

Predicting ambient traffic of a vehicle from road abrasion measurements using random forest

Christian Röger, Irada Ismayilova

Angaben zur Veröffentlichung / Publication details:

Röger, Christian, and Irada Ismayilova. 2020. "Predicting ambient traffic of a vehicle from road abrasion measurements using random forest." In *Proceedings of the 13th ACM SIGSPATIAL International Workshop on Computational Transportation Science - SIGSPATIAL '20, Seattle, WA, USA, November 2020*, edited by Anne Berres and Kuldeep Kurte, 4. New York, NY: ACM. <https://doi.org/10.1145/3423457.3429367>.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:

Deutsches Urheberrecht

Weitere Informationen finden Sie unter: / For more information see:

<https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/>



Predicting Ambient Traffic of a Vehicle from Road Abrasion Measurements using Random Forest

Christian Röger

Irada Ismayilova

christian.roeger@geo.uni-augsburg.de

irada.ismayilova@geo.uni-augsburg.de

University of Augsburg

Augsburg, Germany

ABSTRACT

The development and application of Intelligent Transportation Systems (ITSs) leads to a growing demand of traffic data. Floating Car Observers (FCOs) contribute by providing information about ambient traffic of vehicles while driving. We present an approach to implement an FCO that uses particulate matter sensors for obtaining road abrasion from cars driving ahead of a test vehicle. Using Random Forest (RF), we predict presence and absence of ambient traffic in the vicinity of test vehicle with particulate matter readings ($PM_{0.1}$, $PM_{2.5}$, PM_{10}) as predictor variables. Results show that RF reaches prediction accuracy ranging from 86 to 99 percent for different train/test split options when analysing individual trajectories as well as 88 to 91 percent accuracy when analysing all trajectories combined. We face limitations mainly when merging single trajectories, due to different initial ambient particulate matter values. We conclude that presence and absence of ambient traffic are predictable using Random Forest with road abrasion values as predictor variables. Further, rainfall events (that may cause wash-off effects on roads) do not significantly change the accuracy of our classification. Optimisation of the model and the need of testing more diverse weather and road conditions remain open tasks for future research.

CCS CONCEPTS

• **Hardware** → **Sensors and actuators; Sensor applications and deployments; Sensor devices and platforms;** • **Computing methodologies** → **Machine learning.**

KEYWORDS

Floating Car Observer (FCO), Sensors, Road Abrasion, Machine Learning, Random Forest (RF)

ACM Reference Format:

Christian Röger and Irada Ismayilova. 2020. Predicting Ambient Traffic of a Vehicle from Road Abrasion Measurements using Random Forest. In *13th International Workshop on Computational Transportation Science (IWCTS'20)*, November 3, 2020, Seattle, WA, USA. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3423457.3429367>

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in:

IWCTS'20, November 3, 2020, Seattle, WA, USA

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8166-6/20/11.

<https://doi.org/10.1145/3423457.3429367>

1 INTRODUCTION

Intelligent Transportation Systems (ITSs) aim towards making road systems more efficient, environmentally friendly and safer. Due to a rapid development in the research field of ITSs, there is a growing need of traffic data [6] [34]. Floating Car Observers (FCOs) contribute by providing information about the surrounding traffic of an individual vehicle. This could either be the number of vehicles driving ahead, the number of vehicles surrounding a car or a Boolean representation (absence/presence of traffic). Many attempts have been made to accomplish FCOs, e.g. using proximity sensors like laser scanners [3] [31], scanning for digital signals of nearby devices [12] [29] or filming ambient traffic with a video camera [30] [32]. However, many approaches come with limitations. Video cameras are able to capture ambient traffic accurately, but may be restricted due to privacy guidelines in some regions [25]. Scanning for digital signals in the surroundings of a car (e.g. Bluetooth or Wifi) faces the problem of possible re-detections of similar devices [13]. Proximity sensors like laser scanners generally have high acquisition costs. Consequently, we see potential in approximating ambient traffic of a vehicle using road abrasion measurements as a cost effective and accurate alternative.

We aim to develop an FCO using road abrasion measurements obtained from particulate matter sensors. There is research on mobile particulate matter measurements with sensors mounted on vehicles [8] [33]. However, there is no research on particulate matter based FCOs existing yet. In this paper we aim to prove that ambient traffic (here: presence and absence of other vehicles in the environment of cars) can be predicted using Random Forest (hereafter RF) with road abrasion measurements as predictor variables. Further, we investigate which influences different weather conditions (e.g. rainfall events) and road types (e.g. urban roads, highways) have on the model results.

In this case study, we introduce a multi sensor array for measuring road abrasion. In addition, we place a video camera in front of the windscreen of our test vehicle in order to be able to train our model and validate the results. Using the sensor array and the camera, we create several datasets under different road and weather conditions. Accuracy of predictions, sensitivity, specificity, Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) values will be analysed in order to assess the performance of our approach.

2 MEASURING AND CLASSIFYING ROAD ABRASION

Can road abrasion be measured using low cost particulate matter sensors?

Vehicles do not only emit pollutants in the form of tailpipe emissions. There are also particles that emerge from tire and brake wear and tear. Those particles spread on the road surfaces until they get whirled up by other cars that pass by. This process is called road abrasion and resuspension [9]. [5] emphasise the influence of road abrasion on total particle matter emission of cars. In their experiment, they calculate ratios for particle sources of driving vehicles in an urban and a highway environment. For urban roads, they calculate a ratio of 21 percent brake wear, 38 percent resuspended road dust and 41 percent tailpipe exhaust emissions. For highways, the result is 3 percent brake wear, 56 percent resuspended particles and 41 percent tailpipe exhaust emissions [5].

Depending on the source, particles have different diameters. $PM_{0.1}$, $PM_{2.5}$ and PM_{10} are frequently used subdivisions of particles. The term $PM_{0.1}$ describes very fine particles with a diameter of one μm and smaller. Particles with a diameter of 2.5 μm and less are called $PM_{2.5}$. Bigger particles with a diameter of ten μm and smaller are called PM_{10} [2].

Studies show that road abrasion and resuspension is measurable using particulate matter sensors. Kupiainen and Pirjola (2011) measure road abrasion with a mobile laboratory vehicle. A car is equipped with pipes mounted next to the tires of the vehicle. The pipes lead to particle counters that assess the amount of particulate matter emitted by road abrasion [16]. Zum Hagen et al. (2019) measure particles emitted by brake wear with a cone-shaped sampler that leads air from the break region of a car to its trunk. There, particles get measured and counted [35].

Influences of rainy weather conditions are a possible limitation for our research. In fact, rain may wash away the particles from the road surfaces making them not measurable. In their study, [21] investigate the influence of rainfall events on particles of different sizes and materials. The elements of Manganese (Mn), Iron (Fe), Aluminium (Al), Nickel (Ni), Chromium (Cr), Zinc (Zn), Lead (Pb), Copper (Cu), Antimony (Sb) and Cadmium (Cd) are investigated. Findings indicate that rain only influences particles with a diameter of 75 μm and bigger. Smaller particles are excluded from wash-off effects. Consequently, we expect that rainfall events will not have a significant influence on measurement results of particulate matter sensors when it rains for every particle material.

Not only particles getting washed away during rain events is a possible limitation to our research, but also water particles getting interpreted as fine dust by particulate matter sensors. As [27] proposes, particulate matter readings PM^{wet} can be corrected to dry particulate matter values PM^{dry} by dividing them by a growth function $gf(rh)$:

$$PM^{dry} = \frac{PM^{wet}}{gf(rh)} \quad (1)$$

For selecting growth functions $gf(rh)$, there are several possibilities. [24] propose using the following function:

$$gf_{Soneja} = 1 + \frac{\alpha * rh^2}{1 - rh} \quad (2)$$

with rh as relative humidity and an α value of 0.25 as proposed in the literature [24].

For classifying the measurement results we use the RF method. The choice of RF algorithm is justified by its robustness against noise, missing values and correlation between variables [10]. Moreover, it provides high prediction accuracy through cross validation [4]. As we expect some noise emerging in our fine dust readings, RF seems applicable.

Proposed by Breiman in 2001, RF is an ensemble learning method used for both classification and regression tasks. It is an enhanced version of the widely used Classification and Regression Tree algorithm (CART) [26]. By answering yes/no questions, trees in the forest are grown. Unlike to the boosting and bagging classification trees, each tree in the forest is formed using the best of the random input variables at each node split as well as based on the bootstrap sampling (where one third of the original data is left out and used later for testing) [4] [22]. Forest can grow until the discrete classification is achieved. Then, the final classification decision is made using majority of a classification outcome (vote) of each tree in the forest. RF allows to obtain an average error of prediction of out of bag samples (OOB error). OOB error can also be considered as an independent accuracy assessment parameter since samples used to calculate the error do not appear in bootstrap samples [4].

As a part of general procedure, data is split into train and test datasets. The model is trained using the train subset and validated using the test subset [4]. RF is very user friendly in execution and there are only two parameters that need to be specified [19] and can be tuned so that higher prediction accuracy is achieved. These include the number of classification trees in the forest (n_{tree}) and number of variables used to split internal node (m_{try}) [28]. Variable importance measure provided by a RF model is another estimate which helps in eliminating noisy or less important variables. It is calculated through permutation feature importance function and shows a decrease of a model score when a single feature value is randomly shuffled [4].

In classification tasks, performance of the RF model is possible to verify through a Confusion Matrix (CM) [14]. CM is produced with both OOB error and independent error assessments and shows predicted values versus actual values. Another RF model performance estimate is the Area Under the Receiver Operating Characteristics (AUROC). It explains how good the model is capable to distinguish between predicted classes. ROC curve plots true positive rate against false positive rate. The higher the area under the curve (AUC), the better the model determines different classes [11].

RF is an accurate classifier and robust against noise. Further, it is not sensitive to a small sample size while it is capable to handle big amount of data [17]. There are many examples where RF outperforms other widely used statistical modelling methods. For instance, [1] compared Logistic Regression, Naive Bayes, Support Vector Machine, Random Forest and Neural Networks for detecting anomalies in traffic and concluded that RF had the best prediction performance. In another case study ([7]), which focused on traffic accidents, RF again reached higher prediction accuracy than Artificial Neural Networks and Support Vector Machine.

Therefore, in this paper we present a use case of RF for predicting presence and absence of ambient traffic nearby the test vehicle with the help of particulate matter recordings.

Table 1: Overview of Test Trajectories

| Trajectory | Length | Record Points | Weather Conditions | Road Type | Ø Temperature | Ø Humidity |
|------------|---------|---------------|--------------------|----------------------------|---------------|------------|
| a | 5.37 km | 453 | Moderate Rain | Urban | 15.7 °C | 90.0 % |
| b | 7.13 km | 545 | Cloudy | Urban & Federal Highway | 18.3 °C | 53.3 % |
| c | 5.69 km | 271 | Cloudy | Urban & Federal Highway | 18.0 °C | 52.6 % |
| d | 5.10 km | 453 | Sunny | Suburban | 12.2 °C | 71.1 % |
| e | 9.69 km | 383 | Cloudy | Suburban & Federal Highway | 15.6 °C | 84.9 % |
| f | 9.05 km | 484 | Light Rain | Urban & Federal Highway | 15.9 °C | 82.2 % |
| g | 4.13 km | 310 | Cloudy | Urban | 15.5 °C | 84.4 % |

3 A FLOATING CAR OBSERVER (FCO) USING PARTICULATE MATTER SENSORS AND RANDOM FOREST

We introduce a multi sensor array for gathering road abrasion data. Further, we describe a Random Forest (RF) based model to classify recorded data points into two classes; *ambient traffic* and *no ambient traffic*.

3.1 Data Acquisition

We produce a dataset with particulate matter readings obtained from low cost sensors and information about surrounding traffic of our test vehicle using a video camera. A *PMS3003* particulate matter sensor is placed in front of the ventilation grille of the test vehicle with the aim to sense road abrasion and resuspension emitted by cars driving in front of the car. Further we obtain temperature and humidity readings from a *DHT22* sensor placed next to the particulate matter sensor for eliminating the influence of ambient humidity from fine dust values. Figure 1 shows the particulate matter sensor and the temperature/humidity sensor mounted on our test vehicle.



Figure 1: Particulate Matter and Temperature/Humidity sensors mounted on our test car

In addition, we gather spatio-temporal data. A *NEO-6M* GPS sensor is used to acquire datetime and location data. The sensors

are connected to an *Arduino Nano* microcontroller that logs every record to a micro SD card. We collect a total number of 12 parameters as shown in Table 2.

Table 2: Data obtained for the Case Study

| Parameter | Unit |
|------------------|----------------------|
| Datetime | Datetime |
| Location | Latitude & Longitude |
| PM_{01}^{wet} | $\mu g/m^3$ |
| $PM_{2.5}^{wet}$ | $\mu g/m^3$ |
| PM_{10}^{wet} | $\mu g/m^3$ |
| PM_{01}^{dry} | $\mu g/m^3$ |
| $PM_{2.5}^{dry}$ | $\mu g/m^3$ |
| PM_{10}^{dry} | $\mu g/m^3$ |
| Temperature | Degrees Celsius |
| Humidity | % |
| Ambient Traffic | Boolean (0/1) |

Table 2 shows the parameters obtained from our sensors. *Date-time* and *Location* values are used for joining the particulate matter values to the *ambient traffic* values gathered from the recorded video stream. We use a Boolean representation with the value 0 for *no ambient traffic* and the value 1 for *one or more cars* surrounding our test vehicle. PM_{01}^{wet} , $PM_{2.5}^{wet}$, PM_{10}^{wet} values grouped into three particle sizes (PM_{01}^{wet} , $PM_{2.5}^{wet}$, PM_{10}^{wet}) are directly obtained from the particulate matter sensor. Since humidity influences particulate matter readings, we apply a humidity compensation to our sensor values. Using humidity readings and the humidity compensation formulas proposed by [27] and [24], we are able to compute particulate matter values under dry condition (PM_{01}^{dry} , $PM_{2.5}^{dry}$, PM_{10}^{dry}). Using our sensor setup, we perform seven test drives to generate data for our case study. The test trajectories and their conditions (trip length, number of record points, road types, weather conditions, average temperature and average humidity) are presented in Table 1. All trajectories vary between 4.13 kilometres (trajectory g) and 9.69 kilometres (trajectory e) trip length with a total record point count between 271 (trajectory c) and 545 (trajectory b) points. We cover several road types such as urban roads, suburban roads and federal highways. By driving under different weather conditions we want to determine if there are limitations to the usage of

our sensor array for specific weather events like rainfall or dryness. Our dataset covers rainy as well as sunny conditions.

3.2 Model Design

We conduct a Case Study assessing the use of RF to predict ambient traffic of a car using road abrasion measurements. Figure 2 provides a flowchart with input data, processing steps and model outputs.

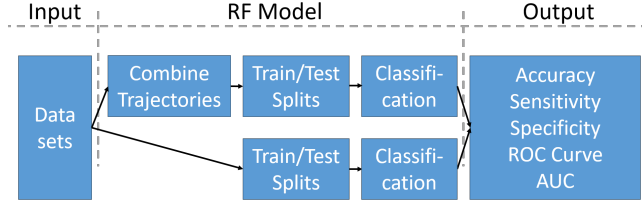


Figure 2: Flowchart with input data, processing steps and output parameters of the model

Figure 2 shows input data, processing steps and output parameters of our model. We conduct the study both with the single trajectories of our dataset and with the entire dataset. We use only PM_{01}^{dry} , $PM_{2.5}^{dry}$ and PM_{10}^{dry} as predictors to train our model. In order to join the single trajectories, we perform a field calibration of our measurements to bring the measurement values of PM_{01}^{dry} , $PM_{2.5}^{dry}$ and PM_{10}^{dry} to similar levels. This is a necessary step, as each trajectory comes with individual initial ambient particulate matter levels within the air, triggered by the amount of pollen, water particles and wind that lets air circulate. This step is not performed for single trajectories since we only look at relative changes of particulate matter values. Next, we split both the entire dataset and the single trajectories into train and test subsets. There is no particular recommendation in the literature which split option to choose. Therefore, we use 50/50, 60/40 and 70/30 splits. We train the model using the train subsets and test the model performance with the test subsets. We use the default parameter values of the R randomForest function for the model, with n_{tree} (500) and $mtry$ (\sqrt{p} , where p is the number of predictor variables) [18] and then tune those parameters if required to achieve better classification results.

Outputs of the model include the accuracy of predictions, the sensitivity, the specificity, a ROC curve and the AUC value. Output parameters are then used to assess the performance of the predictions and validate model results.

3.3 Results of the Classification

First, we analyse the single trajectories individually in order to examine the influence of different weather and road conditions on the model performance. Output parameters of the model are shown in Table 3.

Table 3: Accuracy values for different train/test splits of the test trajectories ID_{Traj}

| ID_{Traj} | Split | Accuracy | Sensitivity | Specificity | AUC |
|-------------|-------|----------|-------------|-------------|--------|
| a | 50/50 | 0.9251 | 0.8846 | 0.9593 | 0.97 |
| a | 60/40 | 0.9396 | 0.9398 | 0.9394 | 0.9634 |
| a | 70/30 | 0.9485 | 0.9508 | 0.9467 | 0.9237 |
| b | 50/50 | 0.9194 | 0.9247 | 0.9167 | 0.9686 |
| b | 60/40 | 0.9404 | 0.9420 | 0.9396 | 0.9772 |
| b | 70/30 | 0.9329 | 0.9123 | 0.9439 | 0.9583 |
| c | 50/50 | 0.9338 | 0.6897 | 1.0000 | 0.9826 |
| c | 60/40 | 0.9725 | 0.8571 | 1.0000 | 0.9775 |
| c | 70/30 | 0.9878 | 0.9231 | 1.0000 | 0.9695 |
| d | 50/50 | 0.8987 | 0.9024 | 0.8942 | 0.9651 |
| d | 60/40 | 0.9121 | 0.8932 | 0.9367 | 0.9723 |
| d | 70/30 | 0.9338 | 0.9600 | 0.9016 | 0.9871 |
| e | 50/50 | 0.9635 | 0.9697 | 0.9570 | 0.9906 |
| e | 60/40 | 0.961 | 0.9747 | 0.9467 | 0.9799 |
| e | 70/30 | 0.9565 | 0.9655 | 0.9474 | 0.9854 |
| f | 50/50 | 0.9558 | 0.9667 | 0.9457 | 0.983 |
| f | 60/40 | 0.97 | 0.9794 | 0.9612 | 0.9883 |
| f | 70/30 | 0.9733 | 1.0000 | 0.9518 | 0.9883 |
| g | 50/50 | 0.8581 | 0.9296 | 0.7976 | 0.9323 |
| g | 60/40 | 0.8871 | 0.9180 | 0.8571 | 0.9341 |
| g | 70/30 | 0.871 | 0.8776 | 0.8636 | 0.974 |

Table 3 shows accuracy, sensitivity, specificity and AUC values for different splits of the trajectories a to g . Prediction accuracy values range from 86 percent (trajectory g , 50/50 split) to 99 percent (Trajectory c , 70/30 split) showing that the overall prediction performance is good. Sensitivity ranges from 69 percent (Trajectory c , 50/50 split) to 100 percent (trajectory f , 70/30 split). Trajectory c with 50/50 split shows a bad true positive rate, meaning that points with actual ambient traffic could not be predicted well out of particulate matter values. The rest of sensitivity values show good performance in detecting true positives. Specificity ranges from 86 percent to 100 percent meaning that true negatives (in this case *no ambient traffic*) are detected sufficiently. ROC curves imply high AUC values which range from 93 percent (Trajectory g , 50/50 split) to 99 percent (Trajectory e , 50/50 split). Overall, 70/30 splits perform best with the used datasets. Nevertheless, 60/40 splits perform only few percent worse. 50/50 splits also perform well with the exception of sensitivity in Trajectory c .

We cannot observe any differences of performance values when taking into account weather and road conditions of our trajectories. We assume that the humidity correction we conducted when producing the datasets eliminates the influence of rain. There is also no influence of different road types (suburban roads, urban streets, highways) visible in the results.

Further, we train and test the model with the entire dataset. Due to different ambient particulate matter values while recording single trajectories, the overall levels of the values differ. This can be caused by pollen, water particles or different wind conditions. Thus, we calibrate particulate matter values from each trajectory to bring the single trajectory readings to the same level. A correction factor

is calculated by taking into account mean values of the particulate matter readings of each trajectory. After calibration, we merge the trajectories to a single dataset. Using different train/test splits (50/50, 60/40, 70/30) and standard settings of n_{tree} and $mtry$, we run the RF model again. Model outputs are shown in Table 4.

Table 4: Accuracy values for different train/test splits of the entire dataset

| Split Option | Accuracy | Sensitivity | Specificity | AUC |
|--------------|----------|-------------|-------------|--------|
| 50/50 | 0.8823 | 0.8494 | 0.9054 | 0.9588 |
| 60/40 | 0.8777 | 0.8815 | 0.8745 | 0.9507 |
| 70/30 | 0.913 | 0.8690 | 0.9465 | 0.961 |

Table 4 shows accuracy of classification, sensitivity, specificity and AUC values for all three train/test options. Accuracy ranges from 88 percent (50/50 split) to 91 percent (70/30 split). Sensitivity values are quite similar for each split options going from 85 percent (50/50 split) to 88 percent (60/40 split). Specificity ranges from 88 percent (60/40 split) to 95 percent (70/30 split). ROC curves are shown in Figure 3.

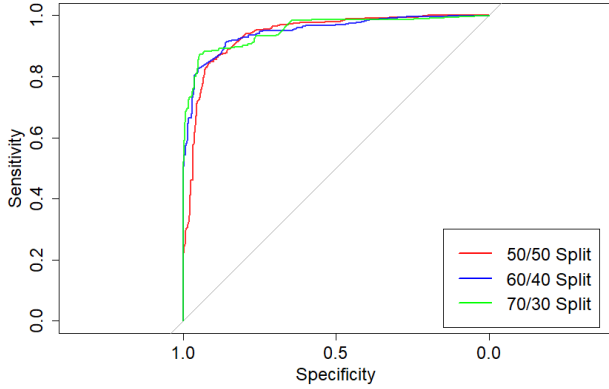


Figure 3: Receiver Operating Characteristic (ROC) curves for sensitivity and specificity of the RF model (entire dataset)

Figure 3 shows ROC curves for each split options of the entire dataset. Curves run very near the upper left corner, which indicates a valid model. Especially the 70/30 option shows good performance with an AUC value of 96 percent. As for the single trajectories, the 70/30 split option provides us with the best results. Other tested split options do not significantly perform worse.

4 DISCUSSION AND LIMITATIONS

Our study shows promising results. Prediction accuracy, sensitivity, specificity and AUC outputs are on a high level, meaning the model performs good and is valid. Nevertheless, there are some limitations in our research.

For generating data, we introduce a multi sensor array. It includes mainly low cost sensors. While temperature and humidity readings are known to be quite accurate for DHT22 sensors (humidity $\pm 2\%$

RH, temperature ± 0.5 degrees Celsius according to the datasheet [20]), PMS3003 particulate matter sensors tend to give more imprecise output values. According to the PMS3003 datasheet [23], $PM_{2.5}$ readings have a maximum consistency error of $\pm 10 \mu g/m^3$. Response times range from <1 second (single response time) to ≤ 10 seconds (total response time). Since we do not aim to measure absolute particulate matter values, their inaccuracies are acceptable for our research. Still, we see a limitation in the specifications of PMS3003 sensors concerning the response times.

The placement of the sensors in front of the ventilation grille seems reasonable to us. Still it is known that changes in air flow through the sensor also create inaccuracies. While the vehicle is driving, there is an unstable flow of air that comes through the sensor during measurement. In order to stabilise the air flow, other studies propose the use of a critical nozzle [15].

For dealing with influences of humidity in our particulate matter values, we apply a humidity correction. Since trajectories with higher average humidity do not perform significantly worse than trajectories with low average humidity, we conclude that the correction performs well. Figure 4 shows humidity correction exemplary on $PM_{2.5}^{wet}$ and $PM_{2.5}^{dry}$ values of trajectory f.

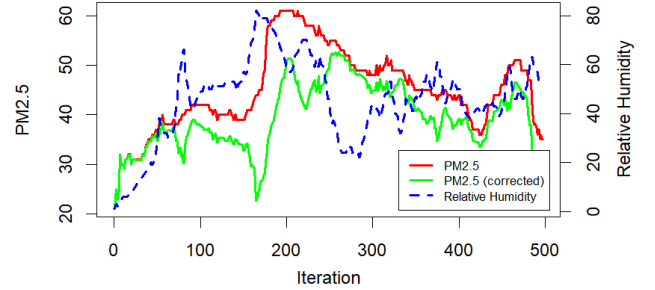


Figure 4: Relative humidity, particulate matter $PM_{2.5}^{wet}$ and humidity corrected particulate matter $PM_{2.5}^{dry}$ for trajectory f

Figure 4 depicts relative humidity (blue dotted line), $PM_{2.5}^{wet}$ before humidity correction (red line) and $PM_{2.5}^{dry}$ after humidity correction (green line) for trajectory f. As ambient humidity rises, the correction formula automatically adjusts the particulate matter values in order to prevent them from influence of water particles. Still high humidity is a factor that affects our research. Some particles get washed away from road surfaces when rainfall events occur. As [21] state, not every type of particle is influenced by rain wash-off events. Though, there are some particle types (e.g. rubber, emerging from tire wear) not tested within the study of [21]. Consequently, a higher number of test drives under high and low humidity conditions might be beneficial to carry out.

We run our model based on seven test trajectories conducted under different road and weather conditions. Although we cover different conditions, there are more heterogeneous characteristics in a real-world environment. For instance, neither very cold nor very hot weather conditions are covered in our data. In addition, our dataset

does not contain road abrasion readings from unpaved streets. Generally, we aim to conduct more test drives using the introduced sensors in order to evaluate the model performance. Hyperparameter tuning of RF model parameters did not significantly increase the prediction accuracy in comparison to the default values set in the RF package. Therefore, we consent with default values and the achieved prediction accuracy. Apart from that, we see potential of optimisation in the calibration of single trajectory particulate matter readings when merging them. Due to varying ambient particulate matter levels in the air during measurements, there are different ranges for PM^{dry} in our trajectories. Ranges vary significantly throughout the dataset. Especially PM_{10}^{dry} values show big differences from trajectory to trajectory. We reach good classification accuracy by levelling the output values as proposed in this study. Still there is a need of taking into account more complex calibration methods. Taking particulate matter measurements from static counters and calculating the relative changes for each single trajectory could provide us with good results. RF has proven itself as an accurate classification method in many studies. Here, we can also conclude that RF performs well in this case study. Still testing other methods, e.g. linear regression, Gradient Boost or Support Vector Machines, might be beneficial.

5 CONCLUSIONS AND FUTURE WORK

In this study we use Random Forest to predict ambient traffic of a vehicle using road abrasion obtained from a low cost particulate matter sensor as predictor. We conclude that this works well with high accuracy for our datasets. The study shows that the model is capable to be tested as a Floating Car Observer. We notice that 60/40 and 70/30 train/test splits work best for our model with prediction accuracy values of up to 99 percent for single trajectories and up to 91 percent for the entire datasets. Sensitivity, Specificity, AUC and ROC outputs prove the validity of the classifications. Limitations exist mainly for the utilised sensors and the different initial ambient particulate matter levels in the air when measuring. Using low cost sensors proves applicable for our research. Since we do not monitor absolute particulate matter values and rather detect changes in the data caused by ambient traffic, we conclude that low cost particulate sensors can be used for detecting changes in road abrasion. Still we see a limitation in the response time specification of the sensors. In our study, rain is no significant limitation for measuring road abrasion. The model performs as good as under dry conditions due to the diameter of particles sensed. For our data, humidity correction is enough to decrease the influence of water particles in ambient air.

Generating more datasets to test the approach under more weather and road conditions will be subject of future research. Furthermore, we will evaluate if our fine dust measurements can also be used for mobile air quality assessment. In addition, we will test more machine learning classification techniques in order to evaluate the performance of our RF-based approach. Finally, we aim to integrate our approach into a Floating Car Observer framework.

REFERENCES

- [1] Brendan Abraham, Abhijith Mandya, Rohan Bapat, Fatma Alali, Don E Brown, and Malathi Veeraraghavan. 2018. A comparison of machine learning approaches

- to detect botnet traffic. In *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.
- [2] V Ariola, A D'Alessandro, F Lucarelli, G Marazzan, F Mazzei, S Nava, I Garcia-Orellana, P Prati, G Valli, R Vecchi, et al. 2006. Elemental characterization of PM₁₀, PM_{2.5} and PM₁ in the town of Genoa (Italy). *Chemosphere* 62, 2 (2006), 226–232.
- [3] Dirk Bank. 2007. *High-resolution ultrasonic sensing for autonomous mobile systems*. Ph.D. Dissertation. Universität Ulm.
- [4] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.
- [5] N Bukowiecki, P Lienemann, M Hill, M Furger, A Richard, F Amato, ASH Prévôt, U Baltensperger, B Buchmann, and R Gehrig. 2010. PM₁₀ emission factors for non-exhaust particles generated by road traffic in an urban street canyon and along a freeway in Switzerland. *Atmospheric Environment* 44, 19 (2010), 2330–2340.
- [6] George Dimitrakopoulos and Panagiotis Demestichas. 2010. Intelligent transportation systems. *IEEE Vehicular Technology Magazine* 5, 1 (2010), 77–84.
- [7] Nejdet Dogru and Abdulhamit Subasi. 2018. Traffic accident detection using random forest classifier. In *2018 15th learning and technology conference (L&T)*. IEEE, 40–45.
- [8] Janis Dröge, Ruth Müller, Cristian Scutaru, Markus Braun, and David A Groneberg. 2018. Mobile Measurements of Particulate Matter in a Car Cabin: Local Variations, Contrasting Data from Mobile versus Stationary Measurements and the Effect of an Opened versus a Closed Window. *International journal of environmental research and public health* 15, 12 (2018), 2642.
- [9] Robert Gehrig, Matz Hill, Brigitte Buchmann, David Imhof, Ernest Weingartner, and Urs Baltensperger. 2004. Separate determination of PM₁₀ emission factors of road traffic for tailpipe emissions and emissions from abrasion and resuspension processes. *International Journal of Environment and Pollution* 22, 3 (2004), 312–325.
- [10] Samuel E Georgian, Owen F Anderson, and Ashley A Rowden. 2019. Ensemble habitat suitability modeling of vulnerable marine ecosystem indicator taxa to inform deep-sea fisheries management in the South Pacific Ocean. *Fisheries Research* 211 (2019), 256–274.
- [11] Ali Golkarian, Seyed Amir Naghibi, Bahareh Kalantar, and Biswajeet Pradhan. 2018. Groundwater potential mapping using C5.0, random forest, and multivariate adaptive regression spline models in GIS. *Environmental monitoring and assessment* 190, 3 (2018), 149.
- [12] Gaby Gurczik. 2016. Bluetooth-Based Floating Car Observer: Model Evaluation Using Simulation and Field Measurements. *German Aerospace Center (DLR) Institute of Transportation Systems* (2016).
- [13] Gaby Gurczik. 2017. Performance measurement of a bluetooth-based floating car observer. *Transportation research procedia* (2017).
- [14] Markus Immitzer, Clement Atzberger, and Tatjana Koukal. 2012. Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. *Remote Sensing* 4, 9 (2012), 2661–2693.
- [15] HD Kim, JH Kim, KA Park, T Setoguchi, and S Matsuo. 2003. Computational study of the gas flow through a critical nozzle. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 217, 10 (2003), 1179–1189.
- [16] Kaarle J Kupiainen and Liisa Pirjola. 2011. Vehicle non-exhaust emissions from the tyre-road interface—effect of stud properties, traction sanding and resuspension. *Atmospheric Environment* 45, 25 (2011), 4141–4146.
- [17] Said Lahssini, Hicham Lahlaoui, Hicham Mharzi Alaoui, Martino Bagaram, Quentin Ponette, et al. 2015. Predicting cork oak suitability in Maamora forest using random forest algorithm. *Journal of Geographic Information System* 7, 02 (2015), 202.
- [18] Andy Liaw and Matthew Wiener. 2018. Package ‘randomForest’. (2018). <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf> Accessed 06/2020.
- [19] Andy Liaw, Matthew Wiener, et al. 2002. Classification and regression by randomForest. *R news* 2, 3 (2002), 18–22.
- [20] Thomas Liu. 2013. Digital relative humidity & temperature sensor AM2302/DHT22. <https://www.cdn-shop.adafruit.com/datasheets/Digital+humidity+and+temperature+sensor+AM2302.pdf> Accessed 06/2020.
- [21] Parvez Mahbub, Godwin A Ayoko, Ashantha Goonetilleke, Prasanna Egodawatta, and Serge Kokot. 2010. Impacts of traffic and rainfall characteristics on heavy metals build-up and wash-off from urban roads. *Environmental science & technology* 44, 23 (2010), 8904–8910.
- [22] Koreen Millard and Murray Richardson. 2015. On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. *Remote sensing* 7, 7 (2015), 8489–8515.
- [23] Plantower. 2016. Digital universal particle concentration sensor. http://download.kamami.pl/p563980-PMS3003%20series%20data%20manual_English_V2.5.pdf Accessed 06/2020.
- [24] Sutyajeet Soneja, Chen Chen, James M Tielsch, Joanne Katz, Scott L Zeger, William Checkley, Frank C Curriero, and Patrick N Breyse. 2014. Humidity and gravimetric equivalency adjustments for nephelometer-based particulate matter measurements of emissions from solid biomass fuel use in cookstoves. *International journal of environmental research and public health* 11, 6 (2014),

- 6400–6416.
- [25] Darius Štītis and Marius Laurinaitis. 2016. Legal regulation of the use of dashboard cameras: Aspects of privacy protection. *Computer Law & Security Review* 32, 2 (2016), 316–326.
 - [26] Pedro Strecht, LuÃs Cruz, Carlos Soares, JoÃ Mendes-Moreira, et al. 2015. A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance. *International Educational Data Mining Society* (2015).
 - [27] Norbert Streibl. 2017. *Influence of Humidity on the Accuracy of Low-Cost Particulate Matter Sensors*. Technical Report. Technical report. DOI: 10.13140/RG [Titel anhand dieser DOI in Citavi-Projekt übernehmen] . 2.2. 21095.75683.
 - [28] Vladimir Svetnik, Andy Liaw, Christopher Tong, J Christopher Culberson, Robert P Sheridan, and Bradley P Feuston. 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of chemical information and computer sciences* 43, 6 (2003), 1947–1958.
 - [29] Touko Tcheumadjeu, Louis Calvin, Andreas Luber, Elmar Brockfeld, Gaby Gurczik, Alexander Sohr, and Anke Sauerländer-Biebl. 2017. Integration of mobile wireless RF sensors into a traffic information system. (2017).
 - [30] Toni Weisheit, Carsten Kühnel, and Robert Hoyer. 2011. Video-and sound parameter based detection of opposing traffic in Floating Car Observer applications. In *Proceedings of the 18th World Congress on Intelligent Transport Systems, Orlando*.
 - [31] Stefan Wender and Klaus Dietmayer. 2008. 3D vehicle detection using a laser scanner and a video camera. *IET Intelligent Transport Systems* 2, 2 (2008), 105–112.
 - [32] Franziska Wolf, Sebastian Naumann, Christoph Engel, and Rene Schonrock. 2008. Low-cost Floating Car Observer Implementation by a Video Camera based Method. In *15th World Congress on Intelligent Transport Systems and ITS America's 2008 Annual MeetingITS AmericaERTICOITS JapanTransCore*.
 - [33] Yizheng Wu, Yuxin Wang, Lewen Wang, Guohua Song, Jian Gao, and Lei Yu. 2020. Application of a taxi-based mobile atmospheric monitoring system in Cangzhou, China. *Transportation Research Part D: Transport and Environment* 86 (2020), 102449.
 - [34] Junping Zhang, Fei-Yue Wang, Kunfeng Wang, Wei-Hua Lin, Xin Xu, and Cheng Chen. 2011. Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems* 12, 4 (2011), 1624–1639.
 - [35] Ferdinand H Farwick zum Hagen, Marcel Mathissen, Tomasz Grabiec, Tim Henricke, Marc Rettig, Jaroslaw Grochowicz, Rainer Vogt, and Thorsten Benter. 2019. On-road vehicle measurements of brake wear particle emissions. *Atmospheric Environment* 217 (2019), 116943.