Prediction of Air Pollutants Concentration Based on an Extreme Learning Machine: The Case of Hong Kong

**Introduction**

* Air pollutants in HK emitted largely by industry and automobiles
* Prediction helps with understanding of air pollution and development of pollution control strategies
* Two methods: deterministic and stochastic
  + Deterministic: models physical and chemical transportation process in terms of meteorological variables (can do either short-term or long term)
  + Some researchers use these to develop integrated air quality model with source, dispersion, and environmental impacts
* Statistical approach learns from historical data, methods include time series analysis, Bayesian filter, ANN, etc.
* Linear regression models not ideal due to reactions between air pollutants and influential factors being nonlinear
* ANN has advantages of incorporating nonlinear relationships between pollutants and meteorological factors, but have drawbacks including:
  + Poor generalization
  + Time consuming
* Used extreme learning machine (ELM), performs better when there is more noise in data
* Feed-forward neural networks based on back propagation (FFANN-BP) is most popular but requires “desired output”
* Need activation function (unit step, linear, sigmoid, relu, etc.)
* Drawbacks include extremely time consuming (all parameters tuned iteratively), prone to get caught in local minimum
* ELM: two-layer network; hidden layer fixed and random, output layer trained
  + Has been widely used from biomedical engineering to computer vision
  + Used for classification, regression, clustering, feature selection, etc

**Methods**

* Inputs included high temp, low temp, diff between hi and lo temp, av. Temp, wind speed, wind direction (radians), rel. humidity, and time variables: day of week, month of year
* Used 10-fold cross validation, average accuracy for 10 iterations recorded, number of hidden nodes was 20

**Results**

* Evaluation metrics: mean absolute error (MAE), root mean square error (RMSE), index of agreement (IA) and R2 value
  + Results w/highest R2 and lowest RMSE is best method
* ELM scored well over linear regression and FFANN-BP for coefficient of determination, yielded lowest RMSE, fastest speed
* Ran again but with a limit value of 50 ug/m3, exceeded this value on 38% of days

**Conclusion**

* ELM performs better in terms of precision, robustness, and generalization, however, no significant differences between the prediction accuracies of each model
* Provided best performance on indicators such as R2 and RMSE

**Citation**

Zhang, J. & Ding, W. (2017). Prediction of Air Pollutants Concentration Based on an Extreme Learning Machine: The Case of Hong Kong. *International Journal of Environmental Research and Public Health*. https://doi.org/10.3390/ijerph14020114

**Purpose**

I will be using this research on my background research paper, as well as to supplement the background information section of my final research paper. Also, I may consider using ELM instead of a traditional ANN, if time permits.