Estimates of $PM_{2.5}$ Using Land-use Regression with Machine Learning and Micro-environmental Models: Comparison and Validation

Ching-Yi Mou[1], Wei-Ting Hsu[1], Chun-Hung Ku[1], Chih-Da Wu[1,2], Yu-Cheng Chen[1*]

[1] National Institute of Environmental Health Science, National Health Research Institutes, Miaoli County, Taiwan

[2]Department of Geomatics, National Cheng Kung University, Tainan City, Taiwan

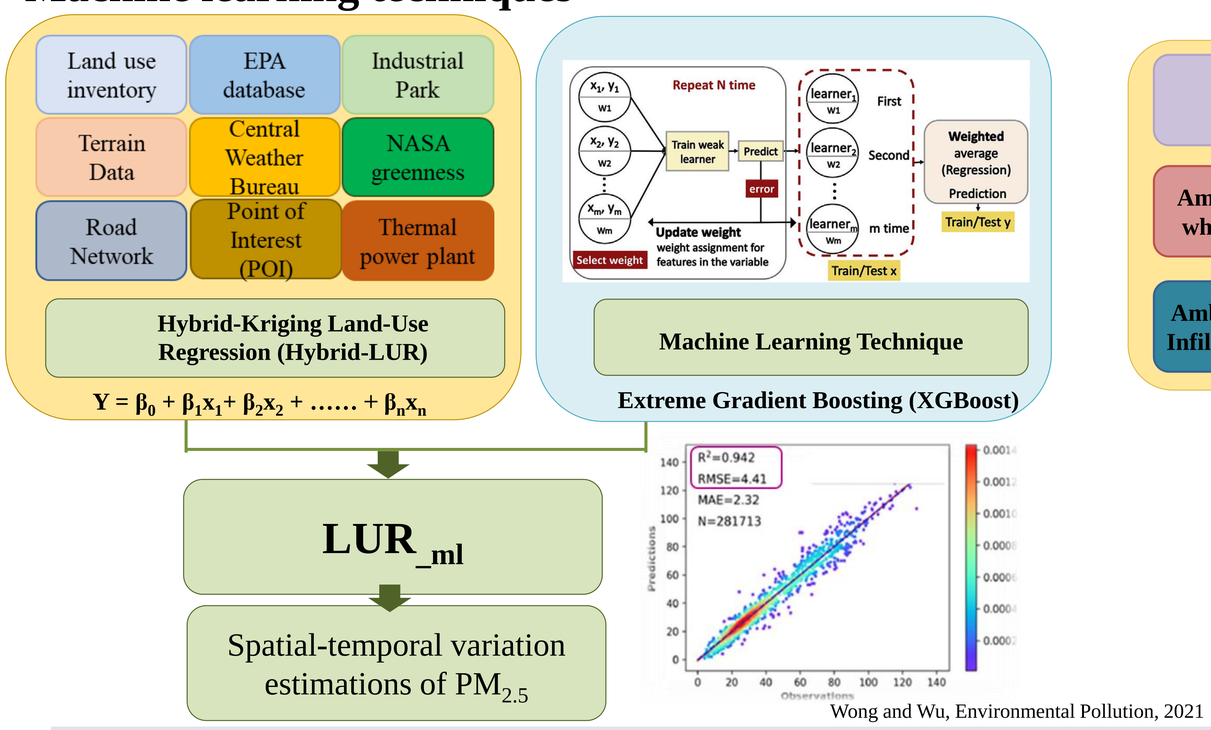
Introduction

The Land use regression model with Kriging incorporating the machine learning algorithm (LUR_ml) and micro-environmental exposure (ME) models are often used to predict the average exposure of $PM_{2.5}$ depending on study design. Little is known the comparison of prediction performance between LUR_ml and ME models for daily exposure to $PM_{2.5}$ across regions and seasons. The agreements of LUR_ml and ME models with measured personal exposure of $PM_{2.5}$ were assessed.

Methods

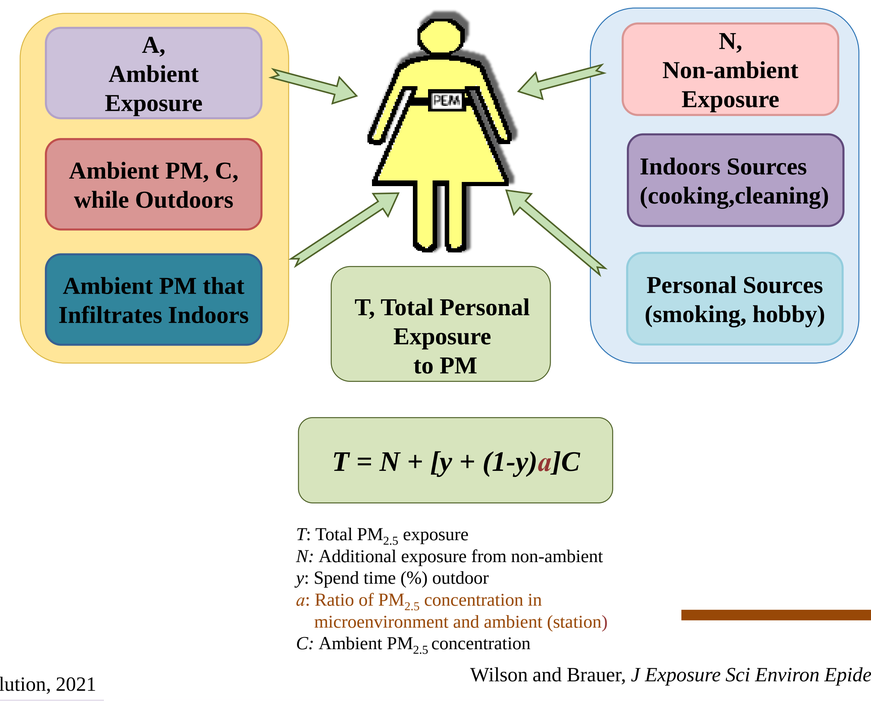
Predictions

Model 1 ：LUR integrated with Kriging and Machine learning techniques



Wong and Wu, Environmental Pollution, 2021

Model 2 ：Micro-environmental Models



T: Total $PM_{2.5}$ exposure
N: Additional exposure from non-ambient
y: Spend time (%) outdoor
a: Ratio of $PM_{2.5}$ concentration in microenvironment and ambient (station)
C: Ambient $PM_{2.5}$ concentration

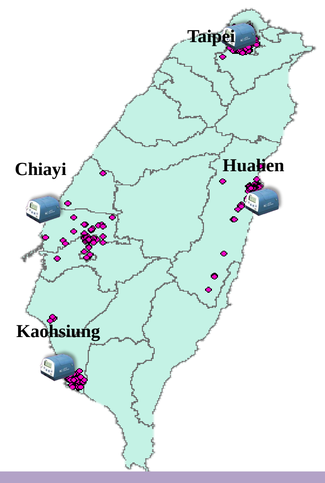
Wilson and Brauer, *J Exposure Sci Environ Epidemiol* 2006

M/S ratio in various type of indoor, outdoor and transit microenvironments

Microenvironments		M/S ratio (median)	
(Hsinchu & Miaoli county)EARUREMEN		Summer	Winter
Indoor	Home	0.56	0.89
	Office	0.69	0.63
	Temple	1.24	2.59
	Restaurant	0.68	1.54
	Hospital	0.88	0.79
	Barbershop	0.46	2.06
	Other public indoor places (bank, convenient store)	0.68	0.76
	Other indoor place (office, bank, etc.)	0.69	0.71
Outdoor	Traditional market	0.80	1.33
	Park	0.76	1.23
	Outdoor playground	0.56	1.04
	Night market	1.35	1.81
	Other outdoor places	0.91	1.75
Transit	Walking on the road	0.90	1.95
	Scooter	0.62	1.14
	Car	0.31	0.94
	Bus	0.51	-
	Train /high speed rail	0.52	0.60

Hsu and Chen, *Environmental Pollution*, 2020

Measurements



24hr SidePak sampling + Questionnaires

- Personal exposure
- Daily time-activity patterns
- Address

- Home indoor
- Home outdoor

- Ambient air

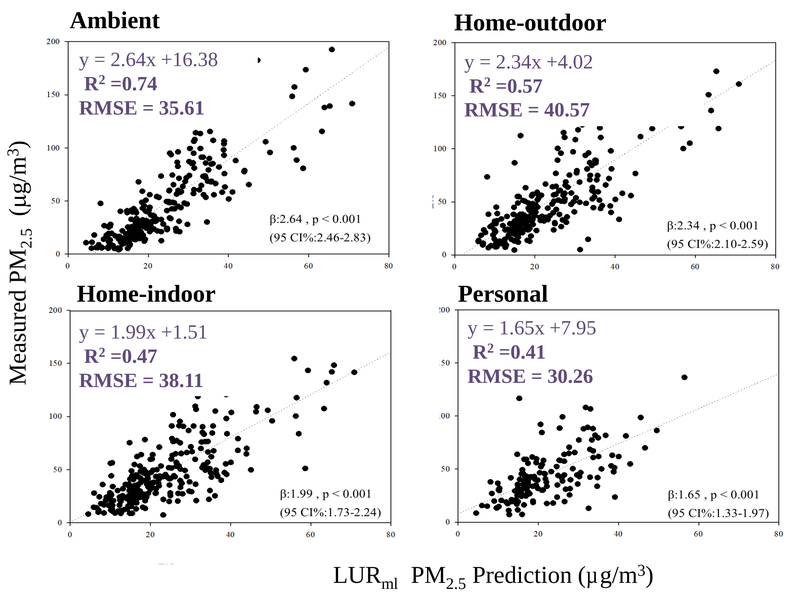
Validation and Comparison

- **Linear regression analysis, explained variance (R^2)**
- **Root-Mean-Square Error, RMSE**

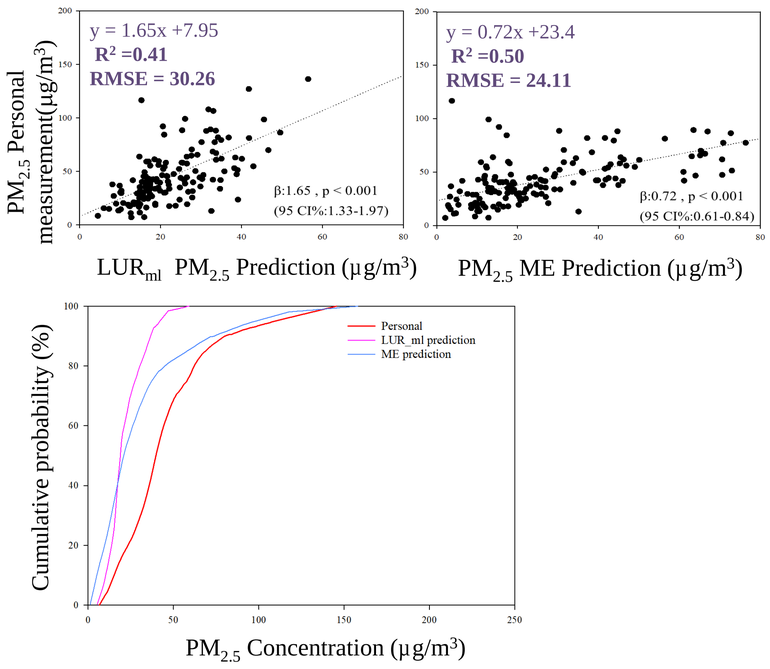
Measurement		Prediction
• Ambient • Outdoor • Indoor • Personal	Validation →	• $LUR_{\_ml}$
• Personal	Validation & comparison →	• $LUR_{\_ml}$ • ME model

Results

$LUR_{\_ml}$ Validation



$LUR_{\_ml}$ & ME Comparison



Conclusion

While most studies estimate outdoor $PM_{2.5}$ levels using spatial prediction models, the ME model takes time-activity patterns and micro-environmental information into account to generate overall exposure. Combining the LUR and ME model may be an advanced approach to improve accuracy in exposure assessment in the future.

2021 Conference on Fine Particulate Matter (PM2.5), Air Quality and COVID-19 Issues (Oral presentation)