Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2019-393-AC1, 2020

Atmospheric Measurement Techniques Discussions

Interactive comment on "Evaluation and Calibration of a Low-cost Particle Sensor in Ambient Conditions Using Machine Learning Technologies" *by* Minxing Si et al.

Minxing Si et al.

kddu@ucalgary.ca

Received and published: 29 February 2020

1. **Comment 1 from Referee No. 3**: Lines 147-150, the authors stated that random search is more efficient, which is also a unique part of this study. Please explain why it is more efficient than manual or grid search in principle and if possible, give some quantitative information.

   Author's response: Manual search can be considered as an automated grid search. Grid search method and manual search method consider every combination of all the hyperparameters to build the learning models and each model needs to be evaluated to find out the one of the highest accuracy for training and prediction. For the XGBoost algorithm used in the manuscript Section 2.3.2, we tuned 7 hyperparameters. Each hyperparameter has 20 different parameters. The grid is a 7 by 20 table. The complete grid search requires 20^7 = 1,280 millions trials. In this study, we used 10 fold cross-validation, which means each trial will run 10 times. So the total runs will be 1,280 million $\times$ 10 = over 12 billion, which is computationally expensive.

For the random search, instead of computing the cases of all possible combinations, random combinations of hyperparameters are selected at each trial. Due to the random nature of sampling, the entire space of the grid could be reached (Zheng 2015).

The higher efficiency of random search can be explained by probability theory: Considering a sample space with a finite maximum, if we need to find a sample that is within the top 5% of all the samples, 60 random observations would give us 95% probability to find the sample. The value of 60 is calculated as follows:

As there are 5% eligible samples in the space, each random observation has 5% chance to find the eligible sample. On another hand, each random observation has $(1 - 5\%)$ chance not to find the eligible sample. If we take n random observations, the chance of not getting the eligible sample would be $(1 - 0.05)^n$, or the chance of getting the eligible sample would be $1 - (1 - 0.05)^n$. Let

$$1 - (1 - 0.05)^n > 95\%$$

And we can solve for n = 60 Therefore, the random search method would significantly save computation resources but still have a good chance to guess the close-to-optimal combination of hyperparameters.

Author's changes in manuscript We added a reference below in line 150 to explain the rationale of random search method. Zheng (2015) explained that random search with 60 samples will find a close-to-optimal combination with 95% of probability.

2. **Comment 2 from Referee No. 3** Figure 1, I would appreciate a geographical map showing where is the monitoring location.
   Author's response a geographical map is added as Figure 2.
   Author's changes in manuscript Added a new figure - Figure 2.

3. **Comment 3 from Referee No. 3** CRAZ also monitors NOx, NMHC, ozone, and wind data, which may also influence the PM concentration. Why these data were not included in the machine learning?
   Author's response The low-cost sensor evaluated in this study only measured temperature (T) and relative humidity (RH). The ultimate goal of low-cost sensor application is to provide same quality data as the reference method using available information provided by the low-cost sensor. Therefore, we only used the parameters that the sensor measured. Next phase of the study would be testing other types of low-cost sensors, which may provide other parameters than T and RH. In that case, we would include those parameters in machine learning.
   Author's changes in manuscript Not applicable.

4. **Comment 4 from Referee No. 3** Line 195: there is a typo: "SHAPR" should be "SHARP".
   Author's response Corrected
   Author's changes in manuscript Corrected to SHARP

5. **Comment 5 from Referee No. 3** SHARP was used as the reference method for PM monitoring. How often was SHARP calibrated to ensure its data quality
   Author's response The SHARP instrument is regulated by the provincial air monitoring directive. It was calibrated monthly.
   Author's changes in manuscript We added a clarification in Line 171 The instrument was calibrated monthly

6. **Comment 6 from Referee No. 3** Line 228: how the hyperparameters were determined?

Author's response The hyperparameters were determined by the XGBoost algorithm itself. Detailed explanation of each hyperparameter is provided in the XGBoost documentation https://xgboost.readthedocs.io/en/latest/parameter.html
Author's changes in manuscript We added the following reference in line 238 Detailed explanation of each hyperparameter is provided in the XGBoost documentation (XGBoost developers, 2019)

7. **Comment 7 from Referee No. 3** Figure 3: will you explain what the shape of erlenmeyer flask means in the plot?
Author's response The plots outside of the boxplots in Figure 3 is called violin plot. The violin plot is to describe the density of data. More details can be found in the following link: https://mode.com/blog/violin-plot-examples/
Author's changes in manuscript The following sentences were added in Line 264. The violin plot in Figure 3 describes the distribution of the PM2.5 values measured by the low-cost sensor and SHARP using density curve. The width of each curve represents the frequency of PM2.5 values at each concentration level.

8. **Comment 8 from Referee No. 3** One aspect of the uniqueness of this study is that its study covers different seasons. I would like to see a brief discussion how season influence the results of low-cost sensors.
Author's response We added a section to discuss the seasonal impact in Section 3.5.2

We assessed the seasonal impact on the low-cost sensor by comparing the mean of absolute daily average between the sensor values and the SHARP values in winter (December 2018 to February 2019) and spring (March 2019 to April 2019). A descriptive statistic is presented in Table 7.

We used a two-sample t test to assess if the mean of absolute differences for winter and spring were statistically significant. The p value of the t test was

0.754. Because $P = 0.754 > \alpha = 0.05$, we retained the null hypothesis. There was not sufficient evidence at the $\alpha = 0.05$ level to conclude that the means of absolute differences between the low-cost sensor and SHARP PM values were significantly different for winter season and spring season.

Author's changes in manuscript Added a section 3.5.2

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-393, 2019.

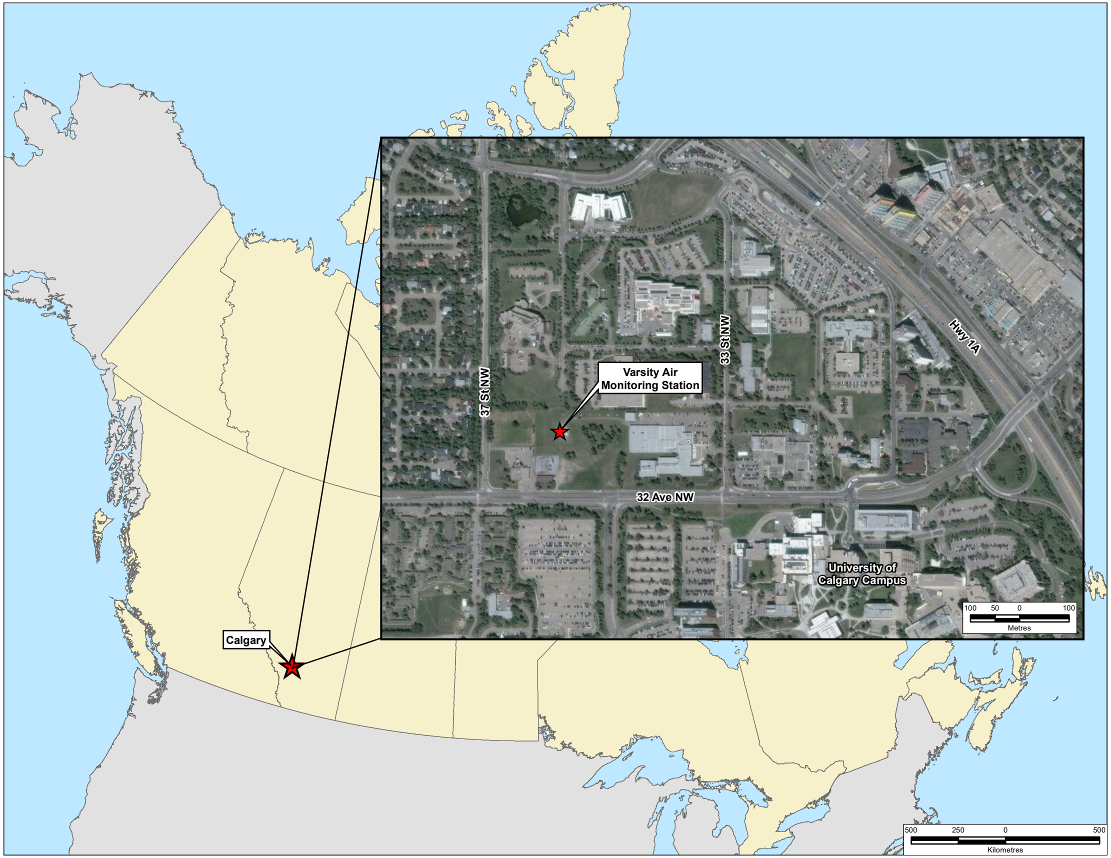

Fig. 1.

Interactive comment

Printer-friendly version

Discussion paper