# Project 5 Pollution Vision

*This manuscript (permalink) was automatically generated from Saran-Wang/dsproject@0ef85a2 on December 6, 2020.*

## Authors


- **Shiyuan Wang**
  · Saran-Wang
  Department of Civil and Environmental, University of Illinois

- **Weiqi Ni**
  · weiqini
  Department of Civil and Environmental, University of Illinois; Department of Environmental and Resources, Zhejiang University

- **Gemma Clark**
  · 441gclark
  Department of Civil and Environmental, University of Illinois

- **Xueao Li**
  · XueaoLi
  Department of Civil and Environmental, University of Illinois

# Abstract

# Introduction

# Literature Review

There are many studies using digital camera and advanced algorithm to estimate the concentrations of Particulate Matters. Hong et al. [1] developed a novel method of predicting the concentrations and diameters of outdoor ultrafine particles using street-level images and audio data in Montreal, Canada. Convolutional neural networks, multivariable linear regression and genralized additive models were used to make the predictions.

# Exploratory Data Anlysis

1. Variables Explanation
2. Data Cleaning

- Delete the useless columns in the dataset
- Delete the rows with equipment error during sampling

3. Visualization of the distributions of varibales Figure 1 shows that “Wind_Speed”, “Camera_Angle”, “Distance_to_Road” and “Elevation” are all in discrete distributions, while “Temp(C)” are in continuous distribution. “Pressure(kPa)” has four clusters. It should also be noted that the “Dead Time” almost shares the same distribution as “Total”.

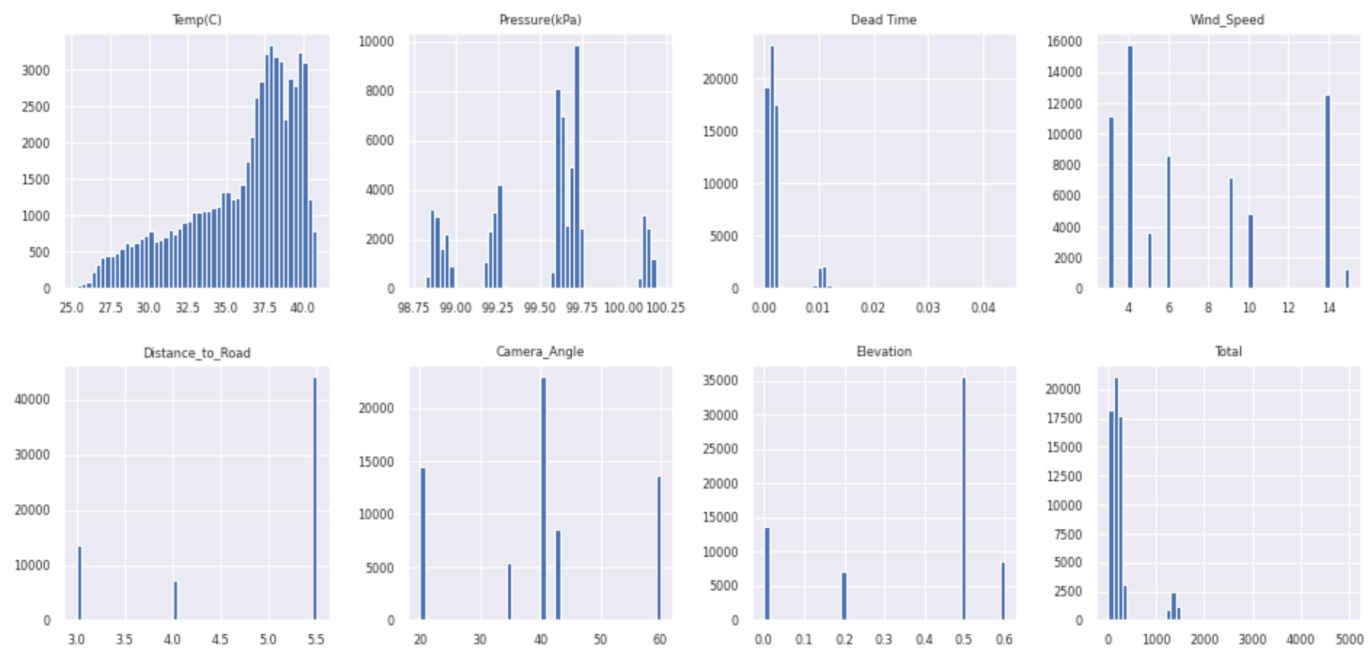


**Figure 1: Variables Distribution**

4. Correlations among variables From the correlation map 2 we could see that “Dead Time” are extremely correlated with “Total”, with a coefficient of 1, followed by “Camera_Angle”, “Pressure(kPa)” and “Distance_to_Road”, with coefficient of 0.52, 0.49, 0.44 respectively. Here you may be curious why “Dead Time” could be so closely related to “Total”, and there is one possible explanation: Actually, “Dead Time” is an instrument parameter, and if there are more PM concentrations in the air, the instrument need more time to process, and vice versa.

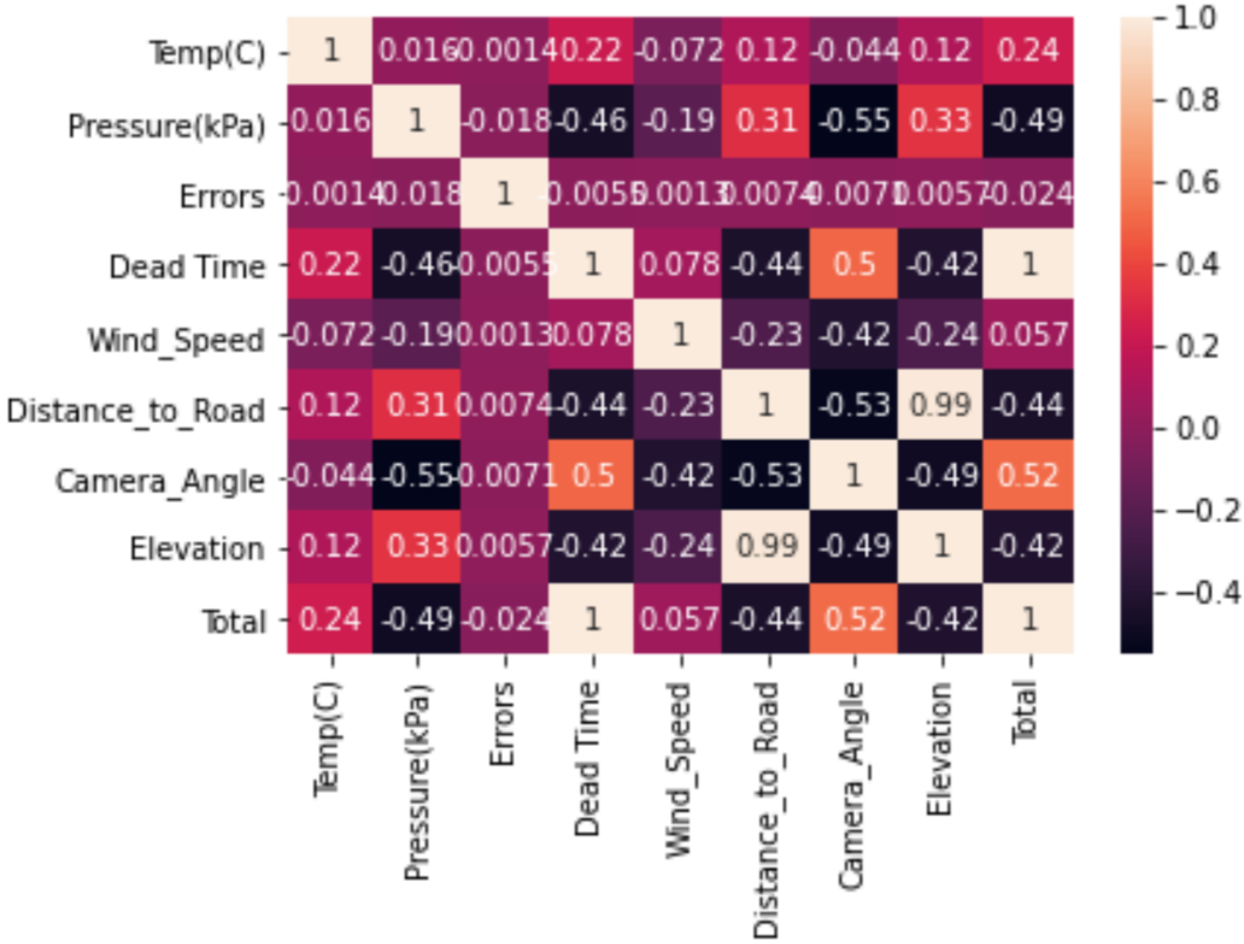


**Figure 2: Variables Correlations**

# Model

## Shiyuan’s Model

My model setup splits into two part, the first is image data extraction, the second is the selection of appropriate model to fit this dataset.

## Image Extraction

First I want to digitize images by extracting image features, there are mainly 6 features I want to extract: RGB, image luminance, image contrast, image entropy, transmission and amount of haze removed and number of cars on streets.

1. RGB

The RGB color model is one of the most straightforward parameters describing an image. Intuitively, in this case, we may expect more blueness and greenness if the PM concentrations are low since the color of tree and sky would be brighter when the air conditions are good. For each image, after deriving the RGB of each pixel, we take the average of them, and then divide each value by 255 to normalize it. The figure below 3 shows the distributions of RGB in this dataset. We can see that they are nearly normally distributed with mean 0.45, 0.55 and 0.35 respectively. For blueness, we could see a second peak at around 0.42.

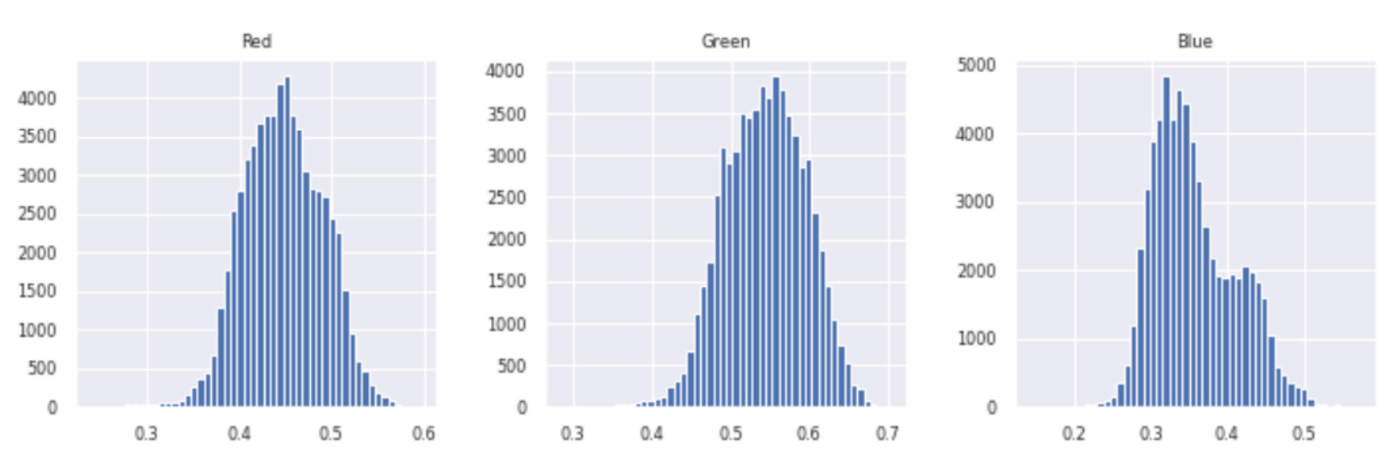


**Figure 3: RGB Distribution**

2. Luminance

Like RGB, luminance is also a very basic parameter describing an image, which could be an indicator of how bright the image will appear. The luminance of each image is calculated by taking the average of the luminance intensity of each pixel. From figure 4 we could also see that it’s also normally distributed with a mean of around 130.

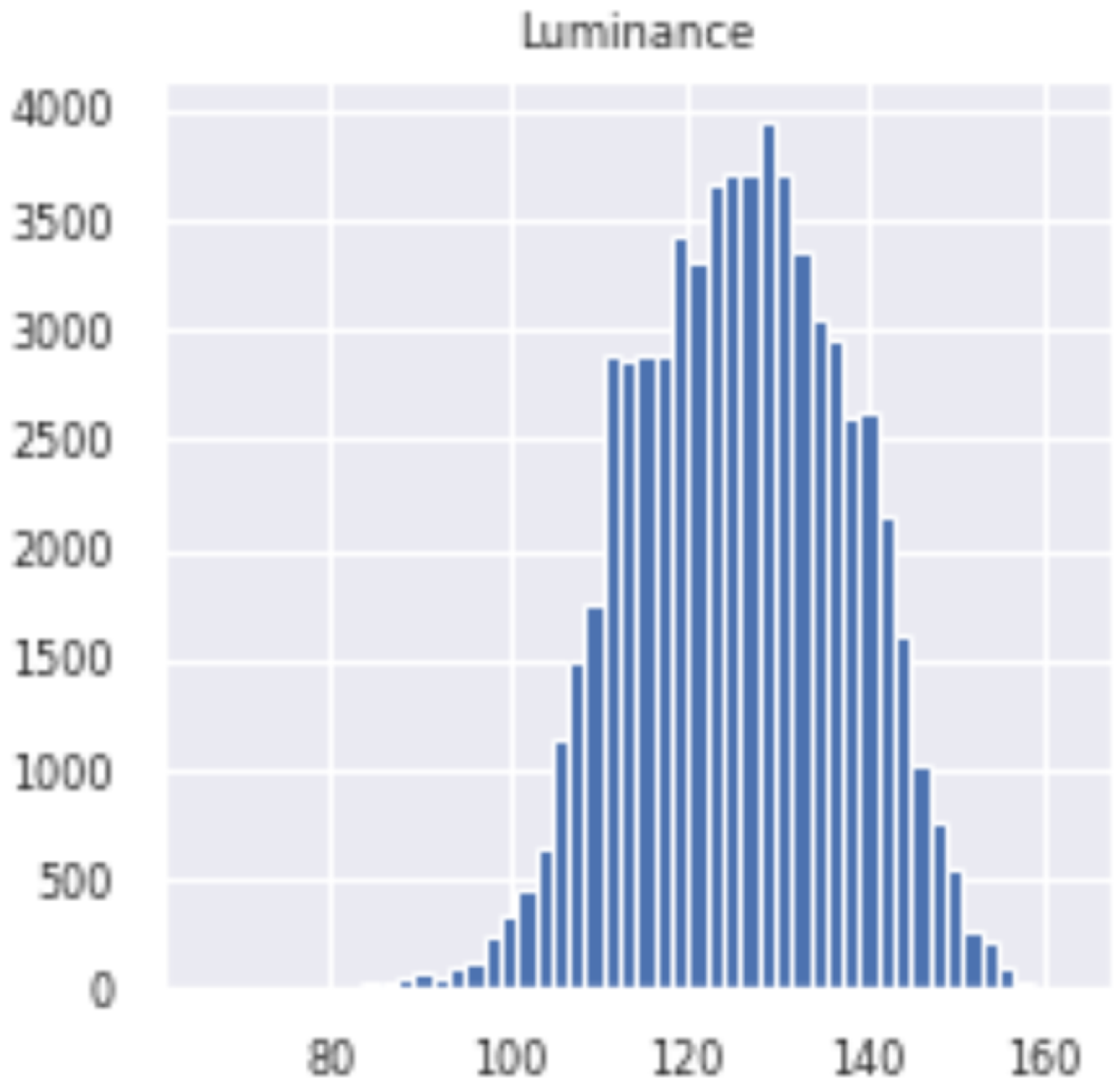


**Figure 4: Luminance Distribution**

3. Contrast
4. Entropy
5. Transmission and amount of haze removed
6. Number of cars on streets

## Model Selection

## Gemma’s Model

## Weiqi's Model

## Xueao’s Model

## Conclusion

## References


1. **Predicting outdoor ultrafine particle number concentrations, particle size, and noise using street-level images and audio data**
Kris Y. Hong, Pedro O. Pinheiro, Scott Weichenthal
*Environment International* (2020-11) https://doi.org/ghnh6n
DOI: 10.1016/j.envint.2020.106044 · PMID: 32805577