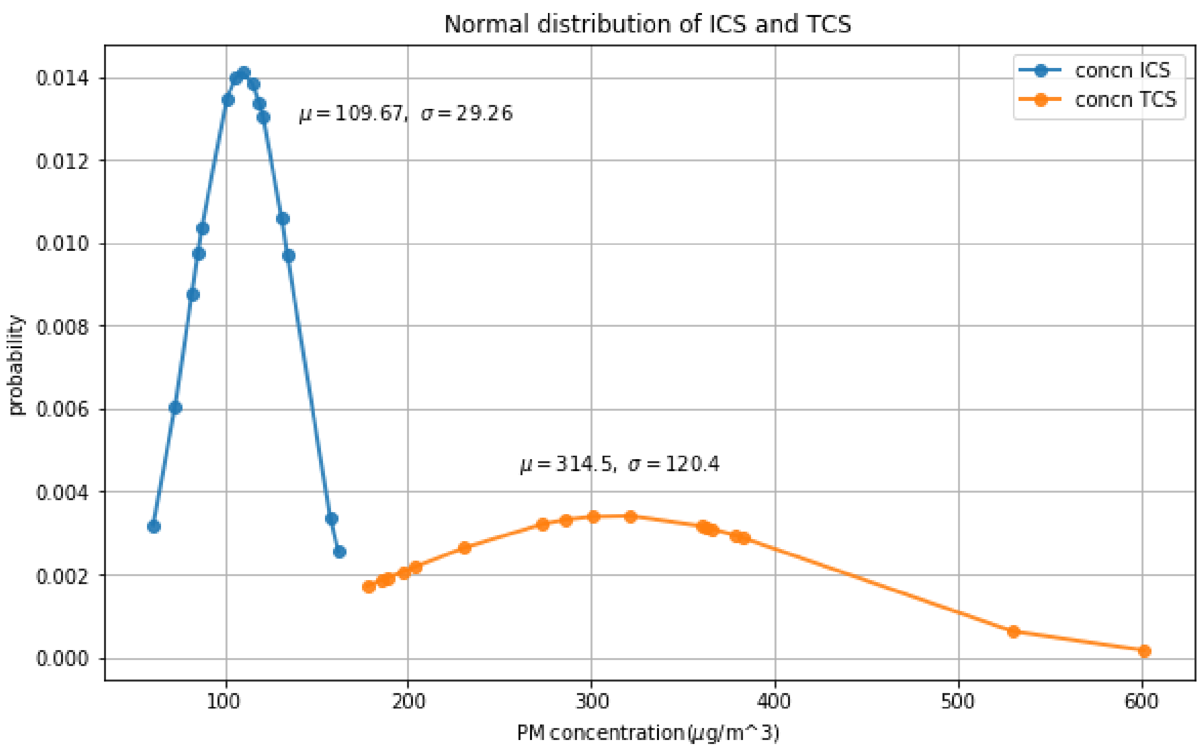


Figure 4.7: Comparison of Normal Distribution given by ICS and TCS.

4.4 Ventilation characteristics and CO prediction:

4.4.1 Ventilation Characterization

Air Exchange Rates (AER) were calculated using the CO decay curve method (Equation 4) to characterize the ventilation conditions in the sampled kitchens. The calculated AER values ranged from 0.2 ACH to 11.6 ACH across the households (Table 4.3 summarizes data used for these initial calculations). Figure 4.1 shows the relationship between the calculated AER and the time taken for CO concentrations to decay; the weak correlation (R^2 = 0.2142) suggests considerable variability not captured by this simple relationship, potentially due to limitations in the decay method application or complex airflow dynamics. Similarly, Figure 4.8 illustrates the expected inverse trend between AER and average CO concentration during cooking, but again with significant scatter, implying other factors like emission rate fluctuations strongly influence in-kitchen levels. Comparing the calculated AER values to the ASHRAE recommendation of 5 ACH for residential buildings revealed that less than 50% of the sampled kitchens met this guideline, indicating generally low ventilation levels. It is important to acknowledge that the two-point decay method used for these initial AER characterizations can be sensitive to noise and non-ideal conditions, potentially affecting the accuracy and reliability of individual AER estimates.

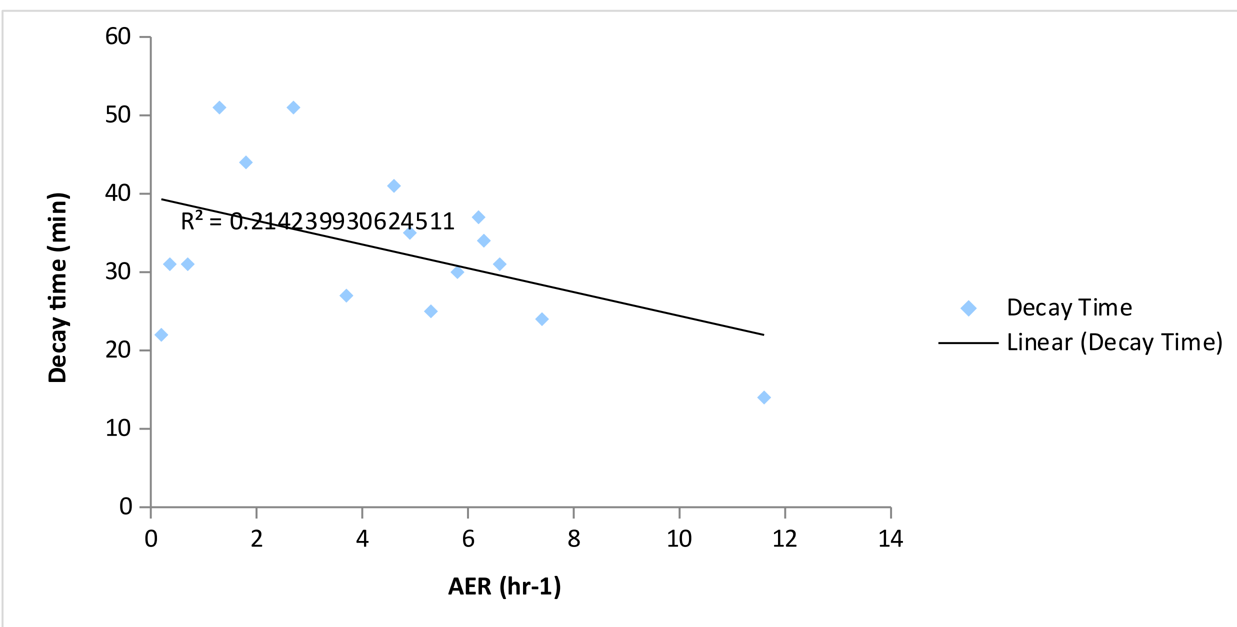


Figure 4.8: Relationship between AER and CO decay time

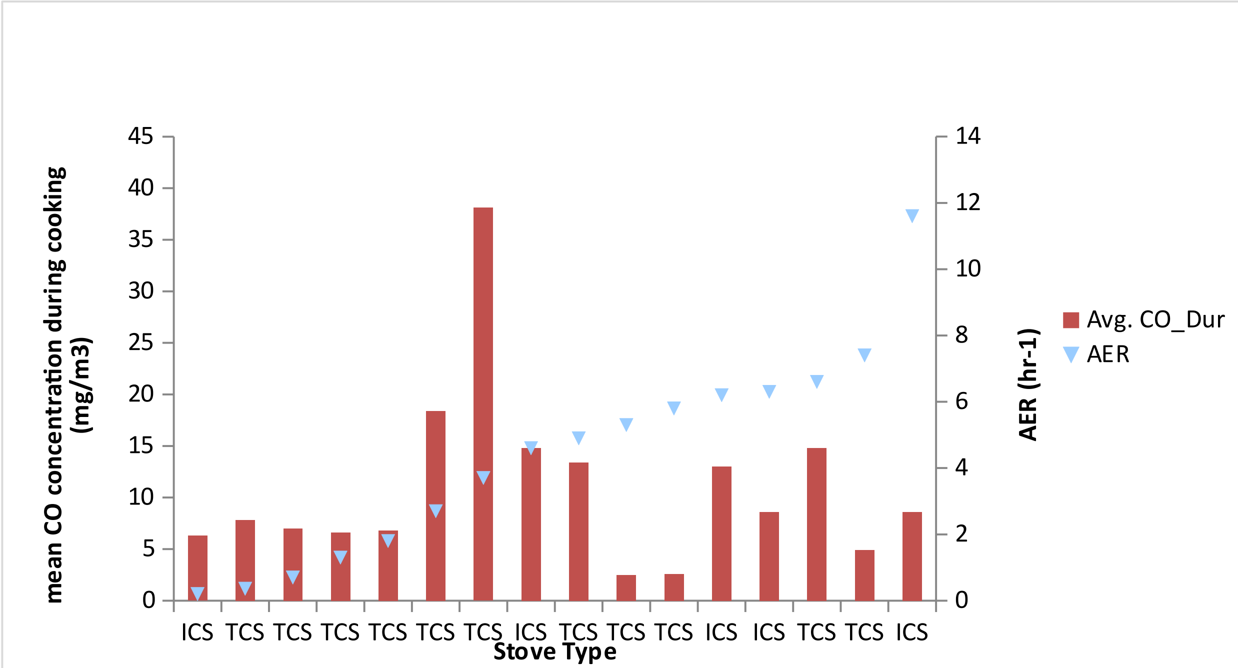


Figure 4.9: AER vs CO concentration during cooking for different stove types in different sample houses

Table 4.1: Relationship between AER, CO decay time and concentration during cooking.

Stove Type	AER	Avg. CO_Dur	Decay Time	House No.	Remarks
ICS	0.2	6.3	22	15	
TCS	0.4	7.8	31	25	
TCS	0.7	7	31	12	
TCS	1.3	6.6	51	24	
TCS	1.8	6.8	44	23	
TCS	2.7	18.4	51	20	
TCS	3.7	38.1	27	22	Kerosene lamp burnt
ICS	4.6	14.8	41	11	Secondary rocket type stove used
TCS	4.9	13.4	35	19	
TCS	5.3	2.5	25	18	Roof opening
TCS	5.8	2.6	30	17	Straw Walls
ICS	6.2	13.0	37.0	5	
ICS	6.3	8.6	34	13	
TCS	6.6	14.8	31	8	
TCS	7.4	4.9	24	9	
ICS	11.6	8.6	14	7	

4.4.2 CO Prediction using Random Forest

Given the challenges in accurately predicting CO using the simple box model, a Random Forest (RF) machine learning model was developed as described in Section 3.5.1.

a) Model Performance:

The RF model was trained on 75% of the households and evaluated on the remaining 25% held-out test set. The performance metrics achieved were:

- Root Mean Squared Error (RMSE): 16.69 ppm
- Mean Absolute Error (MAE): 5.62 ppm
- Coefficient of Determination (R^2): 0.915

An R^2 value of 0.915, as shown in figure 4.10 indicates that the model explains approximately 91.5 % of the variance in CO concentrations in unseen households. The RMSE of 16.69 ppm represents the typical magnitude of prediction error, which compares favorably well to the observed peak CO concentrations (ranging up to 250 ppm in the test set). These metrics suggest the data-driven RF model provides a substantially better predictive capability compared to the initial box model attempts.

```
--- Starting Random Forest Training & Evaluation ---
Loading ML-ready data from: ML_Ready_Data.csv
  Successfully loaded 2306 rows and 15 columns.
  Columns: ['HouseholdID', 'Timestamp', 'Volume_m3', 'AmbientCO_ppm_Est', 'HourOfDay', 'MinuteOfHour', 'IsCooking', 'TimeSinceCookStart_min', 'StoveT
ype_ICS', 'StoveType_TCS', 'CO_ppm_Lag1', 'CO_ppm_Lag2', 'CO_ppm_Lag5', 'CO_ppm_Lag10', 'CO_ppm']
Defining features and target...
  Target: 'CO_ppm'
  Features (12): ['Volume_m3', 'AmbientCO_ppm_Est', 'HourOfDay', 'MinuteOfHour', 'IsCooking', 'TimeSinceCookStart_min', 'StoveType_ICS', 'StoveType_T
CS', 'CO_ppm_Lag1', 'CO_ppm_Lag2', 'CO_ppm_Lag5', 'CO_ppm_Lag10']
Splitting data into train/test sets (75%/25%), keeping households intact...
  Training set: 1729 rows from 15 households.
  Test set:     577 rows from 5 households.
  Test Households: [ 5  6 13 21 23]

Training RandomForestRegressor model...
  Parameters: n_estimators=150, max_depth=15, min_samples_leaf=5, n_jobs=-1
  Training complete. Time taken: 0:00:00.503607
Making predictions on the test set...

Evaluating model performance on the test set...

--- Test Set Evaluation Metrics ---
RMSE: 16.69 ppm
MAE:  5.62 ppm
R²:   0.915
```

Figure 4.10: Performance metrics of test set.

b) Feature Importance:

The relative importance of the input features, as determined by the trained RF model, is shown in Figure 4.11. The most influential feature was *CO_ppm_Lag1* highlighting the strong auto-correlation of CO levels minute-to-minute.

Other important predictors included *CO_ppm_Lag2, TimeSinceCookStart_min, CO_ppm_Lag10* among others, indicating the significance of recent concentration history, active cooking status, and kitchen characteristics in determining subsequent CO levels. Features like *Volume_m3, StoveType_TCS and StoveType_ICS* had negligible importance in this model.

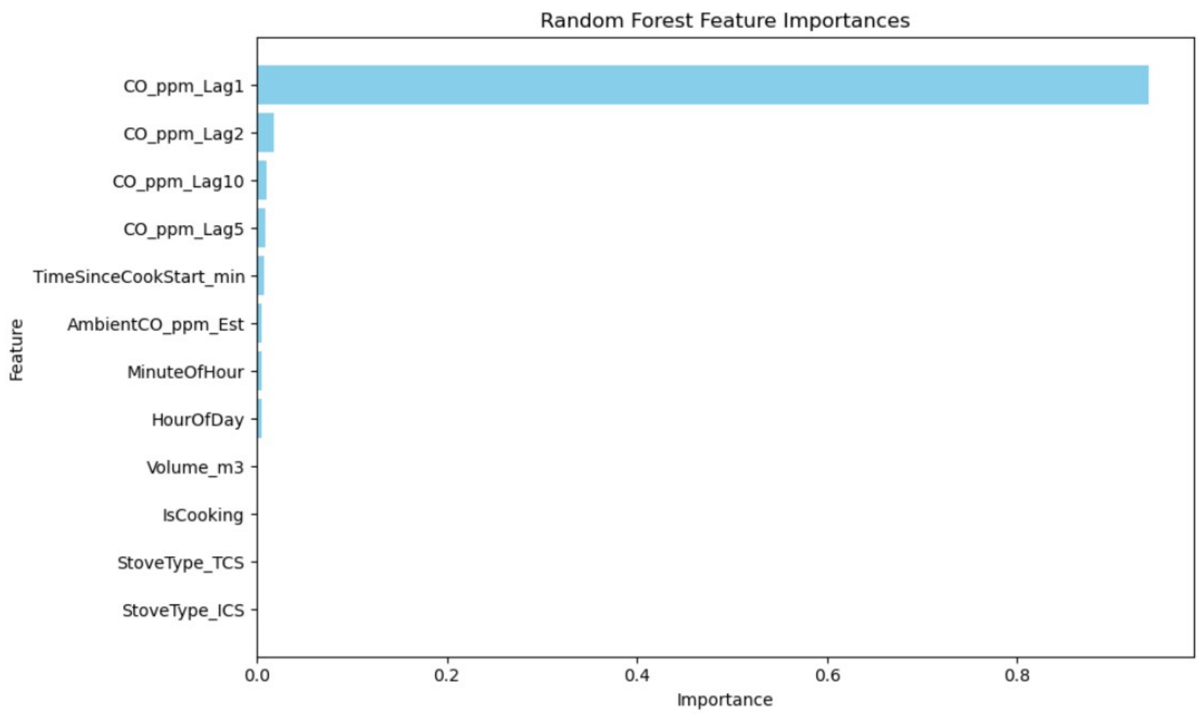


Figure 4.11: Feature importances of our Random Forest model

c) Time-Series Prediction:

Figures 4.12 a-d illustrate the model's predictions (red dashed line) against the observed CO concentrations (blue line) for four sample households from the test set.

Household ID 5, TCS: For Household 5, the model successfully captures the timing and approximate magnitude of the main CO peak during the cooking phase, as well as the subsequent decay. However, it tends to spike out some of the low-ranging, short-term peaks observed during active cooking, suggesting quite a random behavior in the emission for TCS, making it difficult for the model to predict accurately.

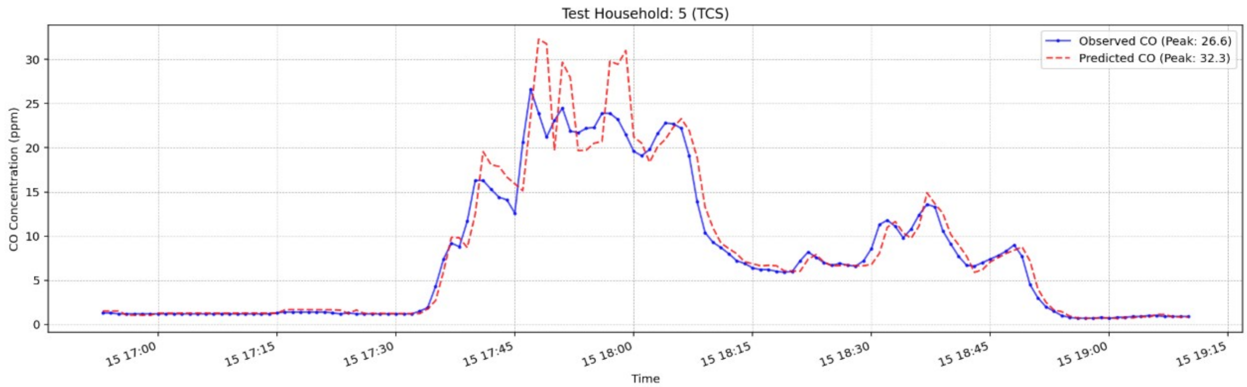


Figure 4.12 a: Observed vs Modeled CO for Household 5 (TCS)

Household ID 13, ICS: In Household 13, which exhibited lower overall concentrations, the model again follows the general trend effectively. The prediction during the decay phase closely matches the observed values. The modeled values behave more accurately in comparison to the TCS, which could lead to the strong possibility that ICS has a more regulated airflow and CO decay rate, leading to a much accurate prediction.

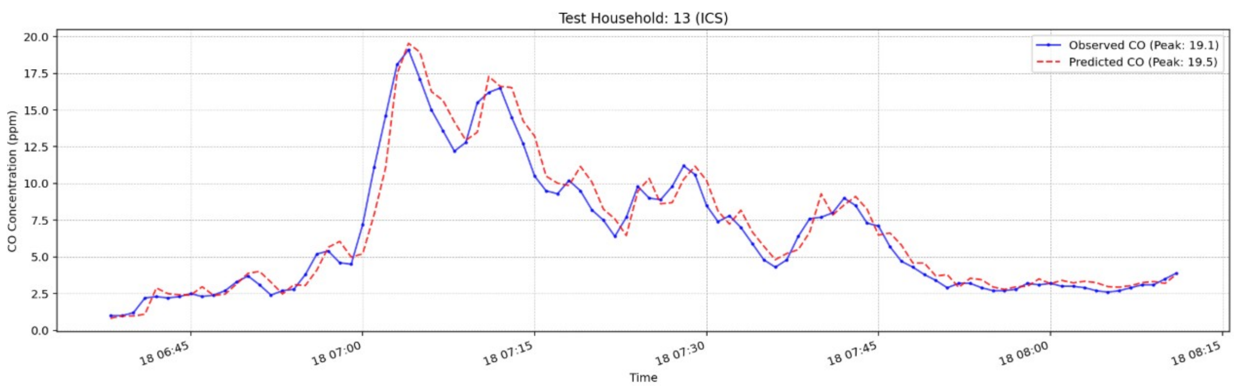


Fig 4.12 b: Observed vs Modeled CO for Household 13 (ICS)

Household ID 6, TCS*:* Household 6 (TCS) had a comparatively weak model performance, with a lot of concentration spikes peaking during the cooking phase not addressed properly. The predicted CO concentration had a drastic difference. This could potentially be due to a more random airflow and CO decay in the TCS cooking setup.

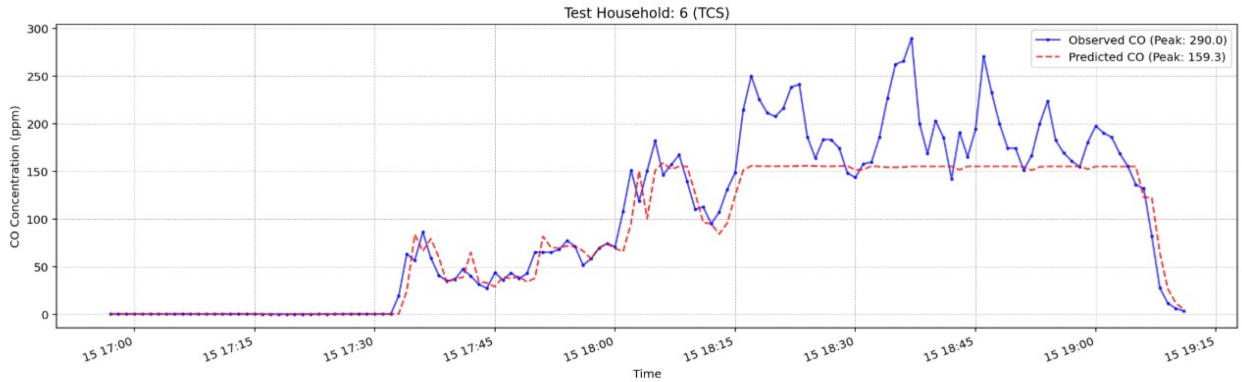


Figure 4.12 c: Observed vs Modeled CO for Household 6 (TCS)

Household ID 23, ICS: Household 23 (ICS) had a relatively low concentration of CO observed. The model effectively followed the general trend, making sure the prediction closely matched the observed values. In this instance as well, the modeled values behaved more accurately than the TCS, which leads to a strong evidence of ICS behaving more predictably than traditional ones.

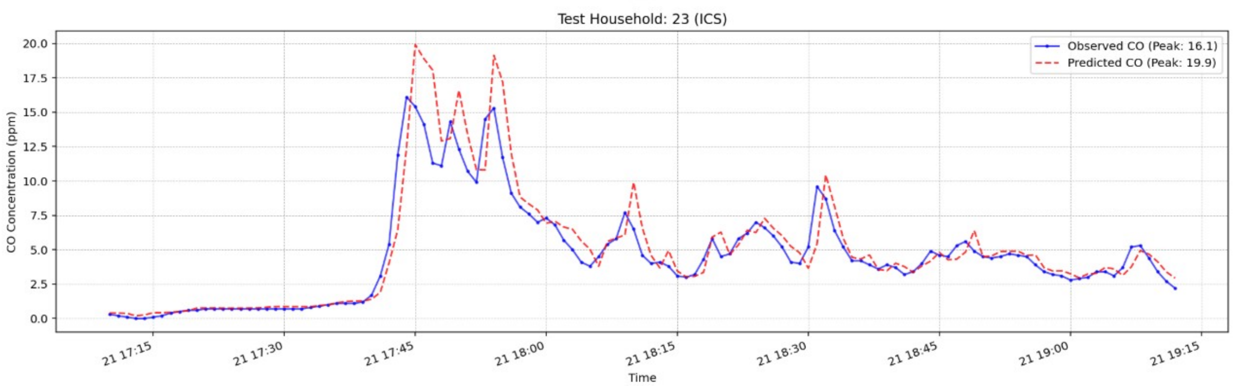


Figure 4.12 d: Observed vs Modeled CO for Household 6 (ICS)

Overall, these examples demonstrate the RF model's capability to learn complex temporal patterns from the data, significantly outperforming simple mechanistic assumptions, although precise prediction of rapid fluctuations remains challenging. Furthermore, the differences in the predictability of the TCS and ICS showed a strong suggestion that there is a significant variation in the air flow and decay rate patterns of the two stove types: TCS had a more unpredictable flow, with peaks of real-time CO concentration occurring at unexpected instances, making it harder for the model to predict an accurate result. ICS on the other hand, gave a remarkably accurate prediction of the CO concentration trend, suggesting that the emission from ICS could have a more organized variation in the airflow and CO decay rate.

For instance, in figure 4.12c, the peak CO concentration observed for household 6 (TCS) had a massive difference (> 100 ppm) in comparison with the modeled peak CO concentration. Whereas, the same difference can be considered almost negligible (≤ 4 ppm) when observing figure 4.12d for household 23 (ICS). These findings also suggest that cooking in ICS gave away lesser concentration of CO and had a more regulated airflow of pollutants, making the occupants' exposure to these pollutants comparatively minimal.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATION

After the completion of this project, we can come to a conclusion that through the installation of ICS, we could see a significant reduction in indoor air pollutants when compared with TCS. However, both the stoves seem to have emission levels exceeding the national and WHO threshold values. Hence, further study on stove types should be conducted to identify more ways to further reduce occupant's exposure to these pollutants.

To address CO prediction, initial box modeling proved insufficient. A Random Forest machine learning model, however, demonstrated good capability in predicting CO concentrations based on past levels, operational status, and kitchen characteristics (R^2= 0.915 on test data). Furthermore, while observing the CO emission trends for both the stove types, the model showed a more accurate predictability for ICS, while TCS had a comparatively weak predictability, with a lot of concentration spikes peaking during the cooking phase not addressed properly. This suggested a more random airflow and decay rate of CO while cooking in traditional setups, in comparison to the improved one. This leads to the possibility that cooking in ICS shows proper ventilation and airflow of pollutants when comparing with the TCS, helping it provide a better estimate of the emissions.

Although this improves prediction, switching to cleaner options (LPG, bio fuels) remains the most effective emission reduction strategy. Given the socioeconomic context, improved cook stoves (ICS) offer a feasible interim measure, contributing to lower exposure, especially when coupled with adequate ventilation facilitated by features like chimneys.

To reduce emission, switching to better cooking options such as LPG, bio fuels, etc. is the ultimate solution. But, due to an inconvenient socio economic background, ICS seems to be the most feasible stove choice for households in rural areas. Due to presence of 1.2 m high chimney in ICS, better ventilation was noted when comparing it with TCS.

Implementing better designs in cook stoves can help reduce the pollutant emission to a much larger scale. Because the major population using these stove designs have a low economic status, design of innovative, low cost cook stoves should be recommended for production.

Further investigation using larger datasets could refine the machine learning models for CO and potentially PM2.5 prediction, exploring additional features like fuel characteristics or specific ventilation opening details. While AER calculation via CO decay proved challenging here, alternative methods like tracer gas techniques could provide more robust ventilation data for future studies aiming to explicitly model its impact. Both cleaner cook stoves and proper ventilation are equally important to prevent backflow and reduce IAP. The introduction of ICS in kitchen without any proper ventilation will ultimately decrease the IAQ.

REFERENCES

Household Air Pollution and Health. (2018, May 08). Retrieved July` 24, 2018, from World Health Organization: http://www.who.int/en/news-room/fact-sheets/detail/household-air-pollution-and-health

Nepal Energy Situation. (2018, July 20). Retrieved from Energypedia: https://energypedia.info/wiki/Nepal_Energy_Situation

C.R. Chu, Y. C. (2015). Wind-Driven Natural Ventilation for Buildings with Two Openings on the Same External Wal. *ENB*.

C.-R. Chu, Y.-H. C.-T.-L. (2015). Title: Wind-Driven Natural Ventilation for Buildings with Two Openings on the Same External Wall. *ENB*.

Mozaffarian, R. (2009). *natural ventilation in buildings and the tools for analysis.* University of Florida.

Parajuli, I., Lee H., Shrestha K. (2016). Indoor Air Quality and ventilation assessment of rural mountainous households of Nepal. *International Journal of Sustainable Built Environment*.

Ramaswami A, J. B. (2005). *Integrated Environmental Modeling.* John Wiley & Sons.

World Health Organization. (2005). Indoor air pollution from solid fuels and risk of low birth weight and stillbirth. *Annual Conference of the International Society for Environmental Epidemiology (ISEE).* Johannesburg.

Zhai, Z., Mankibi, M. E., & Zoubir, A. (2015). Review of natural ventilation models. *6th International Building Physics Conference, IBPC 2015.*

Albalak, R. (2001). Indoor respirable particulate matter concentrations from an open fire, improved cook stove and LPG/open fire combination in a rural Guatemalan community. *Environ. Sci. Technology* , 2650-2655.

Alternative Energy Promotion Centre. "Rural Energy Policy", Alternative Energy Promotion Centre (AEPC), Kathmandu, Nepal. http://www.aepc.gov.np/images/pdf/REPolicy-2006.pdf.

Armendáriz et al. (2010). Indoor particle size distributions in homes with open fires and improved Patsari cook stoves. *Atmospheric Environment 44*

Bruce, N., Perez-Padilla, R., Albalak, R. (2000). Indoor air pollution in developing countries: a major environmental and public health challenge. *Bulletin of the World Health Organization* 78 (9), 1078e1092

Chen, C., Zeger, S. (2016). Estimating Indoor $PM_{2.5}$ and CO concentrations in Households in Southern Nepal: *The Nepal Cook stove Intervention Trials*. *PLoS ONE* , 11(7): e0157984. doi:10.1371/journal.pone.0157984.

Clark, M. (2010). Indoor air pollution, cook stove quality, and housing characteristics in two Honduran communities. *Environ.Res.110* , 12-18.

KC, B. (2015). The effect of enhanced stove design on 'real life' exposure to $PM_{2.5}$ and CO in rural dwellings in Salambu, Nepal

Parajuli, I., Lee H., Shrestha K. (2016). Indoor air quality and ventilation assessment of rural mountainous households of Nepal. *International Journal of Sustainable Built Environment*

Parikh, J., Balakrishnan K., Laxmi V., Biswas H. (2001). Exposure from cooking with biofuels: pollution monitoring and analysis for rural Tamil Nadu, India

Reid HF, Smith KR, Sherchand B. (2016). Indoor smoke exposures from traditional and improved cookstove comparisons among rural Nepali women. *Mt. Res. Dev*. 1986; 6(4): 293–303.

Rehfuess, E., Mehta, S, Pruss-Ustun, A. (2006). Assessing household solid fuel use: multiple implications for the millennium development goals. *Environ. Health Perspect*. 114 (3), 373–378.

Singh, Ashish, Tuladhar, Bhushan, Bajracharya, Karuna, Pillarisetti, Ajay (2012). Assessment of effectiveness of improved cook stoves in reducing indoor air pollution and improving health in Nepal. *Energy Sustainable Dev.* 16, 406–414.

Smith, KR. (2002). Indoor air pollution in developing countries: recommendations for research. *Indoor Air* 12 (3), 198–207.

Thapa, R.B., Shrestha, R.M. (2013). Metallic Improved Cook Stoves Dissemination in Mountain Region of Nepal: Experience, Financial Viability, Opportunity & Challenges

ANNEXES

1. Code for Random Forest Algorithm Training and Evaluation:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import GroupShuffleSplit, train_test_split # To split by household
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
import joblib # For saving the model
import sys
import datetime

# --- Configuration ---
ML_DATA_FILE = 'ML_Ready_Data.csv' # ML data file
TEST_SET_SIZE = 0.25            # Proportion of households to hold out for testing (e.g., 0.25 = 25%)
N_ESTIMATORS = 150              # Number of trees in the forest (more trees -> potentially better but slower)
MAX_DEPTH = 15                  # Max depth of trees (None=unlimited, controls complexity, try e.g., 10, 15, 20)
MIN_SAMPLES_LEAF = 5            # Minimum samples required at a leaf node (controls complexity, try 3, 5, 10)
N_JOBS = -1                     # Use all available CPU cores for training (-1)
RANDOM_STATE = 42               # For reproducible results
N_HOUSEHOLDS_TO_PLOT = 4        # How many test households to visualize
SAVE_MODEL_FILENAME = 'random_forest_co_model.joblib' # Optional filename to save the trained model
# --- End Configuration ---

print("--- Starting Random Forest Training & Evaluation ---")

# --- Load Data ---
try:
    print(f"Loading ML-ready data from: {ML_DATA_FILE}")
    ml_data = pd.read_csv(ML_DATA_FILE)
    # Ensure Timestamp is datetime object if needed for analysis later (though not a direct feature here)
    if 'Timestamp' in ml_data.columns:
        ml_data['Timestamp'] = pd.to_datetime(ml_data['Timestamp'])
    print(f"  Successfully loaded {len(ml_data)} rows and {len(ml_data.columns)} columns.")
    print(f"  Columns: {ml_data.columns.tolist()}")
except FileNotFoundError:
    print(f"FATAL ERROR: File not found: {ML_DATA_FILE}")
    sys.exit(1)
except Exception as e:
    print(f"FATAL ERROR loading data: {e}")
    sys.exit(1)

# --- Define Features (X) and Target (y) ---
print("Defining features and target...")
try:
    target_column = 'CO_ppm'
    # Features are all columns except the target and potentially HouseholdID/Timestamp
    # Explicitly list features based on the previous preparation script's output
    feature_columns = [col for col in ml_data.columns if col not in [target_column, 'Timestamp', 'HouseholdID']]
```

Contd.

```python
# --- Define Features (X) and Target (y) ---
print("Defining features and target...")
try:
    target_column = 'CO_ppm'
    # Features are all columns except the target and potentially HouseholdID/Timestamp
    # Explicitly list features based on the previous preparation script's output
    feature_columns = [col for col in ml_data.columns if col not in [target_column, 'Timestamp', 'HouseholdID']]

    # Verify all expected feature columns exist
    if not all(col in ml_data.columns for col in feature_columns):
         missing = set(feature_columns) - set(ml_data.columns)
         print(f"WARNING: Some expected feature columns missing: {missing}. Excluding them.")
         feature_columns = [col for col in feature_columns if col in ml_data.columns]

    if not feature_columns:
        print("FATAL ERROR: No feature columns identified. Check column names in CSV.")
        sys.exit(1)

    X = ml_data[feature_columns]
    y = ml_data[target_column]
    groups = ml_data['HouseholdID'] # Needed for group-based splitting

    print(f"  Target: '{target_column}'")
    print(f"  Features ({len(feature_columns)}): {feature_columns}")

except KeyError as e:
    print(f"FATAL ERROR: Column mismatch. Missing column: {e}")
    print(f"Make sure '{ML_DATA_FILE}' contains the target and all feature columns.")
    sys.exit(1)
except Exception as e:
     print(f"FATAL ERROR preparing features/target: {e}")
     sys.exit(1)


# --- Split Data: Train/Test Holding Out Households ---
print(f"Splitting data into train/test sets ({1-TEST_SET_SIZE:.0%}/{TEST_SET_SIZE:.0%}), keeping households intact...")
# Using GroupShuffleSplit to ensure all data from a household is in EITHER train or test
gss = GroupShuffleSplit(n_splits=1, test_size=TEST_SET_SIZE, random_state=RANDOM_STATE)
train_idx, test_idx = next(gss.split(X, y, groups))

X_train, X_test = X.iloc[train_idx], X.iloc[test_idx]
y_train, y_test = y.iloc[train_idx], y.iloc[test_idx]
groups_train, groups_test = groups.iloc[train_idx], groups.iloc[test_idx] # Keep track of households in each set
timestamps_test = ml_data['Timestamp'].iloc[test_idx] # For plotting

train_households = groups_train.unique()
test_households = groups_test.unique()

print(f"  Training set: {len(X_train)} rows from {len(train_households)} households.")
print(f"  Test set:     {len(X_test)} rows from {len(test_households)} households.")
print(f"  Test Households: {np.sort(test_households)}")
```

Contd.

```
if len(X_test) == 0:
    print("FATAL ERROR: Test set is empty. Maybe TEST_SET_SIZE is too large or too few households?")
    sys.exit(1)

# --- Train Random Forest Model ---
print(f"\nTraining RandomForestRegressor model...")
print(f"  Parameters: n_estimators={N_ESTIMATORS}, max_depth={MAX_DEPTH}, min_samples_leaf={MIN_SAMPLES_LEAF}, n_jobs={N_JOBS}")

# Initialize the model
rf_model = RandomForestRegressor(
    n_estimators=N_ESTIMATORS,
    max_depth=MAX_DEPTH,
    min_samples_leaf=MIN_SAMPLES_LEAF,
    random_state=RANDOM_STATE,
    n_jobs=N_JOBS,
    oob_score=False, # Can set to True for out-of-bag estimate on training data
    max_features=1.0 # Consider all features at each split (default in newer sklearn)
)

# Train the model
try:
    start_time = datetime.datetime.now()
    rf_model.fit(X_train, y_train)
    end_time = datetime.datetime.now()
    print(f"  Training complete. Time taken: {end_time - start_time}")
except Exception as e:
    print(f"FATAL ERROR during model training: {e}")
    sys.exit(1)

# --- Make Predictions on Test Set ---
print("Making predictions on the test set...")
try:
    y_pred = rf_model.predict(X_test)
except Exception as e:
    print(f"FATAL ERROR during prediction: {e}")
    sys.exit(1)

# --- Evaluate the Model ---
print("\nEvaluating model performance on the test set...")
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"\n--- Test Set Evaluation Metrics ---")
print(f"RMSE: {rmse:.2f} ppm")
print(f"MAE:  {mae:.2f} ppm")
print(f"R²:   {r2:.3f}")
print(f"-----------------------------------")
print(f"(Compare RMSE/MAE to the range of CO_ppm values in your data)")
```

Contd.

```
# --- Visualize Results ---

# 1. Feature Importances
print("\nCalculating and plotting feature importances...")
try:
    importances = rf_model.feature_importances_
    feature_importance_df = pd.DataFrame({
        'Feature': feature_columns,
        'Importance': importances
    }).sort_values(by='Importance', ascending=False)

    plt.figure(figsize=(10, max(6, len(feature_columns) * 0.3))) # Adjust height based on num features
    plt.barh(feature_importance_df['Feature'], feature_importance_df['Importance'], color='skyblue')
    plt.xlabel('Importance')
    plt.ylabel('Feature')
    plt.title('Random Forest Feature Importances')
    plt.gca().invert_yaxis() # Display most important at the top
    plt.tight_layout()
    plt.show()
    print("  Top 5 Features:")
    print(feature_importance_df.head())
except Exception as e:
    print(f"  Warning: Could not generate feature importance plot: {e}")


# 2. Observed vs. Predicted Time Series (for Sample Households)
print(f"\nPlotting observed vs. predicted for {N_HOUSEHOLDS_TO_PLOT} sample test households...")
if len(test_households) > 0:
    households_to_plot = np.random.choice(test_households, size=min(N_HOUSEHOLDS_TO_PLOT, len(test_households)), replace=False)

    plt.figure(figsize=(15, N_HOUSEHOLDS_TO_PLOT * 5))
    plt.suptitle('Observed vs. Predicted CO (Random Forest) - Test Set Samples', fontsize=16, y=1.02)

    plot_count = 0
    for i, household_id in enumerate(households_to_plot):
        plot_count += 1
        ax = plt.subplot(N_HOUSEHOLDS_TO_PLOT, 1, plot_count)

        # Find indices corresponding to this household in the *original test set*
        idx_in_test = groups_test[groups_test == household_id].index

        # Use these indices to get the correct timestamps, observed, and predicted values
        household_timestamps = ml_data.loc[idx_in_test, 'Timestamp']
        household_y_test = y_test.loc[idx_in_test]
        # Need to align y_pred with the original index of X_test before filtering
        household_y_pred = pd.Series(y_pred, index=X_test.index).loc[idx_in_test]
```

```
        if household_y_test.empty: continue # Skip if somehow empty

        ax.plot(household_timestamps, household_y_test, label=f'Observed CO (Peak: {household_y_test.max():.1f})', color='blue', marker='.', linestyle='
        ax.plot(household_timestamps, household_y_pred, label=f'Predicted CO (Peak: {household_y_pred.max():.1f})', color='red', marker=None, linestyle=

        stove_type_plot = X_test.loc[idx_in_test, 'StoveType_TCS'].iloc[0] == 1 if 'StoveType_TCS' in X_test.columns else X_test.loc[idx_in_test, 'Stove
        stove_label = 'TCS' if stove_type_plot else 'ICS' if stove_type_plot != 'Unknown' else 'Unknown'

        ax.set_title(f'Test Household: {household_id} ({stove_label})')
        ax.set_ylabel('CO Concentration (ppm)')
        ax.set_xlabel('Time')
        ax.legend(loc='best')
        ax.grid(True, which='both', linestyle='--', linewidth=0.5)
        plt.xticks(rotation=20, ha='right')


    plt.tight_layout(rect=[0, 0.03, 1, 0.98]) # Adjust layout
    plt.show()
else:
    print("  No households available in the test set to plot.")


# --- Optional: Save the Trained Model ---
if SAVE_MODEL_FILENAME:
    print(f"\nSaving trained model to: {SAVE_MODEL_FILENAME}")
    try:
        joblib.dump(rf_model, SAVE_MODEL_FILENAME)
        print("  Model saved successfully.")
    except Exception as e:
        print(f"  Error saving model: {e}")
```

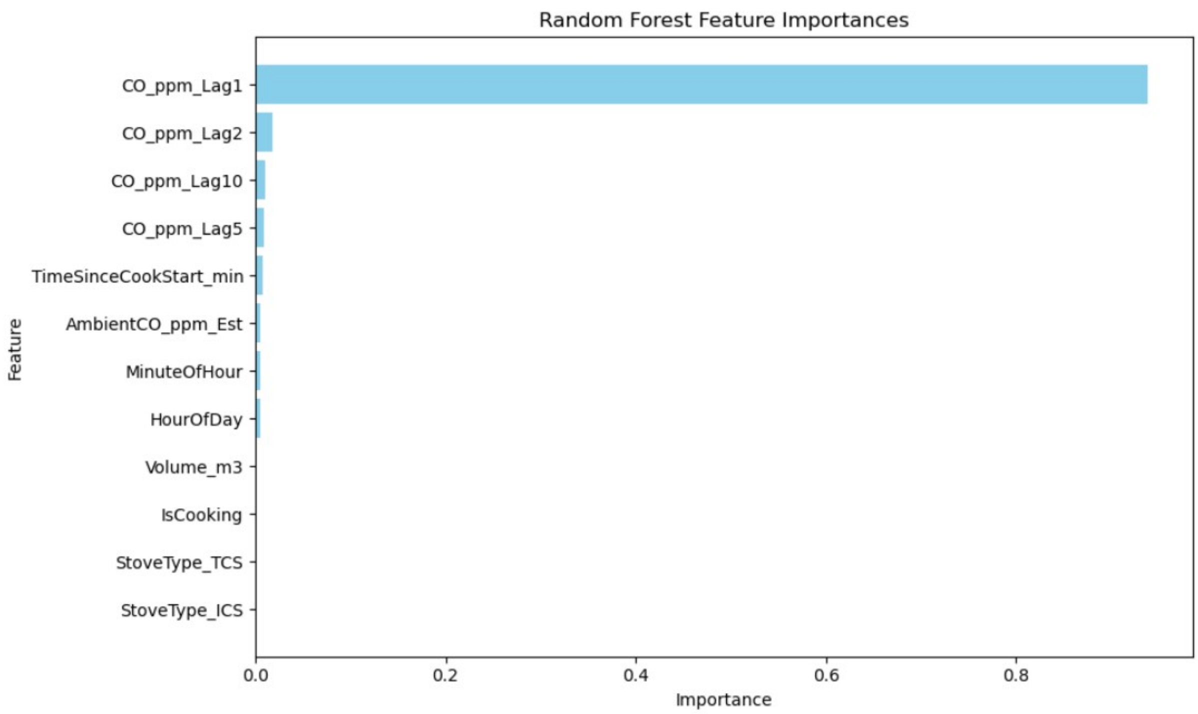
2. Output from the RF code:

```
--- Starting Random Forest Training & Evaluation ---
Loading ML-ready data from: ML_Ready_Data.csv
  Successfully loaded 2306 rows and 15 columns.
  Columns: ['HouseholdID', 'Timestamp', 'Volume_m3', 'AmbientCO_ppm_Est', 'HourOfDay', 'MinuteOfHour', 'IsCooking', 'TimeSinceCookStart_min', 'StoveT
ype_ICS', 'StoveType_TCS', 'CO_ppm_Lag1', 'CO_ppm_Lag2', 'CO_ppm_Lag5', 'CO_ppm_Lag10', 'CO_ppm']
Defining features and target...
  Target: 'CO_ppm'
  Features (12): ['Volume_m3', 'AmbientCO_ppm_Est', 'HourOfDay', 'MinuteOfHour', 'IsCooking', 'TimeSinceCookStart_min', 'StoveType_ICS', 'StoveType_T
CS', 'CO_ppm_Lag1', 'CO_ppm_Lag2', 'CO_ppm_Lag5', 'CO_ppm_Lag10']
Splitting data into train/test sets (75%/25%), keeping households intact...
  Training set: 1729 rows from 15 households.
  Test set:     577 rows from 5 households.
  Test Households: [ 5  6 13 21 23]

Training RandomForestRegressor model...
  Parameters: n_estimators=150, max_depth=15, min_samples_leaf=5, n_jobs=-1
  Training complete. Time taken: 0:00:00.503607
Making predictions on the test set...

Evaluating model performance on the test set...

--- Test Set Evaluation Metrics ---
RMSE: 16.69 ppm
MAE:  5.62 ppm
R²:   0.915
```




```
Top 5 Features:
                   Feature  Importance
8              CO_ppm_Lag1    0.939103
9              CO_ppm_Lag2    0.017958
11            CO_ppm_Lag10    0.009888
10             CO_ppm_Lag5    0.008820
5   TimeSinceCookStart_min    0.007064

Plotting observed vs. predicted for 4 sample test households...
```

3. Averages of PM2.5 before, during and after cooking for indoor condition

House	BEFORE	DURING	AFTER	SD BEF	SD DUR	SD AFT	Type
3	35.223	118.8875	36.23529	12.168	105.9678	23.68417	Improved
6	25.4567	601.7936	68.1667	4.906	315.083	18.793	Traditional
7	49.0138	105.9677	43.3375	21.68473	31.92823	7.34028	Improved
8	47.6	363.1	109.8	7.4	331.2	56.8	Traditional
10	34.917	115.19	54.22	23.877	35.15	39.263	Improved
13	49.7636	155.541	43.6793	27.624	143.1364	5.875	Improved
15	30.43	120.874	71.394	6.37	54.266	30.9362	Improved
16	21.5345	300.34	35.694	4.373	261.063	25.424	Traditional
17	23.47	82.072	45.27	6.28662	80.25	9.234	Improved
18	29.524	60.957	37.81875	8.0282	76.643	10.99	Improved
19	36.954	177.8757	68.21667	8.4276	233.8	65.75	Traditional
20	36.978	273.1721	127.1455	17.5098	190.513	202.3053	Traditional
21	25.2286	109.51	41.47429	5.5866	46.23	11.65	Improved
22	25.98	360.24	58.3	24.66	302.645	24.9988	Traditional
23	26.7025	101.24	54.9784	7.8324	78.2145	27.02144	Improved
24	35.7	72.81	33.624	13.97	56.603	17.14	Improved
25	40.962	157.64	33.922	10.977	207.3	16.183	Improved
1	47.376	188.77	50.455	22.0732	203.26	26.081	traditional
2	43.78	85.397	64.51	18.566	218.298	40.776	improved
3	66.055	365.573	50.0172	51.973	244.202	46.88	traditional
4	87.626	134.064	30.185	59.683	122.464	7.971	improved
5	53.656	530.31	91.05	131.0116	431.35	252.184	traditional
8	37.307	286.13	71.012	30.484	290.306	88.9715	traditional
9	26.7342	203.527	40.6547	24.814	218.69	95.477	traditional
11	26.01073	131.124	26.29	22.183	156.62	11.573	improved
12	19.291	197.152	52.205	6.282	222.613	48.86	traditional
14	39.815	87.48	46.277	9.5655	40.556	10.763	improved
18	79.456	378.848	45.627	27.711	362.3304	17.666	traditional
22	42.02	230.63	62.271	4.899	313.3699	19.70553	traditional
23	41.336	185.466	26.70217	34.472	282.886	1.145	traditional
24	59.815	320.981	102.67	11.811	306.183	250.6306	traditional
28	18.012	382.63	20.844	7.983	374.31	4.424	traditional

4. Averages of PM2.5 before, during and after cooking for personal condition

Houses	BEFORE	DURING	AFTER	SDBEF	SDDUR	SDAFT	Type
3	29.3371	104.2425	34.55	5.46	99.742	9.86	improved
4	33.1	136.24	43.432	4.98	76.51	26.64	traditional
6	23.374	332.92	55.84	3.2965	274.82	13.435	traditional
7	45.623	93.272	41.642	4.94	43.76	6.61	improved
8	50.72	263.45	42.82	5.961	312.79	5.82	traditional
10	32.12	133.16	28.78	21.28	41.04	4.23	improved
13	50.65	156.47	25.423	14.4525	134.81	6.6241	improved
14	59.475	104.126	32.01	22.724	65.851	5.782	improved
15	25.614	107.864	35.21	6.756	51.66	5.337	improved
17	29.6886	97.6772	54.8	8.496	127.113	22.94	improved
18	39.233	70.894	43.45	15.49	76.953	14.32	improved
21	37.81	130.11	43.744	45.021	59.705	10.092	improved
22	24.586	180.41	51.267	1.811	222.22	17.33	traditional
23	31.3	74.79	42.792	22.0011	53.0043	13.583	improved
24	39.597	91.661	29.252	14.14	99.696	11.973	improved
25	41.172	147.91	29.66	$1_{2.5}5$	230.25	16.94	improved
1	$4_{2.5}7$	154.57	34.02	23.18	118.74	14.79	traditional
3	48.7	231.12	29.09	28.5	230.4	20.51	traditional
4	66.98	109.47	19.34	47.16	131.39	5.21	improved
5	16.74	191.69	37.2	5.66	182.89	14.98	traditional
6	69.57	145.67	70.84	80.21	97.54	50.65	traditional
8	16.48	238.94	23.38	5.27	206.33	15.44	traditional
10	21.08	196.97	43.56	8.18	269.48	31.76	traditional
11	20.89	179.18	20.08	14.75	244.81	7.19	traditional
18	83.99	395.83	49.68	38.75	299.82	25.47	traditional
23	34.75	596.22	32.93	16.43	535.83	8.67	traditional
24	66.91	201.88	37.4	26.15	277.73	8.11	traditional
27	35.73	367.14	45.43	12.79	205.81	18.37	traditional
28	56.46	170.71	49.89	19.9	147.78	16.02	traditiona

5. Code to calculate AER and generate IAP trend during cooking

```python
import pylab as p
import numpy as np
import xlrd
import math
from matplotlib import pyplot as plt
from scipy.interpolate import spline

file_location="F:/CO data/salambu_house 5.xls"
workbook = xlrd.open_workbook(file_location)
sheet=workbook.sheet_by_index(0)
#calculating AER using carbon monoxide decay time
C=[sheet.cell_value(r,6) for r in range(121,157)]
Co=0.5*1.15

i=len(C)

AER= [((math.log((C[0]-Co)/(C[t]-Co))))/t for t in range(1,i)]

AER=np.array(AER)

AER1=AER.mean()

print("AER",AER1)
S=17.4*60
V=39.984

Ct1=[(((S/(AER1*V))*(1-math.exp(-AER1*t))))/8+Co*(math.exp(-AER1*t))for t in range (0,71)]
Ct2=[Co*(math.exp(-AER1*t))for t in range (72,106)]
Ct=Ct1+Ct2
Cob=[sheet.cell_value(r,6) for r in range(50,157)]
plt.plot(Cob,'r x', label="Observed")
plt.plot(Ct,'b',label='Modeled')
plt.xlabel('time(min)')
plt.ylabel('CO concentration (mg/m^3)')
plt.legend()
plt.show()
```

6. Code for box plot

```
In [1]: import pandas as pd
        import xlrd
        import numpy as np

        from matplotlib import pyplot as plt

In [2]: data=pd.read_excel("C:\\Users\\Bholu\\Desktop\\project final\\Mean and sd salambu.xlsx", sheet_name=3)

In [13]: data.boxplot('DURING', by='type', figsize=(12, 8),fontsize=12)
         j=1
         for i in ["improved","traditional"]:
             y=data.DURING[data.type==i].dropna()
             x=np.random.normal(j,0.04,size=len(y))
             j=j+1
             plt.plot(x,y,'o',alpha=0.9, lw=1.5)
             x1, x2 = 1, 2   # columns 'improved' and 'traditional' (first column: 0)
         y, h, col = data.DURING.max() + 10, 6, 'C0'
         plt.plot([x1, x1, x2, x2], [y, y+h, y+h, y], lw=1.5, c=col)
         plt.xlabel('Stove type',fontsize=12, fontweight='bold')
         plt.ylabel('PM 2.5 Concentration ($\mu$g/m^3)',fontsize=12,fontweight='bold')
         plt.text((x1+x2)*.5, y+h, "****", fontsize=15,ha='center', va='bottom', color='k')

Out[13]: Text(1.5,617.794,'****')
```

7. Code for normal distribution

```
In [2]: import numpy as np
        import xlrd
        import pandas as pd

In [3]: import scipy.stats as stats
        import matplotlib.pyplot as plt
        from scipy.interpolate import spline
        data=pd.read_excel("C:\\Users\\Bholu\\Desktop\\project final\\Mean and sd salambu.xlsx", sheet_name=9, header=1)
        beta=pd.read_excel("C:\\Users\\Bholu\\Desktop\\project final\\Mean and sd salambu.xlsx", sheet_name=10,header=1)
        Hmean=np.mean(data)
        Hstd=np.std(data)

In [4]: pdf=stats.norm.pdf(data,Hmean,Hstd)
        Gmean=np.mean(beta)
        Gstd=np.std(beta)
        pgf=stats.norm.pdf(beta, Gmean, Gstd)

In [5]: plt.figure(figsize=(10, 6))
        plt.plot(data, pdf, '-o',lw=2,alpha=0.9,label='concn ICS')
        plt.plot(beta, pgf, '-o',lw=2,alpha=0.9, label='concn TCS')


        plt.xlabel('PM concentration($\mu$g/m^3)')
        plt.ylabel('probability')



        plt.legend(loc='best',frameon=True)

        plt.title('Normal distribution of ICS and TCS')
        plt.text(140, .013, r'$\mu=109.67,\ \sigma=29.26$')
        plt.text(260, .0045, r'$\mu=314.5,\ \sigma=120.4$')
        plt.grid()
```

8. Codes for Sparklines

```
In [30]: import numpy as np
         import scipy.stats as stats
         import pandas as pd
         import matplotlib.pyplot as plt
         import matplotlib.patches as mpatches

         import xlrd
         data=pd.read_excel("C:\\Users\\Bholu\\Desktop\\project final\\Mean and sd salambu.xlsx", sheet_name=7, header=1)
         beta=pd.read_excel("C:\\Users\\Bholu\\Desktop\\project final\\Mean and sd salambu.xlsx", sheet_name=8, header=1)

In [32]: x = ["Before","During", "After"]
         y = [data.Before,data.During,data.After]
         m=["Before", "During", "After"]
         n=[beta.Before,beta.During,beta.After]
         plt.figure(figsize=(15, 9))

         plt.plot(x, y,'C1',lw=3,alpha=0.8)
         plt.plot(m,n,'C4',lw=3,alpha=0.6)
         plt.xlabel('Cooking Duration', fontsize=12, fontweight='bold')
         plt.ylabel('PM 2.5 Concentration ($\mu$g/m^3)', fontsize= 12, fontweight='bold')


         legend_dict = { 'ICS' : 'C1', 'TCS' : 'C4'}
         patchList = []
         for key in legend_dict:
                 data_key = mpatches.Patch(color=legend_dict[key], label=key)
                 patchList.append(data_key)

         plt.legend(handles=patchList)
         plt.show()
```

9. Code for T-test

```
In [1]: import pandas as pd
        import xlrd
        from statsmodels.stats.weightstats import ttest_ind
        from scipy import stats

In [2]: import xlrd

In [3]: data=pd.read_excel("C:\\Users\\Bholu\\Desktop\\project final\\Mean and sd salambu.xlsx", sheet_name=7, header=1)
        beta=pd.read_excel("C:\\Users\\Bholu\\Desktop\\project final\\Mean and sd salambu.xlsx", sheet_name=8, header=1)

In [4]: rad=data["During"]
        bad=beta["During"]

In [5]: stats.ttest_ind(rad,bad,equal_var= False)

Out[5]: Ttest_indResult(statistic=-4.484366845959325, pvalue=0.00034061016857543766)
```

10. Improved cook stove and occupant during cooking

11. Status of ventilation in some households

12. IAQ probe and AEROCET 831 mounted together, with their mouths adjusted at the same point

13. Acceptance email for the IOE Graduate Conference:

[IOEGC16] Editor Decision External Inbox ×

Kobid <conference-noreply@ioe.edu.np>
to me, Nawraj

Thu, Apr 3, 4:19 PM

Binamra Bhusal, Nawraj Bhattarai:

We are pleased to inform you that your manuscript titled " COMPARATIVE ANALYSIS OF EMISSION FROM TRADITIONAL COOK STOVES AND IMPROVED COOK STOVES AND EVALUATION OF VENTILATION USING BOX MODELS" submitted to 16th IOE Graduate Conference is **Accepted** for presentation in the Conference as well as inclusion in the Peer-Reviewed Proceedings. Please note that inclusion in hard copy proceedings is contingent upon your timely response to further edits, if any, during the publication process.

With Warm Regards,
IOEGC-16 Editorial Team

Reply | Reply all | Forward

14. Manuscript for IOE GC:

IOE Graduate Conference
[Placeholder for
Publication
Information]

COMPARATIVE ANALYSIS OF EMISSION FROM TRADITIONAL COOK STOVES AND IMPROVED COOK STOVES AND EVALUATION OF VENTILATION USING BOX MODELS

Binamra Bhusal [a], Nawraj Bhattarai [b]

[a] Department of Applied Sciences and Chemical Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal
[b] Department of Mechanical and Aerospace Engineering, Pulchowk Campus, Tribhuvan University, Nepal
[a] binamra4bhusal@gmail.com, [b] bnawraj@gmail.com

Abstract

In the rural household of Nepal, cooking with open fire has an adverse impact on the health of the occupants, especially women and children. The most prevalent stoves using wood as a primary source of fuel are Traditional Cook-Stoves (TCS) and Improved Cook-Stoves (ICS). TCS is an open fire cook-stove where cooking pot is placed on a cast iron cooking stand, whereas, ICS is a two pot mud type stove having a 1.2 m chimney. Rural households are not mechanically ventilated, and therefore ventilation must occur naturally. Indoor Air Quality measurement devices were used to measure pollutant concentration like $PM_{2.5}$ and Carbon Monoxide (CO). Decay curve method was employed to find the Air Exchange Rate (AER) that helps us assess the situation of ventilation in the occupant's kitchen. This value is compared with respect to American Society of Heating, Refrigeration and Air Conditioning Engineers (ASHRAE) standards. Furthermore, a box model was used to observe the trend of CO emission during the cooking period. Statistical analysis was conducted to verify the significant difference of pollutant emission from the two stoves. The study was carried out at Salambu, Kavre: a remote location in the hilly region of Nepal.

The real time indoor mean concentration and standard deviation of $PM_{2.5}$ during cooking period was 109.7 μg/m^3 (σ =29.3 μg/m^3) for ICS and 314.5 μg/m^3 (σ=120.4 μg/m^3) for TCS. Similarly the mean and standard deviation for personal measurement condition was 109.3 μg/m^3 (σ=26.0 μg/m^3) for ICS and 248.9 μg/m^3 (σ=121.1 μg/m^3) for TCS. The result showed a 65% ($p<0.0001$) and 56.1% ($p<0.001$) reduction in $PM_{2.5}$ for indoor and personal conditions while comparing ICS over TCS. The AER for the sampled households were observed to be in the range of 0.2 ACH to 11.6 ACH where less than 50% of the households meet the standard of 5 ACH set by ASHRAE and the relationship between AER and CO decay time was seen to be linear with R^2=0.21. Although ICS emissions were observed to be less than TCS, both did not meet the WHO standard permissible value for safe health. Furthermore, the backflow of pollutants in the kitchen was not accounted for the model generation although it had a significant impact on the pollutant concentration.

Keywords

ICS, TCS, AER, Ventilation, $PM_{2.5}$, CO

1. Introduction

1.1 Background

Biomass is the main source of fuel in most rural areas throughout the world and cooking in open flames has been one of the major reasons for poor health of a major portion of the world population [1]. About 3 billion people worldwide rely on biomass (wood, animal dung and crop waste) as their primary source of fuel and the count for premature death due to poor indoor air quality is around 4 million people [2]. In the context of Nepal around 76% of the population use biomass for cooking and heating purposes [3]. Some of the notable effects of poor indoor air quality due to biomass burning are, Chronic Obstructive Pulmonary Disease (COPD), Acute Lower Respiratory Infections (ALRI), low birth weight and still birth. Women and children are more prone to these effects as developing countries have a much traditional role of women tending over household works like cooking and taking care of children [4].

Natural ventilation is an effective passive technique to reduce some of the effects of indoor air pollution and ventilation models helps understand and analyze the indoor air quality [5]. Natural ventilation is cost effective and leads to better indoor air quality and creates a productive environment among occupants and is an efficient tool for sustainable, urban development [6]. One effective way to quantify ventilation is calculating the Air Exchange Rate as external air plays a vital role in the ventilation of the building [7]. Air Exchange Rate (AER) or Air Change per Hour (ACH) is defined as the air added of removed from a space divided by the volume of the space. Decay curve technique can be used to calculate the carbon monoxide decay time and ultimately the AER [1]. The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) standards help us compare and understand the condition of ventilation.

Box models are basic and effective tools in analyzing natural processes. Mass balance of carbon monoxide in a box of specified volume which is assumed to be completely mixed can be assessed to predict the concentration of the gas [8]. The predicted values from the model can be compared to the observed values to validate it.

Furthermore, adding the nuisance of indoor air pollution is the use of Traditional Cooking Stove (TCS), negligence in usage of Improved Cook Stoves (ICS) and poor ventilation. Among

15. Plagiarism Check:

Final Report_IOE_UPDATED.pdf

Tribhuvan University

Document Details

Submission ID
trn:oid:::3117:455459692

Submission Date
May 4, 2025, 2:14 PM GMT+5:45

Download Date
May 4, 2025, 2:47 PM GMT+5:45

File Name
Final Report_IOE_UPDATED.pdf

File Size
3.7 MB

73 Pages

14,989 Words

80,614 Characters

Contd.

iThenticate Page 2 of 78 - Integrity Overview Submission ID trn:oid:::3117:455459692

15% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Small Matches (less than 10 words)

Match Groups

76 Not Cited or Quoted 13%
Matches with neither in-text citation nor quotation marks

16 Missing Quotations 2%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

14% Internet sources
6% Publications
0% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

iThenticate Page 2 of 78 - Integrity Overview Submission ID trn:oid:::3117:455459692

Contd.

Match Groups

- **76** Not Cited or Quoted 13%
  Matches with neither in-text citation nor quotation marks
- **16** Missing Quotations 2%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

14% Internet sources
6% Publications
0% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	**conference.ioe.edu.np**	**5%**
2	Internet	**elibrary.tucl.edu.np**	**2%**
3	Internet	**www.aivc.org**	**1%**
4	Internet	**coek.info**	**<1%**
5	Internet	**cyberleninka.org**	**<1%**
6	Publication	**Clark, M.L.. "Indoor air pollution, cookstove quality, and housing characteristics i...**	**<1%**
7	Internet	**www.researchgate.net**	**<1%**
8	Publication	**Armendariz-Arnez, C.. "Indoor particle size distributions in homes with open fires ...**	**<1%**
9	Internet	**link.springer.com**	**<1%**
10	Internet	**www.science.gov**	**<1%**

Contd.

11	Publication	
Abdelaziz Testas. "Distributed Machine Learning with PySpark", Springer Science ...		<1%
12	Internet	
clintransmed.springeropen.com		<1%
13	Internet	
www.ssph-journal.org		<1%
14	Publication	
Othman Alrusaini, Hasan Beyari. "The Use of Machine Learning Algorithms in the ...		<1%
15	Internet	
docslib.org		<1%
16	Internet	
cdn.thomasnet.com		<1%
17	Internet	
pmc.ncbi.nlm.nih.gov		<1%
18	Internet	
repositories.nust.edu.pk		<1%
19	Internet	
www.coursehero.com		<1%
20	Internet	
www.scaler.com		<1%
21	Internet	
www.unicef.org		<1%
22	Internet	
www.tandfonline.com		<1%
23	Internet	
businessdocbox.com		<1%
24	Internet	
mountainscholar.org		<1%

Contd.

25	Internet	
web.realinfo.tv		<1%
26	Internet	
www.mdpi.com		<1%
27	Internet	
www.pure.ed.ac.uk		<1%
28	Internet	
pubmed.ncbi.nlm.nih.gov		<1%
29	Internet	
9pdf.net		<1%
30	Internet	
discol.umk.edu.my		<1%
31	Internet	
eprints.utm.my		<1%
32	Internet	
etd.aau.edu.et		<1%
33	Internet	
hdl.handle.net		<1%
34	Internet	
lup.lub.lu.se		<1%
35	Internet	
mail.bassettcreekwmo.org		<1%
36	Internet	
researchcommons.waikato.ac.nz		<1%
37	Internet	
www.ncasc.gov.np		<1%