International Workshop On Land Cover/Land Use Changes, Forestry, and Agriculture in South/Southeast Asia, Phnom Penh, Cambodia. 8-10 August 2022

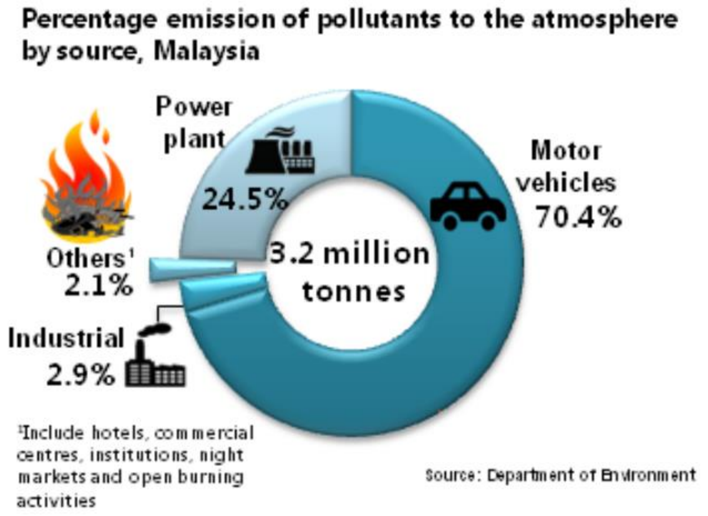
Mapping Particulate Matters <2.5 µm in Malaysia using Himawari-8 and Sentinel 5P satellite products

KasturiDevi Kanniah, Nurul Amalin Fatihah, Dimitris Kaskaoutis*
TropicalMap Research Group
Centre for Environmental Sustainability and Water Security
Fac. Built Environment & Surveying
UNIVERSITI TEKNOLOGI MALAYSIA

* Institute for Environmental Research and Sustainable Development, National Observatory of Athens, 15236 Athens, Greece

INTRODUCTION

- Air pollution is a serious environmental issue in Malaysia
- The main sources of air pollutants:
- PM2.5 and NO2 are main pollutants
- Average PM2.5 is 20 μg m^{-3} in 2019 (>WHO limit of 5 μg m^{-3})
- Inhalable and cause lung and eye diseases, birth defects, premature deaths, allergic and degradation of visibility
- In Malaysia, premature mortality and respiratory problem increase tremendously during the severe haze events.



- PM2.5 data is measured by the Department of Environment, Malaysia since 2017 covering 65 stations across Malaysia
- Main indicator to determine air quality in Malaysia

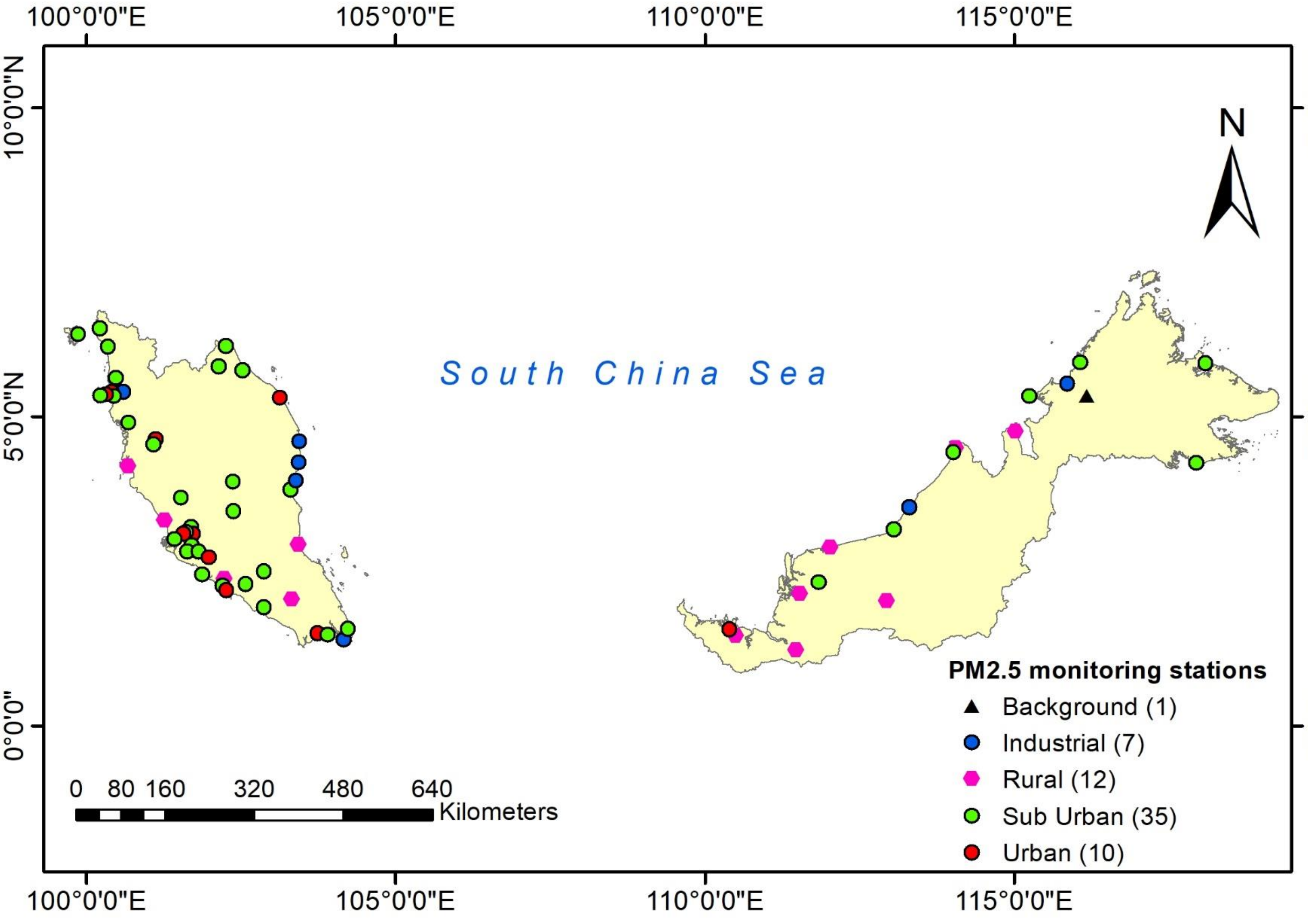
RESEARCH GAP

- PM2.5 data are not available in a spatially explicit manner
- Only few researches on PM2.5 in Malaysia- focus on a small spatial scale (Khan et al., 2015; Amil et al., 2016; Dahari et al., 2020).
- Remote sensing is a valuable tool for investigating the concentration and variability of PM2.5 at regional scale
- Modelling PM 2.5 with gases pollutants can improve estimation

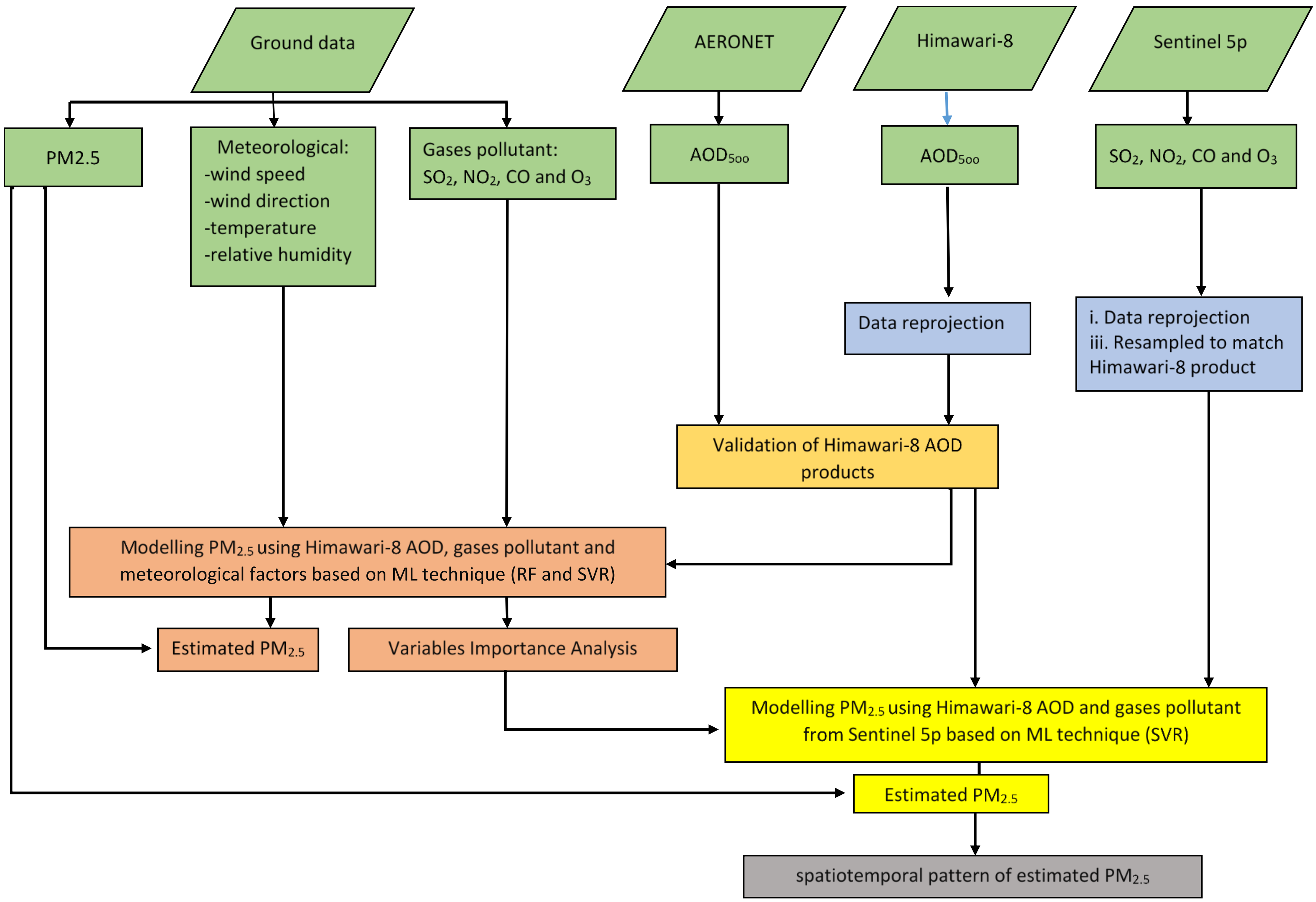
OBJECTIVE

To develop empirical models for estimating PM2.5 concentrations in Malaysia using AOD and gases pollutants

PM2.5 MONITORING STATIONS



WORK FLOW

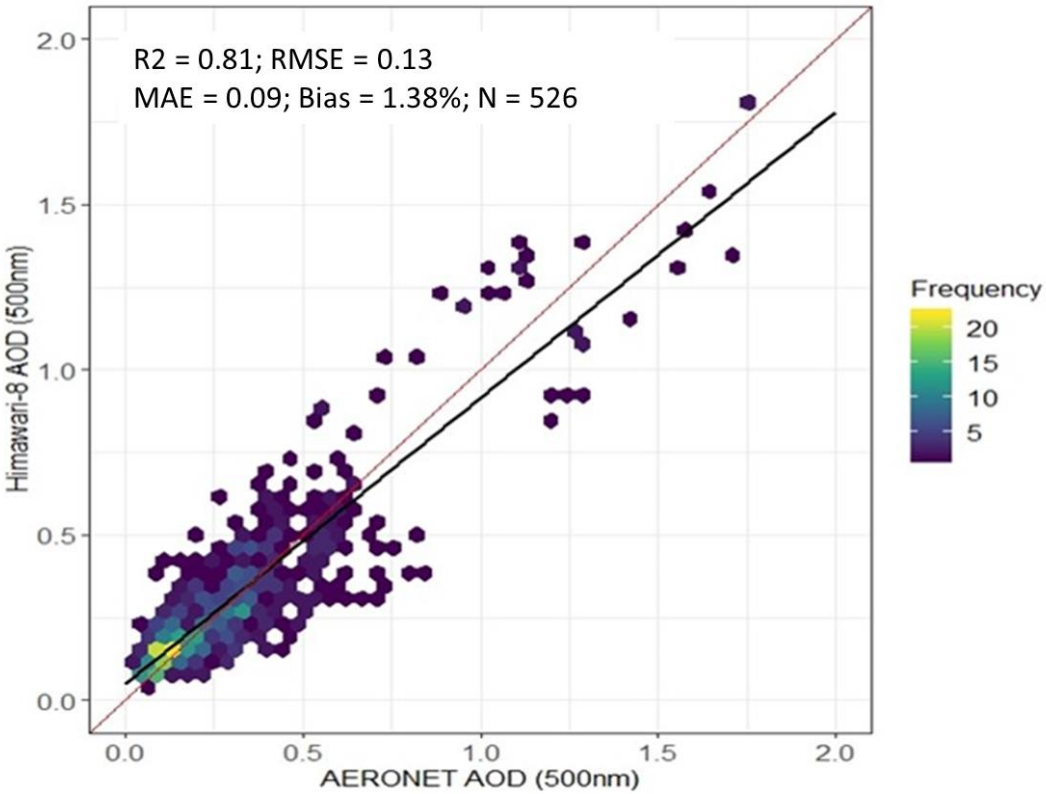


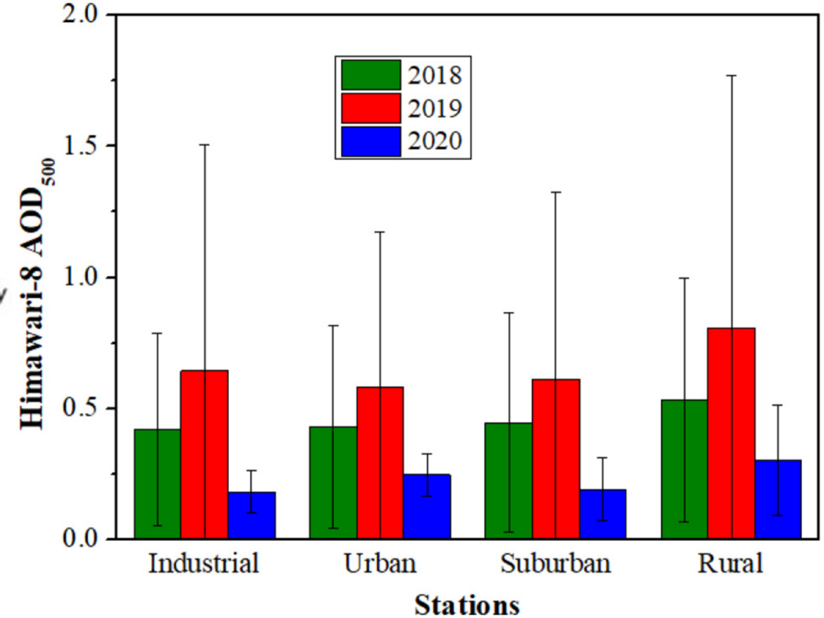
Ground data
AERONET
Himawari-8
Sentinel 5p
PM2.5
Meteorological:
-wind speed
-wind direction
-temperature
-relative humidity
Gases pollutant:
SO_2, NO_2, CO and O_3
AOD_{500}
AOD_{500}
SO_2, NO_2, CO and O_3
Data reprojection
i. Data reprojection
iii. Resampled to match Himawari-8 product
Validation of Himawari-8 AOD products
Modelling $PM_{2.5}$ using Himawari-8 AOD, gases pollutant and meteorological factors based on ML technique (RF and SVR)
Estimated $PM_{2.5}$
Variables Importance Analysis
Modelling $PM_{2.5}$ using Himawari-8 AOD and gases pollutant from Sentinel 5p based on ML technique (SVR)
Estimated $PM_{2.5}$
spatiotemporal pattern of estimated $PM_{2.5}$

EMPIRICAL MODELS

Model	**Spatial/Temporal details**
Model 1	Overall model
Model 2	Urban/industry
Model 3	Suburban/rural
Model 4	Dry season
Model 5	Wet season
Model 6	Inter-monsoon (April–May) season
Model 7	Inter-monsoon (October) season

VALIDATION OF HIMAWARI-8 AOD PRODUCT





Kanniah *et al* (2020)

DESCRIPTIVE STATISTICS OF MEASURED VARIABLES

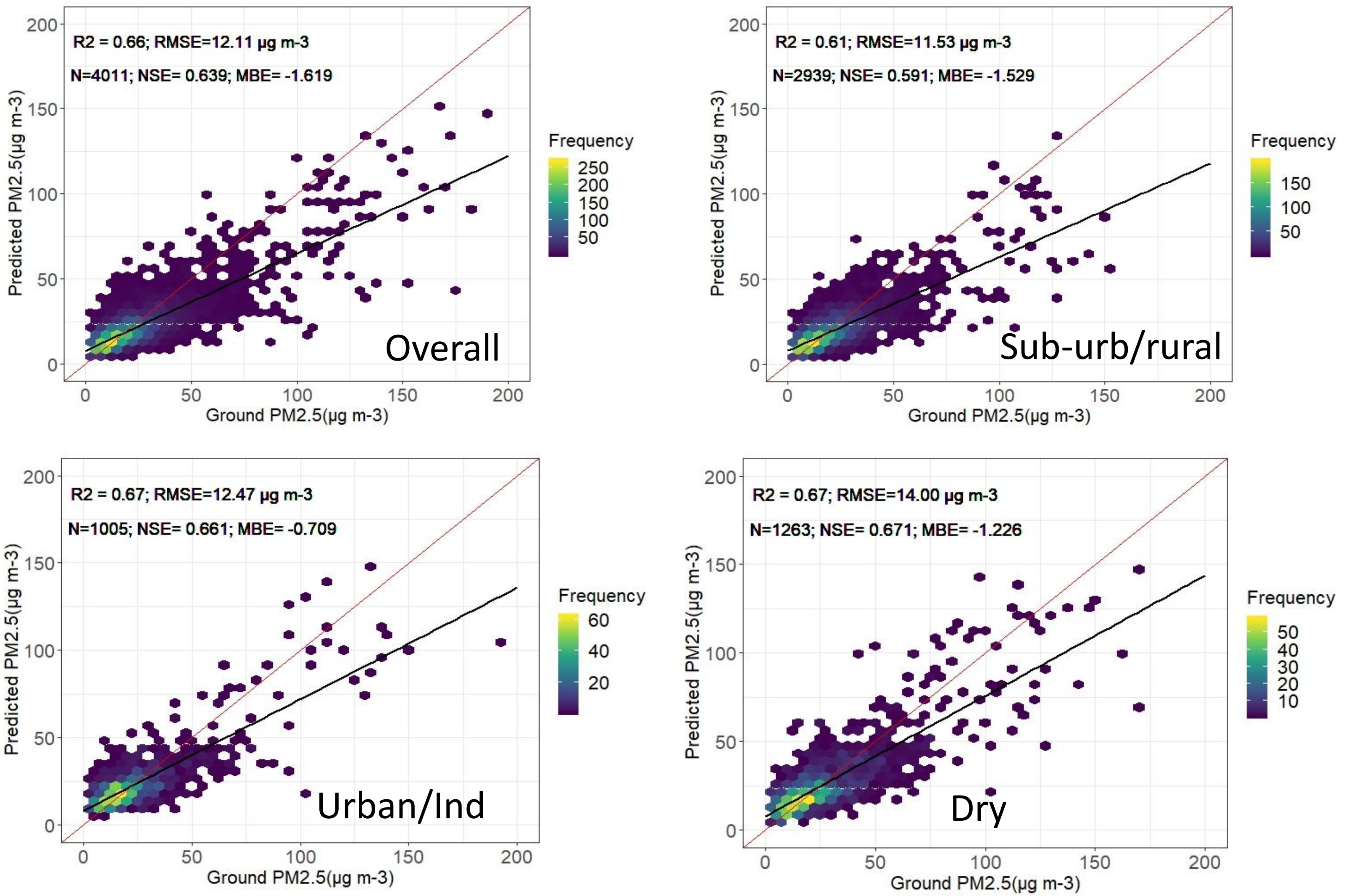
	$PM_{2.5}$ (µg m^{-3})	AOD	SO_2 (ppb)	NO_2 (ppb)	O_3 (ppb)	CO (ppm)	WS (ms^{-1})	RH (%)	TEMP (°C)
Mean	21.86	0.69	1.2	5.23	25.2	0.60	1.73	66.58	30.70
Median	17.07	0.46	1.0	3.81	27.3	0.56	1.60	66.85	30.88
Stdev	19.15	0.68	0.9	6.1	15.1	0.28	1.02	10.35	2.31

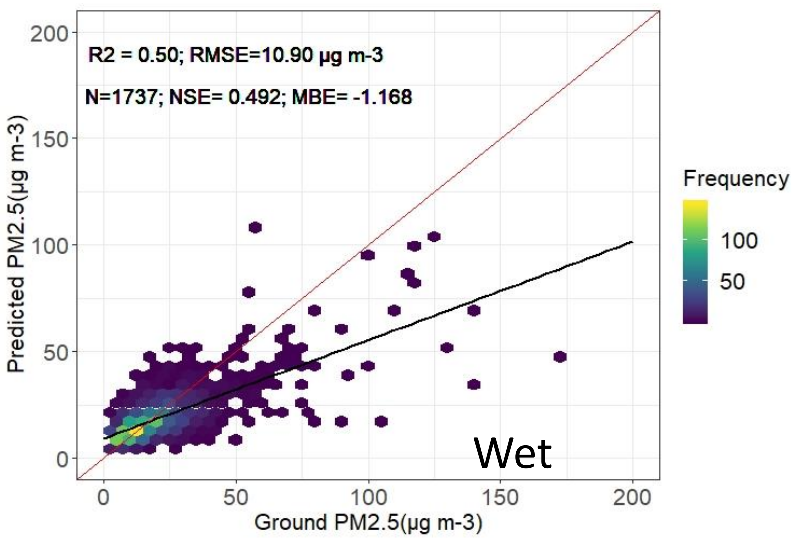
- The columnar AOD500 during 2018–2019 exhibits a mean of 0.69, which is above the median value 0.46 due to many episodic aerosol events with AODs above 2,
- The measured PM2.5 concentrations have higher mean (21.9 µg m−3) than median (17.1 µg m−3) and a maximum value of 230 µg m−3
- These PM2.5 levels are similar to those reported at several sites in Southeast Asia (Dahari et al., 2020; Nguyen et al., 2021)
- NO2 and CO exhibit means of 5.2 ppb and 0.6 ppm
- Tropospheric O3 levels (25.2 ppb) are considered rather high with deleterious effects on human health

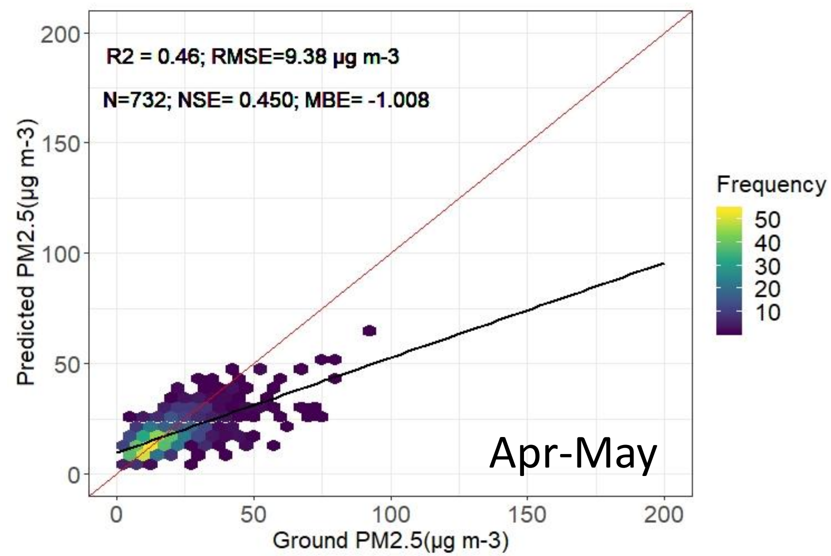
MODELS FOR $PM_{2.5}$ ESTIMATION

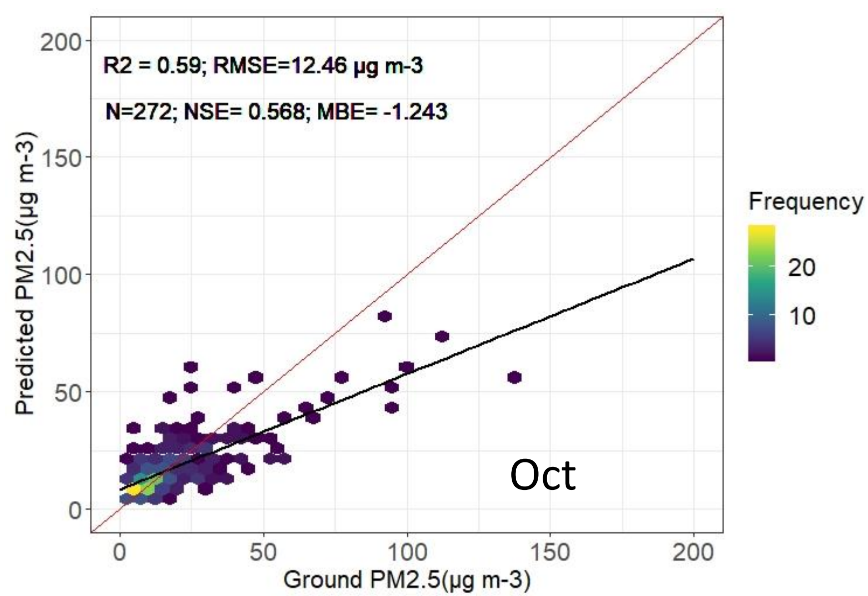
Models	Calibration dataset				
	R^2	RMSE	MBE	NSE	No of samples (N)
1	0.69	10.62	-1.392	0.679	9365
2	0.79	10.23	-1.191	0.777	2352
3	0.69	10.67	-1.416	0.668	6859
4	0.81	11.18	-1.187	0.794	2960
5	0.62	9.96	-1.437	0.599	4060
6	0.58	8.62	-0.994	0.555	1709
7	0.74	11.17	-2.323	0.687	643

PREDICTED VERSUS MEASURED PM2.5

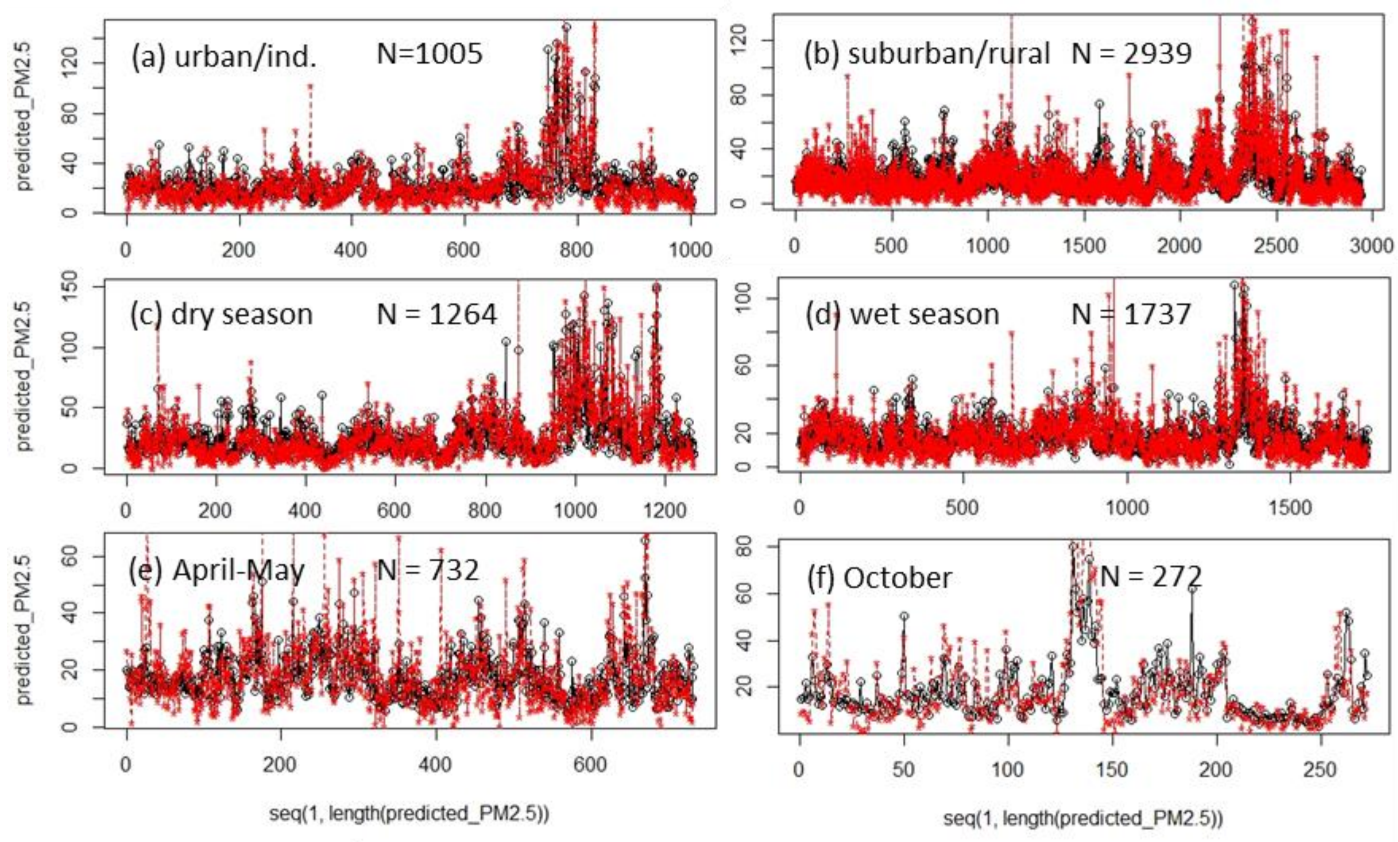


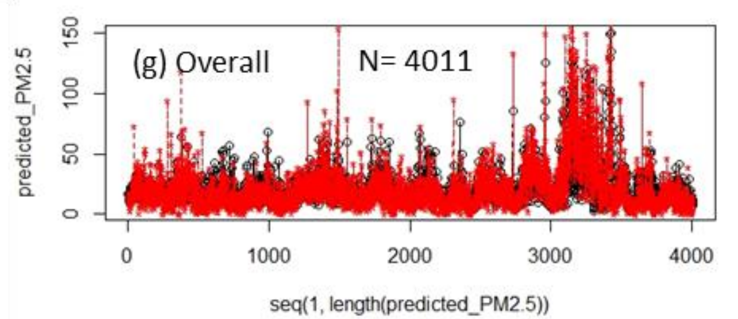

R2 = 0.50; RMSE=10.90 µg m-3
N=1737; NSE= 0.492; MBE= -1.168
Predicted PM2.5(µg m-3)
Ground PM2.5(µg m-3)
Frequency
100
50
Wet
0
50
100
150
200


R2 = 0.46; RMSE=9.38 µg m-3
N=732; NSE= 0.450; MBE= -1.008
Predicted PM2.5(µg m-3)
Ground PM2.5(µg m-3)
Frequency
50
40
30
20
10
Apr-May
0
50
100
150
200


R2 = 0.59; RMSE=12.46 µg m-3
N=272; NSE= 0.568; MBE= -1.243
Predicted PM2.5(µg m-3)
Ground PM2.5(µg m-3)
Frequency
20
10
Oct
0
50
100
150
200

Time series hourly measured and predicted PM2.5

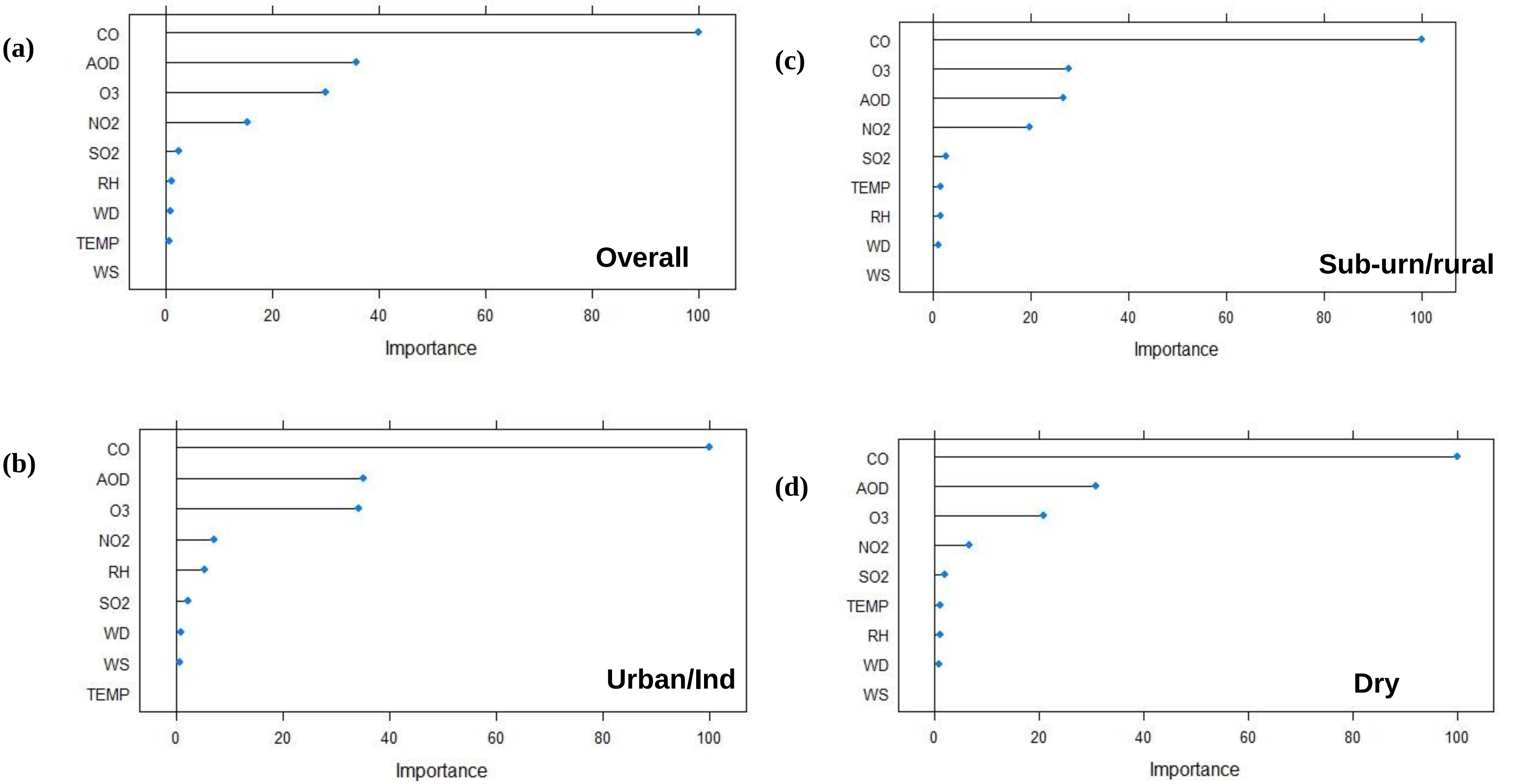




Time series (hourly) of measured (black) and predicted (red) PM2.5 concentrations from the SVR model across Malaysia

Variable Importance Analysis

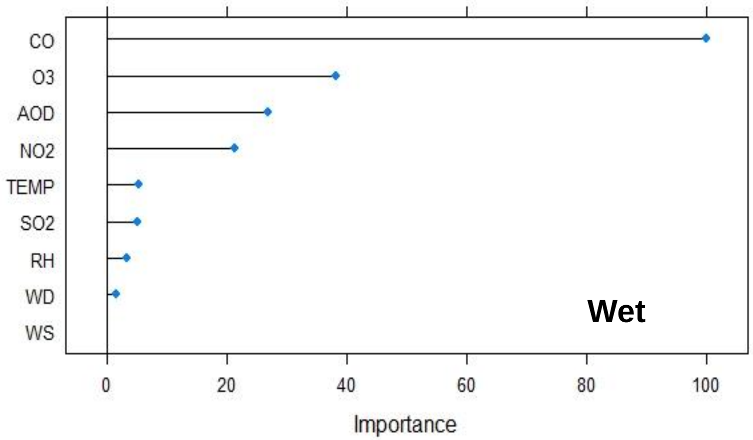
Variable importance statistic was used to analyse the contribution of each variable in PM2.5 estimations



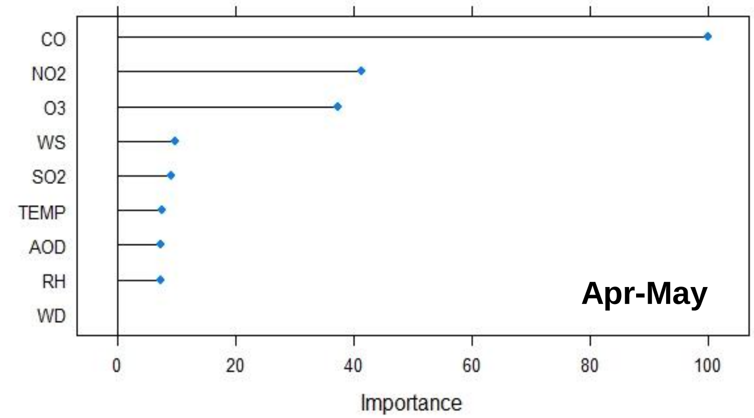
VIA- Meteorological data show importance values <10% and 0% for wind speed

Variable Importance Analysis

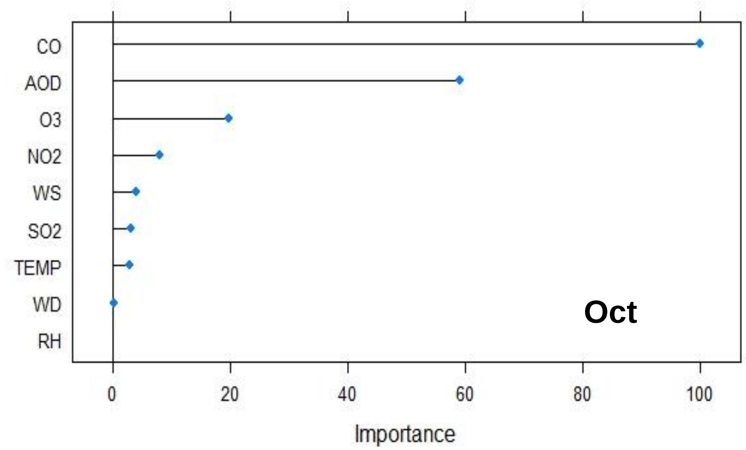
(e)



(f)



(g)



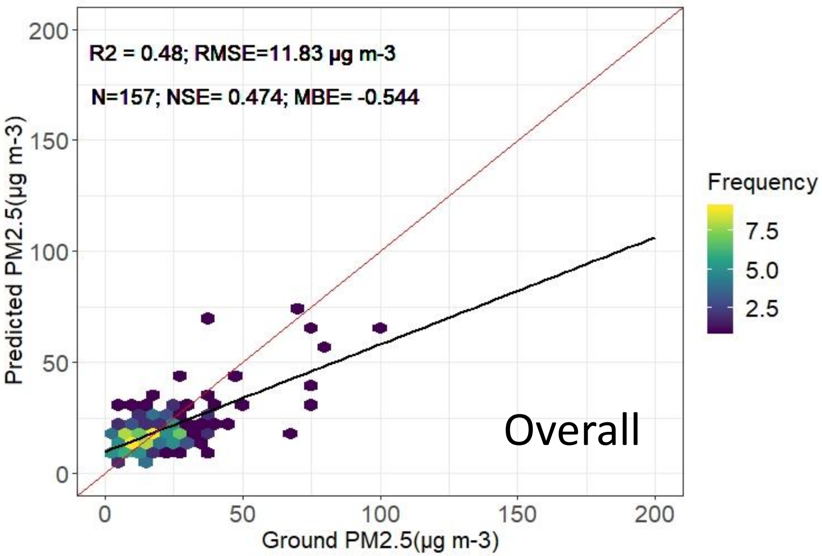
INCORPORATING GASES DATA FROM SENTINEL 5-P

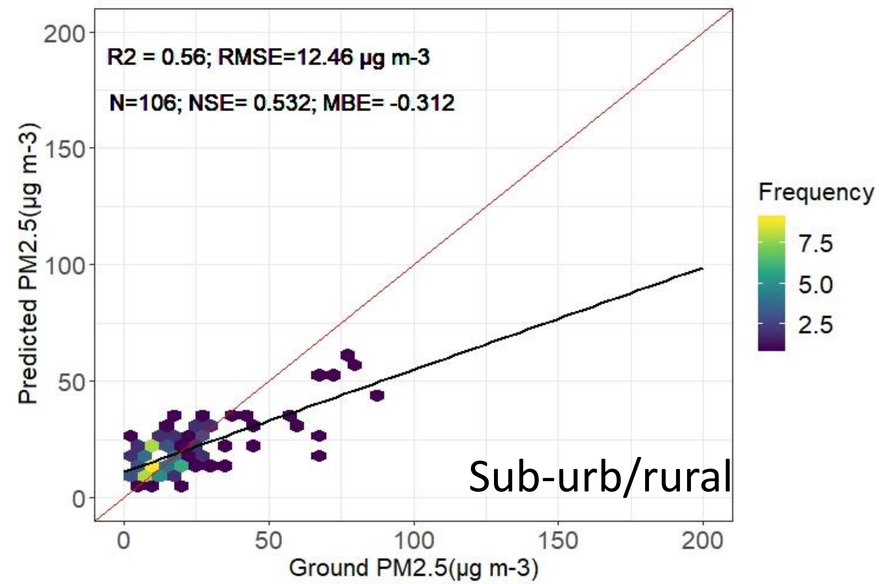
- Variable Importance Analysis shows AOD (31-60%) and gases (CO-100%, O_3-20-40%, NO_2- 9-21% and SO_2- 3-11%) are significant for estimating PM2.5
- Meteorological data had less importance (<10%)
- In Beijing–Tianjin–Hebei, China CO, NO_2, SO_2 and AOD were important variables in predicting $PM_{2.5}$ concentrations (Wang and Sun (2019)
- In Xi'an, China, Song et al. (2015) found CO was the most important variable that explained 21% of the variation in estimated $PM_{2.5}$ concentrations
- In the Klang Valley, Malaysia- CO, NO_2, NO and SO_2 are mostly affected the $PM_{2.5}$ concentrations (Amil *et al.*, 2016).

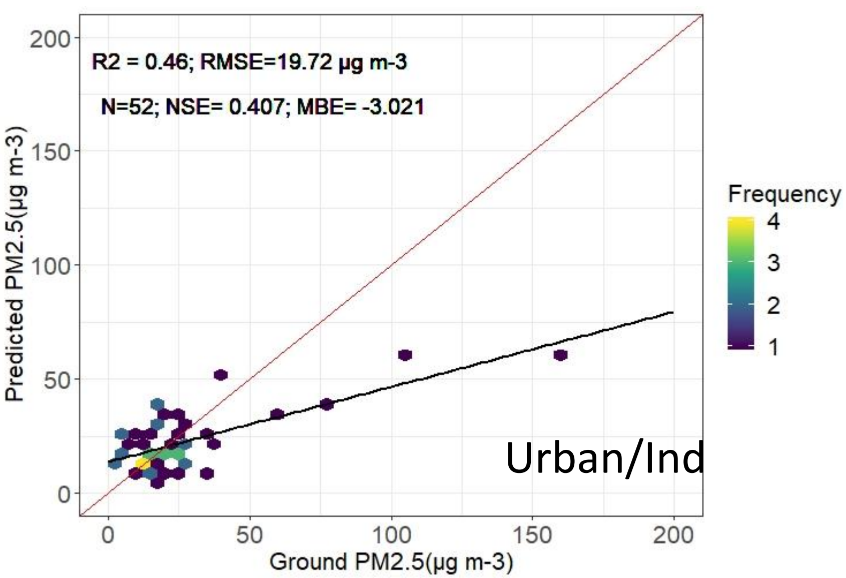
MODELS FOR $PM_{2.5}$ ESTIMATION

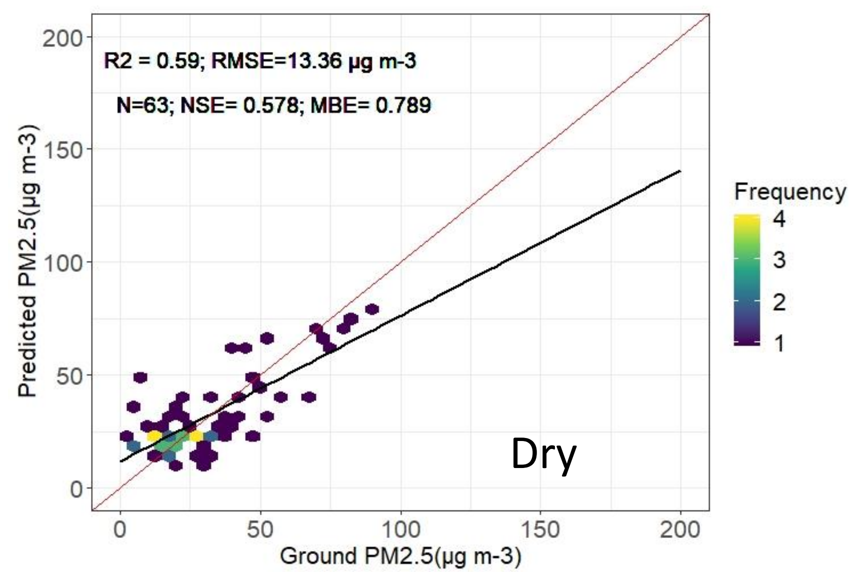
Model	Calibration dataset				
	R^2	RMSE	MBE	NSE	No of sample (N)
1	0.67	13.36	-2.016	0.645	365
2	0.67	17.28	4.435	0.608	118
3	0.61	10.17	-1.268	0.583	246
4	0.76	16.08	-3.072	0.71	147
5	0.59	5.01	-0.477	0.571	129
6	0.56	5.56	-0.507	0.506	72
7	0.86	11.39	-3.406	0.791	21

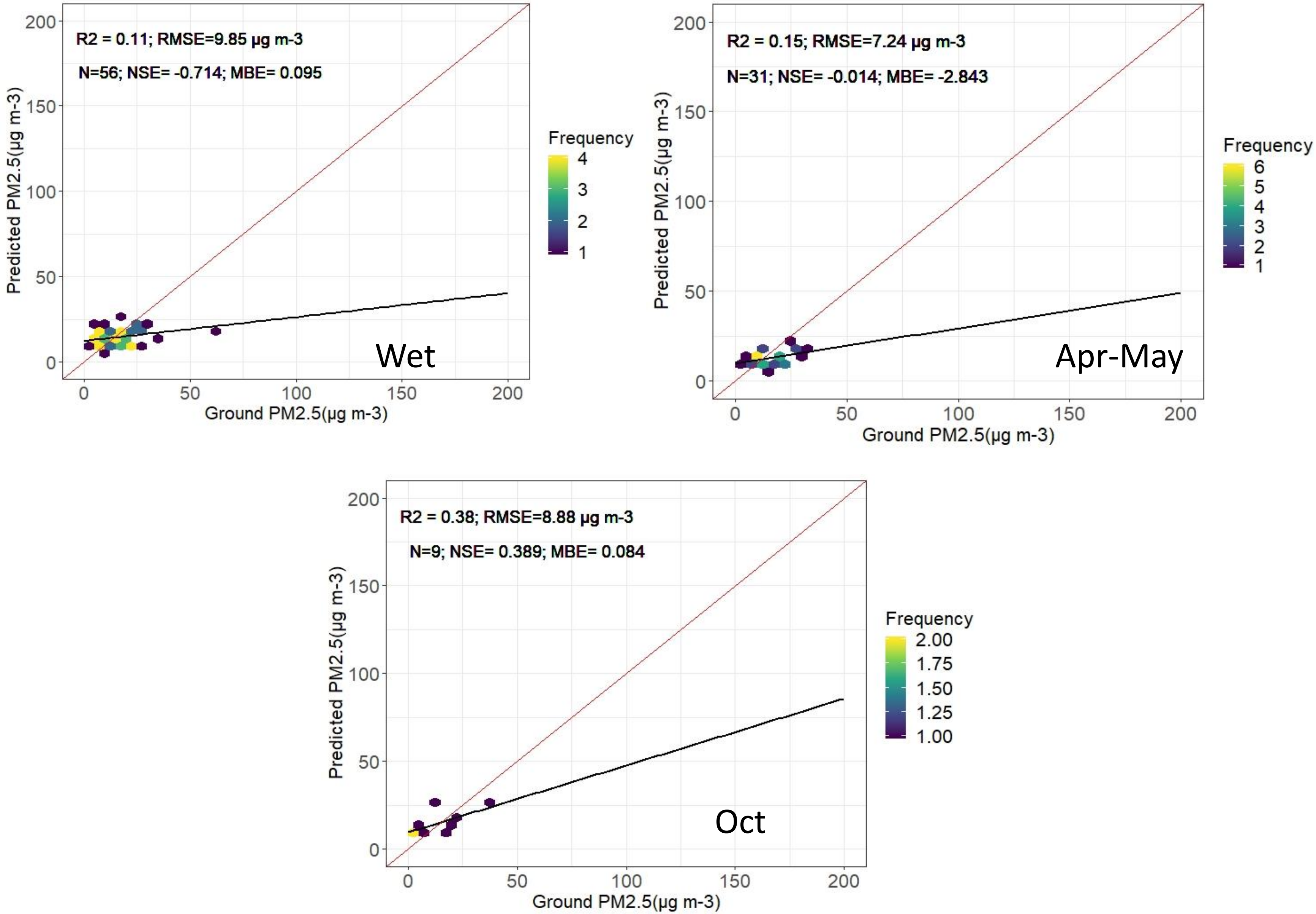
PREDICTED AND MEASURED PM2.5



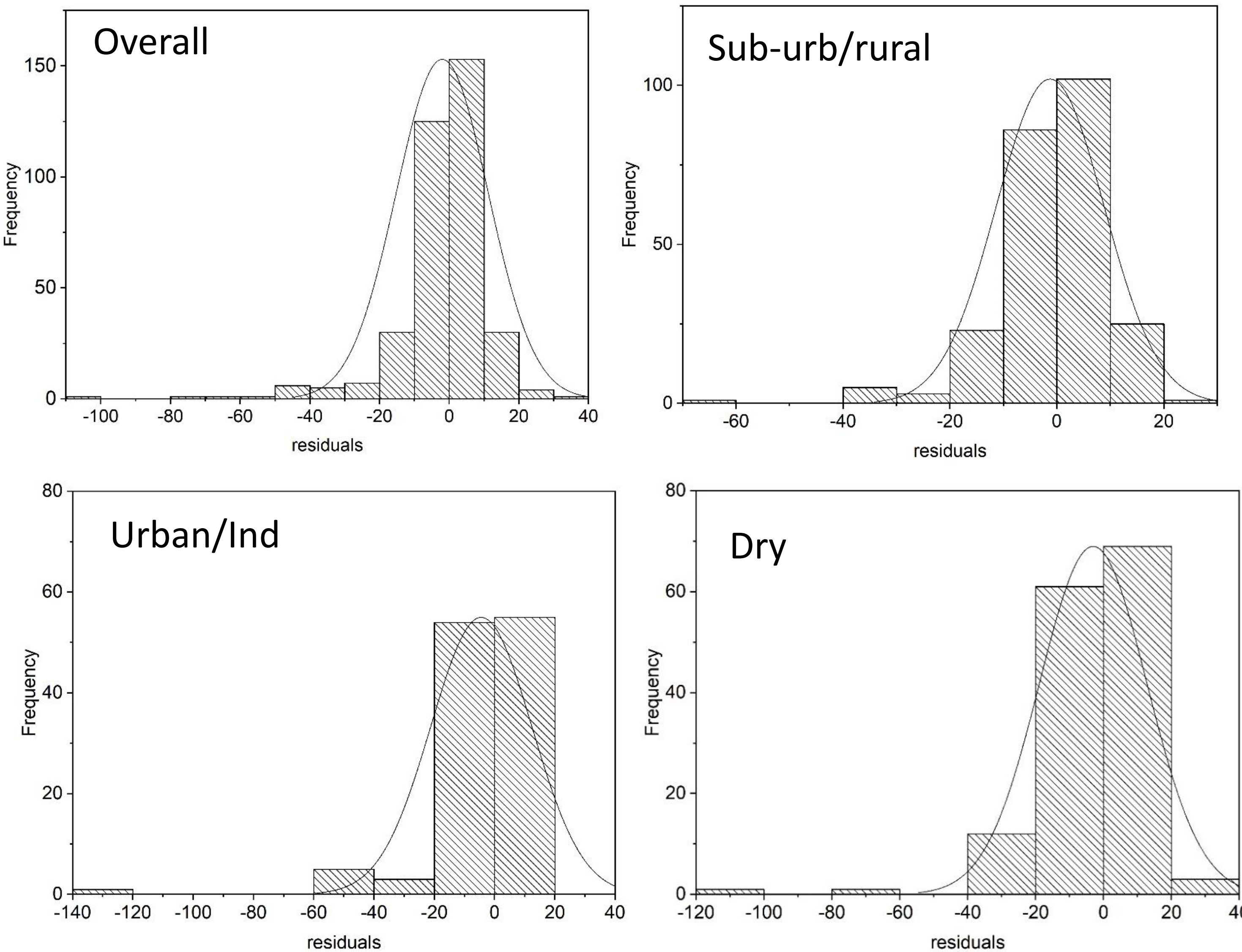




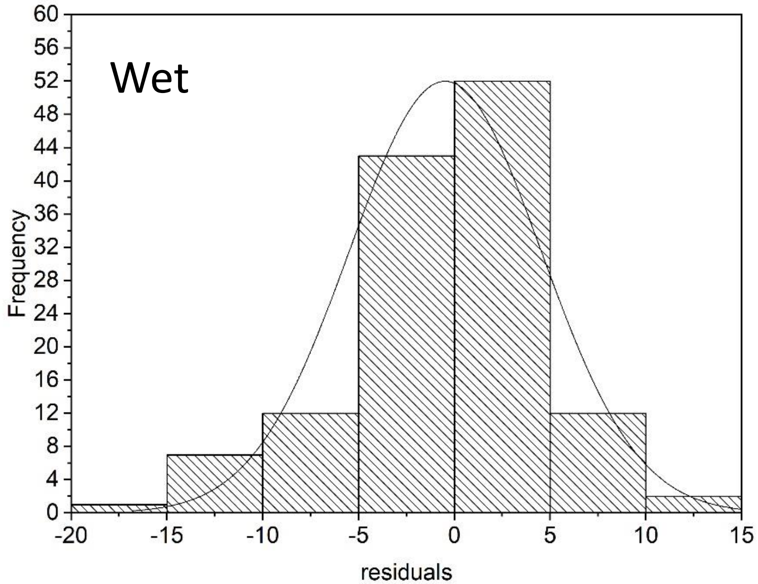


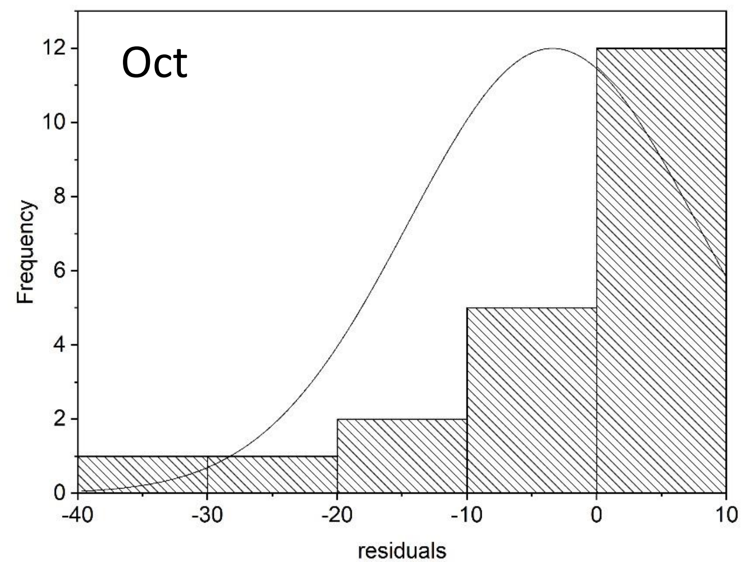

R2 = 0.11; RMSE=9.85 µg m-3
N=56; NSE= -0.714; MBE= 0.095
Predicted PM2.5(µg m-3)
Ground PM2.5(µg m-3)
Frequency
Wet
R2 = 0.15; RMSE=7.24 µg m-3
N=31; NSE= -0.014; MBE= -2.843
Predicted PM2.5(µg m-3)
Ground PM2.5(µg m-3)
Frequency
Apr-May
R2 = 0.38; RMSE=8.88 µg m-3
N=9; NSE= 0.389; MBE= 0.084
Predicted PM2.5(µg m-3)
Ground PM2.5(µg m-3)
Frequency
Oct

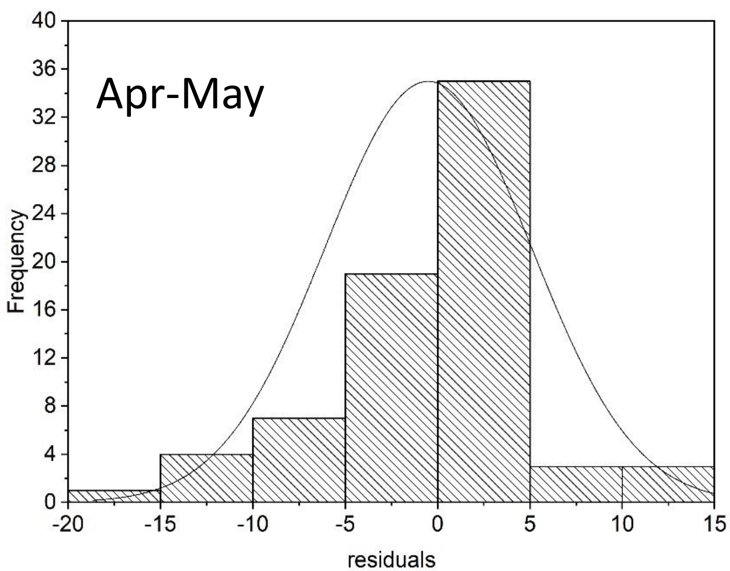
RESIDUAL ANALYSIS



- Estimated PM2.5 minus measured PM2.5
- Frequency of residuals approached 0 (normal) distribution
- Slightly shifted towards negative values







- The frequency of residuals is normally distributed peaking around zero.
- The predictions are quite accurate, indicating that the used machine learning techniques is satisfactory for $PM_{2.5}$ estimations in Malaysia.

CONCLUSION

- New machine-learning model was developed to estimate PM2.5 concentrations across Malaysia for the first time covering the years 2018 and 2019.

- CO was the most influential predictor variable for PM2.5 estimations followed by AOD, and other gases

- In future, with more matching data (PM2.5, AOD, NO2, SO2, CO and O3) more complex machine learning such as Deep Learning can be used to produce better estimation