MY_ATMOS : NOVEL METHOD TO ANALYSE ULTRAFINE PARTICLES USING AN ARTIFICIAL INTELLIGENCE APPROACH.

Yahaya, N.Z. [1], Miles R. Tight[2], James E. Tate.[3], Zul Fadhli Ibrahim[1]
[1]School of Ocean Engineering, Universiti Malaysia Terengganu, MALAYSIA.
[2]School of Civil Engineering, University of Birmingham, UNITED KINGDOM.
[3]Institute for Transport Studies, The University of Leeds, UNITED KINGDOM
Corresponding author : nzaitun@umt.edu.my; Tel No : +609-668 3972

INTRODUCTION TO MY_ATMOS

This presentation will discuss the used of an artificial intelligent method namely the 'stochastic boosted regression trees' (BRT) approach that using an algorithm that applied to an air pollution data namely particle number count concentrations ([PNC]), an ultrafine particles data and particulate matter data case study in United Kingdom and Malaysia.

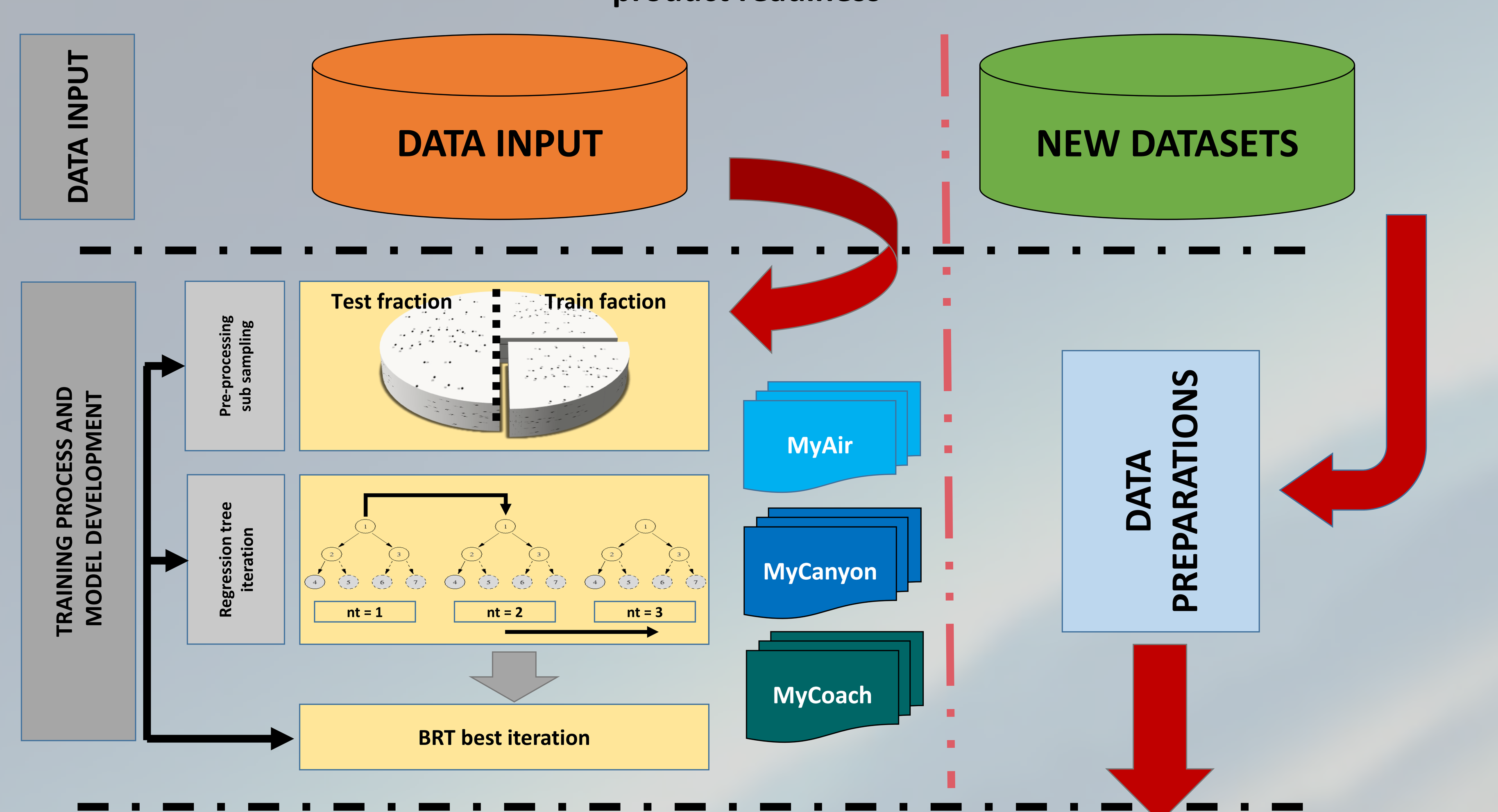
The development of the BRT model involves determining the model algorithm settings of the main model input parameters (*learning rate, number of trees and interaction depth*) that were tested using the R software (version 3.02) by choosing a10-fold cross-validation approach with combination of *lr* 0.05 and *tc* 5 of training set for BRT models.
My_Atmos functions as a predictor tool and applicable to visualized concentration of pollutions overlay on a map which is generated from My_Atmos system. The advantage of the base of My_Atmos, BRT:

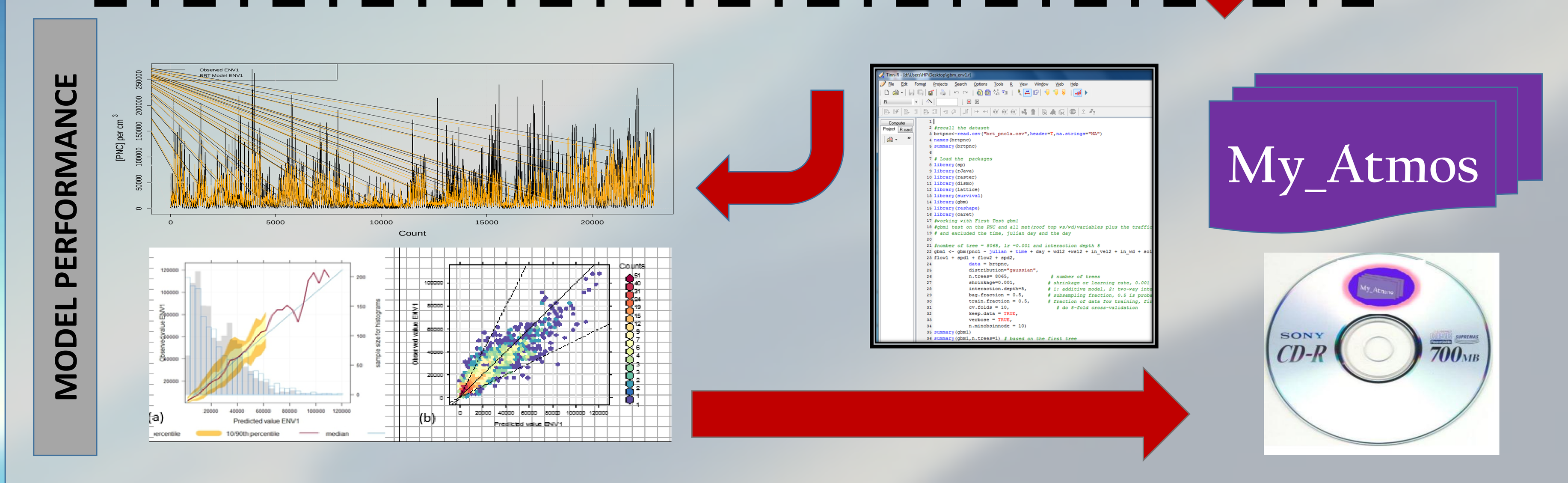
- BRT method can deal with complex data and explain the variability of data;
- BRT model can provide a much smoother gradient, analogous to the fit achieved when using gbm and My_Atmos packages;
- BRT technique handle sharp discontinuities, which is an important advantage when modelling the distributions of pollutants that only occupy a small proportion of the sample environmental space.

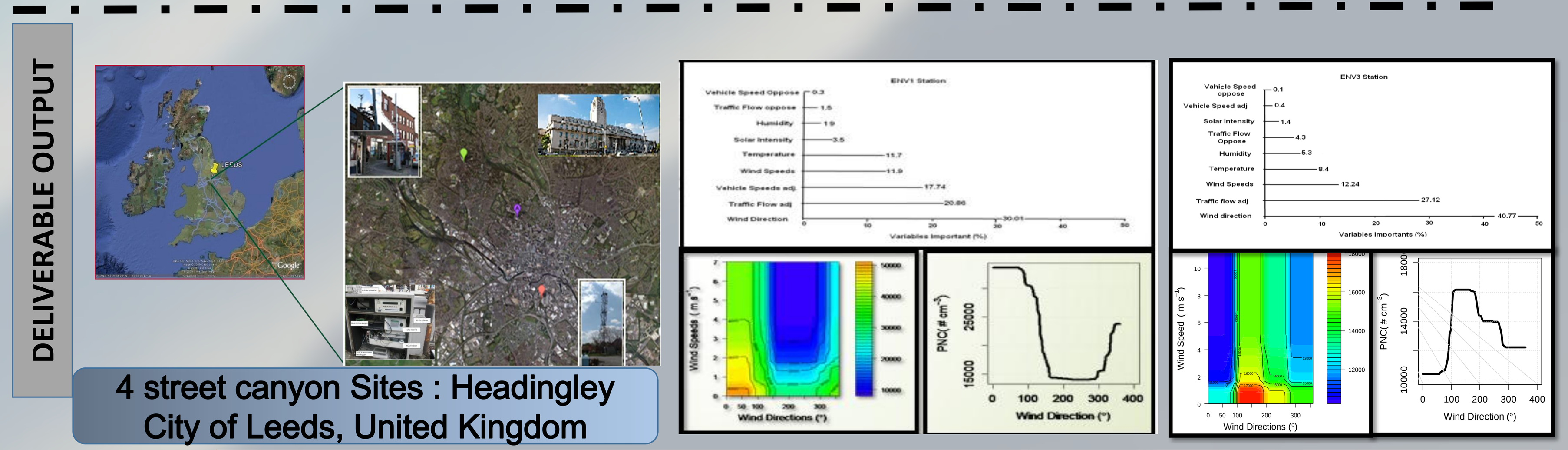
It was found, that the coefficient of determination (R^2) value for the BRT best iteration models were above 0.60 for [PNC] in urban environment. The fine and course particle number (FPNC and CPNC) were found to be 0.75 and 0.72 respectively for one of coastal dataset while R^2 value of 0.78 and 0.85 were obtained for Malaysia data. Further investigated were performed to rank factor influenced. It was found, that Carbon monoxide (30.28 %) gas and followed by temperature (16.81%) and wind direction (16.4%) were found the high factor influenced PM_{10} in urban environment. The interaction index (H-index) between parameters to concentration of pollutions were also examined graphically and in numerical form (H-Index). It was found that the H-Index between parameters 0.3 to 0.4 indicated that the BRT technique able to explain the science of air pollution. The consistent results to produce the best model from the best iteration, able to rank the best parameters that influence most to the concentration of predictor and able to predict interaction between variables premise BRT as one of the method or tools to analyse air pollution data. My_Atmos is interactive and applicable to any industry. BRT own geographical view of its operating system demonstrates applicability for both domestic and international end users as the user can use this model from any part of the world.

MODEL DEVELOPMENTS

The working system is to demonstrates structures, functionality of the algorithm setting and product readiness



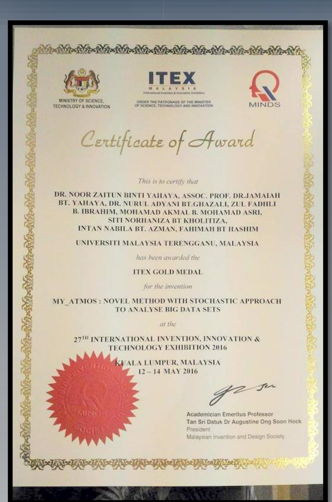




BRT [PNC] algorithm developed Data analysis in two street canyon data sample.

CERTIFICATION

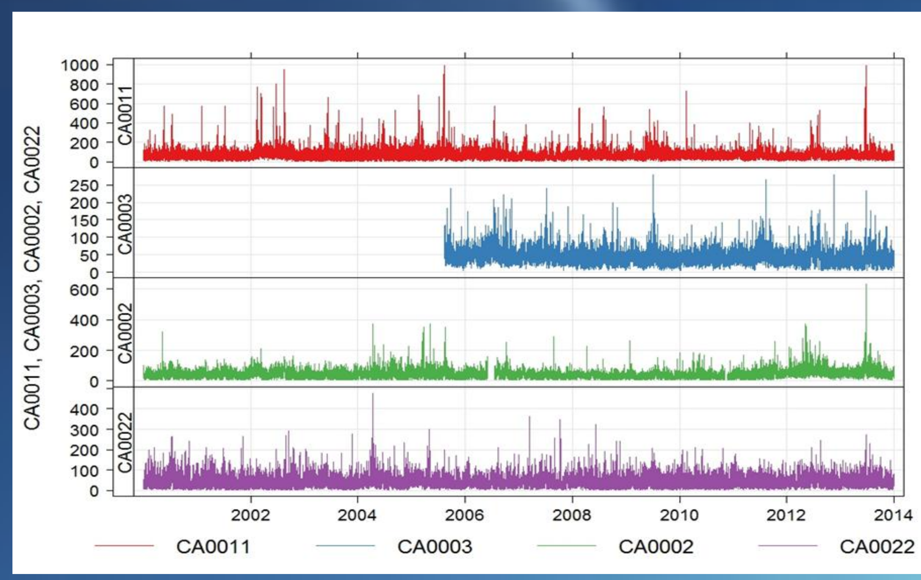

The Queens Anniversary Award Winner for the Higher Education Institution with The Ins. for Transport, University of Leeds, United Kingdom 2010




Gold Medalist at 27th International Invention & Innovation Exhibition (ITEX 2016)

Malaysia Air Quality Monitoring data

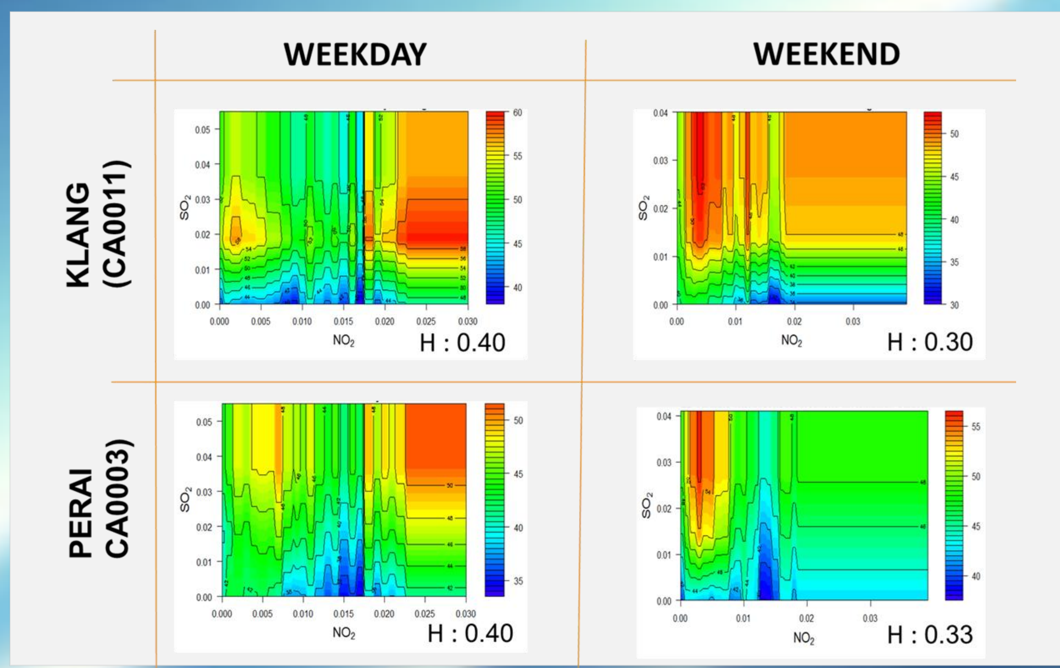
Continuous time series data gathered from AQMS Malaysia from year 2000 – 2013 (with 122,640 hourly data were collected from two monitoring stations, 9 parameters were involved in the model development processes



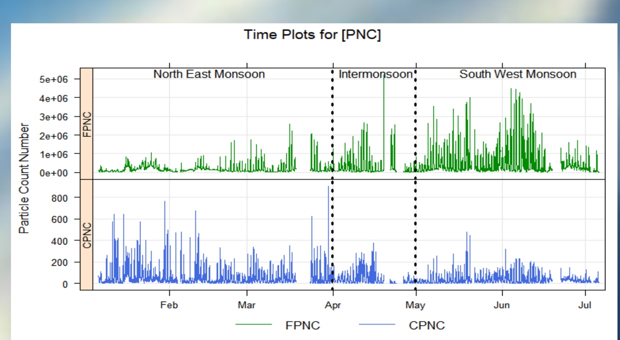
RANK	VARIABLES	UNIT
1	PM_{10}	µg/m^3
2	CO	ppm
3	NO (ppm)	ppm
4	NO_2 (ppm)	ppm
5	SO_2 (ppm)	ppm
6	WS	m/s
7	WD	Degree (°)
8	RH	Percent (%)
9	T	Degree Celcius (°C)

BRT Output

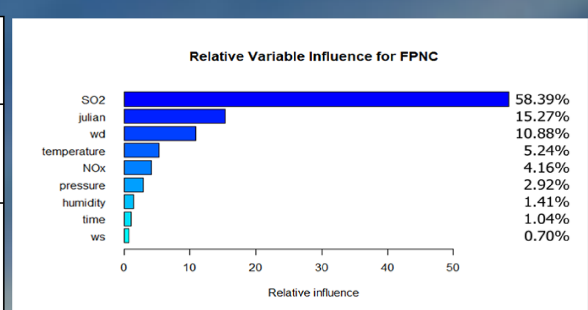
	PERAI STATION		KLANG STATION	
RANK	VARIABLES	% INFLUENCE	VARIABLES	% INFLUENCE
1	PM_{10}lag	16.872	CO (ppm)	36.89
2	t_{jd}	12.251	t_{hour}	14.83
3	NO_2 (ppm)	10.142	PM_{10} lag	14.83
4	CO (ppm)	9.825	NO (ppm)	7.64
5	ws (ms^{-1})	8.13	t_{jd}	5.31
6	wd (degree)	8.003	wd (degree)	4.46
7	SO_2 (ppm)	7.634	temp (°C)	4.08
8	NO (ppm)	7.49	NO_2 (ppm)	3.7
9	temp (°C)	7.47	rh (%)	3.6
10	rh (%)	5.717	ws (ms^{-1})	3.18
11	t_{hour}	4.545	SO_2 (ppm)	2.93
12	$t_{weekday}$	1.92	$t_{weekday}$	0.88

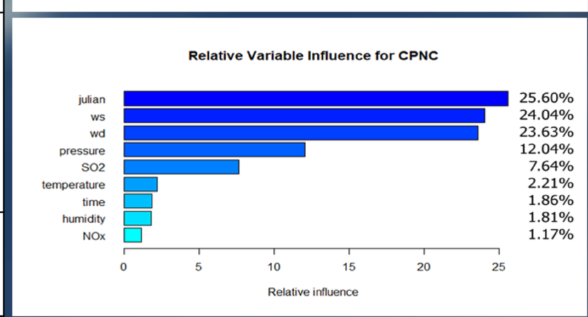


Sampling : Bachok Kelantan, Malaysia (Compilation of 2015 [PNC] datasets



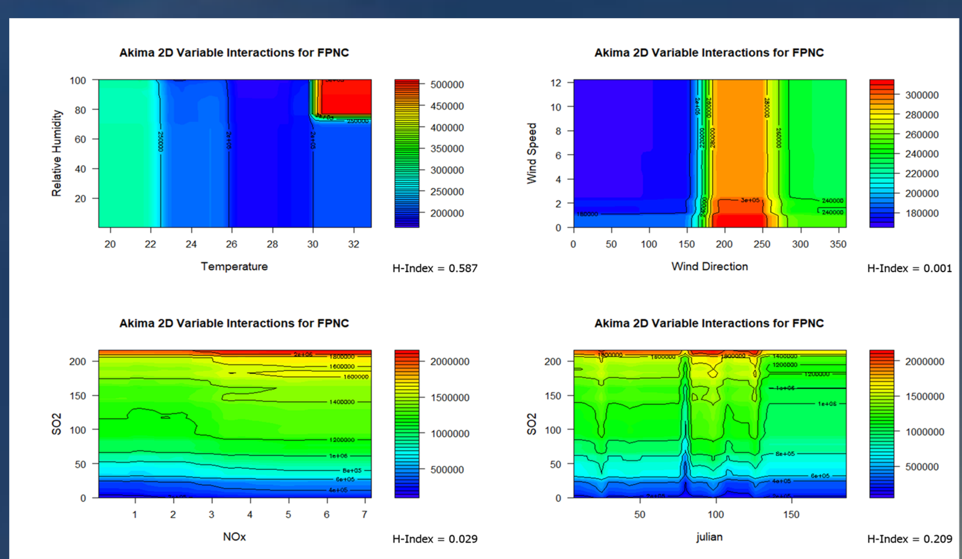
Variables	Response and predictors	Descriptions and units
Response	**[PNC]** FPNC CPNC	[PNC] 0.265 to 0.900 µm [PNC] 2.75 to 9.25 µm
Predictors 1 2	**Time System** Time Julian	Hour of the day (1 to 24 hour) Day of the year (Julian Day)
3 4 5 6 7	**Meteorological Factors** Temperature Humidity Pressure Wind speed Wind direction	Ambient temperature (˚C) Relative humidity (% rh) Pressure (hPa) Wind speed (m/s) Wind direction (degree)
8 9	**Gases** SO_2 NO_x	Sulphur Dioxide (ppb) Nitrogen Oxide (ppb)

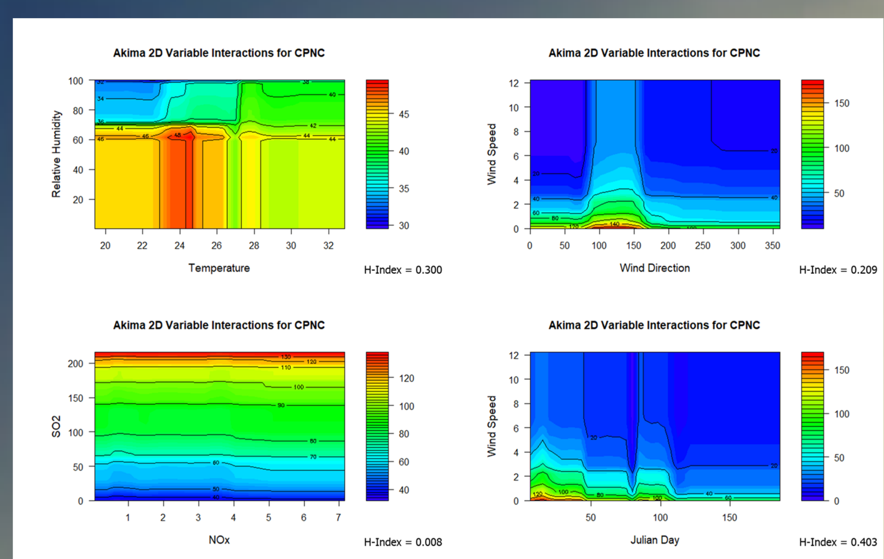




Model	FAC2	RMSE	R	R2	COE	IOA
FPNC	0.803	217,904.94	0.867	0.752	0.565	0.783
CPNC	0.757	27.592	0.846	0.716	0.54	0.77

Sample of the Interaction between variables by using an akima countor plots





CONCLUSION

1. The BRT method has the ability to deal with complex data and explain the variability of particles in environment;
2. The combination of lr 0.05 and tc 5 of training set for BRT models was the lowest error of RMSE compared to other combination for the optimum settings for urban industrial stations.
3. The performance of all the models are within acceptable range which is up to 61% variance explain.
4. BRTs demonstrated the contribution from industrial zone and where by motor vehicles dominant sources of PM_{10} was identified and their interactions (Index up to 0.4) between factors can explain fundamental facts.
5. A clear benefit of BRTs for air pollution applications is their ability to model complex variable interactions and non-linear effects, which are the norm in air pollution and which can be difficult to determine and model using other approaches.

REFERENCES

Carslaw, D.C. & Taylor, P.J. 2009. Analysis of air pollution data at mixed source location using boosted regression trees. Atmospheric environment, 43, 7053-7063.
Friedman, J.H., 2002. Stochastic gradient boosting. Computational Statistics & Data Analysis 38 (4), 367–378.
Sayeh. A. A, Munir, S., Habeebullah, T.M. 2014. Comparing the performance of Statistical Models for predicting PM10 Concentrations. Aerosol and Air Quality Research, 14: 653 – 665.
Yahaya, N.Z., Ghazali, N.A., Ahmad, S., Asri, M.A.M., Ibrahim, Z.F., and Ramli, N.A. 2017. Analysis of Daytime and Nighttime Ground Level Ozone Concentrations Using Boosted Regression Tree Technique. Environment Asia 10(1),*118-129*
Yahaya, N.Z. 2012. A study of the temporal and spatial variation of ultra-fine particles in the urban environment. (Doctoral dissertation).
NZ Yahaya, SM Phang, AA Samah, IN Azman. 2018. Analysis of Fine and Coarse Particle Number Count Concentrations Using Boosted Regression Tree Technique in Coastal Environment. Journal of EnvironmentAsia

Acknowledgements:
- The Engineering and Physical Sciences Research Council (EPSRC) and Institute for Transport studies University of Leeds, United Kingdom
- Ministry of Higher Education Malaysia (MOHE) for Fundamental Research Grant , Malaysia for funding this research;
- Institute for Ocean and Earth Science (IOES), University Malaya, Malaysia for providing data from IOES station;