Published in:

***Science of the Total Environment***

Title:

**Socio-economic development drives solid waste management performance in cities: A global analysis using machine learning**

Authors:

**Costas A. Velis, David C. Wilson, Yoni Gavish, Sue M. Grimes, Andrew Whiteman**

**Data files used to perform the analyses:**

1. *VelisEtAl2023\_WABI\_Input.xlsx*   
   Excel file with 4 sheets:
   1. Info – Additional info on the file.
   2. Meta – Metadata info used by R codes.
   3. Data – Full data set for the WABI, used by the R codes.
   4. Map – Data required for plotting the map in Figure 1.

Available at: <https://zenodo.org/record/7570174#.Y9eeYhfP03s>

DOI: 10.5281/zenodo.7570174

1. *Fig1\_Wabi\_Horizontal.jpg* ; *Fig1\_Wabi\_Vertical.jpg*   
   A JPEG image (300dpi) with the WABI side panel, that will appear on the right of (Horizontal), or below (Vertical) the map in Figure 1.
2. *VelisEtAl2023\_WABI\_Summary.xlsx*   
   Summary Excel file with 2 sheets:
   1. Metadata – Description of the variables.
   2. Summary – Summary of the results of the non-linear analyses.

Available at: <https://zenodo.org/record/7570421#.Y9ee9BfP03s>

DOI: 10.5281/zenodo.7570421

**R scripts used to perform the analyses:**

1. Descriptive
   1. *Get\_Quantiles\_Dep\_IncomeCategory.R*   
      Get the 25th, 50th (median) and 75th quantile of each WABI for all cities together and for each income category ("L", "L-M", U-M", "H") (Table S2 in SI.4).
2. Approach 1 - Non-linear regression analysis:
   1. *FitMod.R*   
      Function to fit a non-linear model against one explanatory variable. Can fit up to 9 different models with minpack.lm::nlsLM. Estimates prediction and confidence intervals with propagate::predcitNLS.Approach
   2. *FitModAllDepAllExp.R*   
      For each combination of dependent variable, explanatory variable and formula, fit a nls model, get predicted values and confidence / prediction intervals. Then compare models with AICc based model selection.
   3. *Add\_ModelSelection\_Options.R*   
      Run the two steps model selection analysis for each WABI. In the first step, the best formula for each explanatory variable is selected, In the second step the 9 best models, one for each explanatory variable, are compared.
3. Approach 2 - Conditional random-forest:
   1. *Fit\_CondRandomForest\_VarImp.R*  
      Fit conditional random-forest for each WABI against the 9 explanatory variables. Extract the regular and conditional variable importance as well as the RMSE. Add the RMSE of the best non-linear model for comparison.
4. Plotting:
   1. *Plot\_WABI\_Panel.R*   
      Plot for a selected WABI a traffic light panel with the colored bands. The panels are used in the WABI panel of figure 1.
   2. *Plot\_WABI\_Histograms.R*Plot a histogram for each WABI (Figure S1 in SI).
   3. *Plot\_Fig1\_Map\_Wabi.R*   
      Plot a world map with points for each city. Around the map add for each city a small WABI indent with the traffic light categories. Add a panel explaining the WABIs color scheme either to the right (Horizontal) or below (Vertical) the map (Figure 1).
   4. *Plot\_Fig2\_ModelSelection\_RF.R*   
      Plot the AICc weights (non-linear regressions) and the variable importance (conditional random forest) of each explanatory variable in each WABI (Figure 2).
   5. *Plot\_Fig3\_WasteGeneration.R*   
      Plot the top 3 non-linear models for the waste generation WABI with the traffic light bands as background. Add a boxplot based on income category (Figure 3).
   6. *Plot\_Fig4\_WasteCollection.R*  
      Plot the best model for the waste collection coverage and quality variables, a boxplot based on income category and the trade-off between coverage and quality (Figure 4).
   7. *Plot\_Fig5\_WasteTreatment.R*  
      Plot the best model for the waste Treatment coverage and quality variables, a boxplot based on income category and the trade-off between coverage and quality.