The Forecasting Waste Generation Model based on Linked Open Data and the DPSIR Framework.
Case study concerning municipal waste in the Czech Republic.

Jiří Hřebíček¹, Jiří Kalina¹, Jana Soukopová¹, Jan Prášek² and Jiří Valta²
¹Masaryk University, Brno, Czech Republic
²Czech Environmental Information Agency, Praha, Czech Republic
Development of forecasting model of waste stream generation includes the following consequent modelling steps:

1. **Identification of the required waste streams** using waste codes of the European List of Waste (ELW) and computation formulas for their amounts.
2. Processing of the historical annual waste streams generation and treatment reports (2009–2013) provided by waste generators and facilities and creating of their data sets.
3. Identification and development of socioeconomic and demographic predictors based on the DPSIR framework (which have influence on waste streams generation) using linked open government data (eGovernment systems) of the Czech Republic.
5. **Forecasting of predictors from the DPSIR framework and calculation of waste stream forecasts.**
6. Processing of sensitivity analyses of predictors of waste stream generation models and scenarios for decision makers.
## Identification of 16 required waste streams

<table>
<thead>
<tr>
<th>Waste stream number</th>
<th>Waste stream</th>
<th>ELW waste codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All wastes</td>
<td>010101-200399</td>
</tr>
<tr>
<td>2</td>
<td>All waste of other category</td>
<td>Codes from 010101-200399 without *</td>
</tr>
<tr>
<td>3</td>
<td>All waste of hazardous category</td>
<td>Codes from 010101-200399 with *</td>
</tr>
<tr>
<td>5</td>
<td>Mixed municipal waste</td>
<td>200301</td>
</tr>
</tbody>
</table>
Processing of the historical annual waste streams generation and treatment

- ISOH - Waste management information system

- ISOH contains data concerning waste generation and treatment by generators and data concerning facilities to treat, recover and dispose of waste. Every year, it records more than 70,000 different generators’ reports in all 6,500 municipalities of the Czech Republic and more than 3,000 facilities’ reports. The annual ISOH database contains more than 50,000 records of municipal waste generation and 10,000 records concerning their treatment.

- The database was available to us in 2009-2014 and we calculated waste streams from previous Table for these years.
Linked open government data of the Czech Republic for models

• eGovernment systems of the Czech Republic provide sources of necessary input data for the DPSIR framework predictors for different waste streams.
• The related linked open data for the DPSIR framework predictors are often available on web sites of Ministry of Environment MoE, CENIA, the Ministry of Finance (MoF), the Ministry of Regional Development (MoRD), and the Czech Statistical Office (CZSO).
• For example, the MoF provides a specialized web information portal MONITOR that allows open public access to budget and accounting information from all public authority levels including every municipality in the Czech Republic.
Main driving forces for MSW

- **Population.** Development of the population, together with relocation of residents with higher purchasing power to cities and agglomerations, also reduce its own waste treatment options (e.g. composting) and create demand for faster replacement of goods, which affects household consumption.
- The number of *pensioners* and the level of *unemployment* as families with small children, some students, pensioners and the unemployed remain near their residence throughout the day where their activities generate waste. Workers and children in kindergartens and schools and some students carry out their daily activities at the place of employment or school where they generate MSW, etc.
- **Consumer behaviour,** including ways of packaging, driven by consumer demand and legal regulations, e.g. hygienic and health protection requirements.
- **Municipal costs of MW** of citizens. We are able to download driving forces data from linked open government data of the Czech Republic.
# State: generation of MSW of the Czech Republic

<table>
<thead>
<tr>
<th>Year</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>amount [tonnes]</td>
<td>5 325 179</td>
<td>5 361 883</td>
<td>5 388 058</td>
<td>5 192 784</td>
<td>5 167 805</td>
<td>5 323 947</td>
</tr>
</tbody>
</table>
Factors in the DPSIR framework for CDW

**Driving Force**
- Prevention
  - Employment in the construction sector
  - Public and EU funding
  - Volume of construction work

**Pressure**
- Residual materials in waste regime
- Waste treatment installations capacity
- Transportation
- Waste management plan indicators
- Integrated pollution register
- Air quality
- Water quality

**State**

**Impact**
- Noise
- Dustiness
- Threats to water
- Reducing retention capacity of the landscape
- Landscape

**Response**
- Legislation
- Spatial restrictions
- Economic instruments
- Certification conditions
Construction and demolition waste (CDW) is mostly generated during major infrastructure projects, especially in construction of highways and urban areas or demolition of large industrial complexes. Driving forces are:

- **Amount of construction work** is essential. In the framework of linear constructions (roads, railways), significant quantities of bulk waste are produced.

- **Prevention** of CDW generation plays an important role. These constructions are usually organized on the basis of public and EU funding.

- Other drivers of CDW production are the number of people employed in construction and prevention programs also have an indispensable role.
# State: generation of CDW of the Czech Republic

<table>
<thead>
<tr>
<th>Year</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>amount [tonnes]</td>
<td>18 520 614</td>
<td>18 480 355</td>
<td>17 387 158</td>
<td>17 318 625</td>
<td>17 904 590</td>
<td>19 124 592</td>
</tr>
</tbody>
</table>
Multi-linear regression model of waste streams generation

Let us assume that for every waste stream $f = 1, \ldots, 16$ (see Table 1), the amount of waste $\hat{w}_{tf}$ and predictors $\hat{A}_{i,tf}, i=1, \ldots, K^f$; in years $t=2009, \ldots, 2014$ are known, where $K^f$ is the number of predictors for the waste stream $f$. Let waste stream generation $w_{f}(t)$ at the given year $t$ fulfils the equation

$$
\log(w_{f}(t)) = a_{0_f} + \sum_{i=1}^{K^f}(a_{i_f} \log(A_{i,f}(t))) + \epsilon_{tf}
$$

where

$A_{i,f}(t), i=1, \ldots, K^f$ are predictors in the given year $t$ derived from the DPSIR analysis of the waste stream $f$.

$\epsilon_{tf} = \log(w_{f}(t)) - \log(\hat{w}_{tf})$, for $t=2009, \ldots, 2014$ are approximation errors.

Coefficients $a_{0_f}, \ldots, a_{K^f}$ in (1) for each waste stream $f$ are calculated using multiple regression on the basis of the values of waste generation $\hat{w}_{tf}$ and predictors $\hat{A}_{i,tf}, i=1, \ldots, K^f; t=2009, \ldots, 2014$. 
Multi-linear regression model of waste streams generation

- Approximation errors $\epsilon_{f,t}=2009,\ldots,2014$, have the mean equal to 0 and the normal distribution.

- If we want to establish the confidence interval of predictors $A_{i,f,t}$, it will be necessary to restrict their number to $K_f \leq 4$, since we only have a time series of six past known values. If we have the values of $\hat{w}_{f,t}$ and $\hat{A}_{i,f,t}$ for the next years $t=2015,\ldots$, the model (1) will be more accurate and an approximation error $\epsilon_{f,p,s}$ will be smaller.

- Furthermore, we assume that the predictors $A_{i,f}(t), i=1,\ldots,K_f$, for $t=2015,\ldots,2024$ have either known values (GDP, population, household consumption, etc.) from the eGovernment systems or are determined by an appropriate extrapolation method or are chosen by decision makers.
The sensitivity of estimate of waste generation

The *local sensitivity* of the model (1) for the $i^{th}$ predictor $Aif(t)$, in the year $t$ and waste stream $f$ can be estimated using the partial derivatives of the waste stream $wf(t)$ in according to the $i^{th}$ predictor $Aif(t)$:

\[ aif \cdot \frac{wf(t)}{Aif(t)}. \]  

(2)

It follows from (2) that if the value of predictor $Aif(t)$ is increased by 1 percent then the amount of waste $wf(t)$ in waste stream $f$ will increase or decrease by $aif$ percent, if $aif > 0$ or $aif < 0$, for $i=1,\ldots,K^f$.

This knowledge is important for users of the model. We continue in the further analysis of the developed model, i.e. assessment of the statistical significance of predictors.
Statistical significance of predictors

We use statistical software where it is more common to calculate the test p-value, which we denote $p_i$. This is the smallest level of the F-test in which we would reject the hypothesis $H_0: \{s^2 = s_i^2\}$. We set this value as $p_i = 1 - H(F_i)$. This procedure is repeated gradually for other predictors $A_i(t)$, for which we calculate $F_i$ statistics and the test p-values $p_i$, $i = 2, ..., K_f$.

Let us choose the level of significance $\alpha$ (values 0.05 or 0.1 are usually selected) of the predictors. We calculate p-values $p_i, i = 1, ..., K_f$ and compare them with this level of significance $\alpha$:

- If $p_i > \alpha$ => the null hypothesis $H_0: s^2 = s_i^2$ is rejected. Conclusion: the variances of different models are statistically significant and the $i$th predictor $A_i(t)$ is significant.
- If $p_i < \alpha$ => we cannot reject the hypothesis $H_0$. Conclusion: the variances of both models are not statistically significantly different (i.e., the selections originated from the same basic model with the common variance $s^2$) and the $i$th predictor is not significant.
Extrapolation of the predictors

For the calculation of the forecast waste stream generation \( w_f(t) \) in the years \( t=2015,\ldots,2024 \) it is necessary to know the values of predictors \( A_{if}(t), i=1,\ldots, K_f, t=2015,\ldots,2024 \). These values, however, may not always be listed in the sources (linked open data in the eGovernment systems) from which we draw the data predictors \( \hat{A}_{itf}, i=1,\ldots,K_f; t=2009,\ldots,2014 \). In this case, the procedure is the following:

• Enter the values of the predictors based on expert’s estimates or other appropriate sources;
• On the basis of the values of the predictors \( \hat{A}_{itf}, i=1,\ldots,K_f; t=2009,\ldots,2014 \) the values of predictors \( A_{if}(t), i=1,\ldots, K_f, t=2015,\ldots,2024 \) are calculated using either linear or exponential extrapolation.
Municipal solid waste forecast generation analysis

1. In the first step, in Figure, there are boxes of basic data inputs over the time $t=2009,\ldots,2014$ for modelling: previous MSW generation and time series of the known values of the relevant DPSIR analysis predictors (implicit values are parsed from linked open data) and forecasting model outputs.

2. In the second step, Figure, users can specify the values of the predictors (pre-filled values are available with hyperlinks to the linked open data sources) expected in the model (1) or choose their possible linear/exponential extrapolation. Users can also input expected prevention measures (three possible scenarios of MSW prevention are available).

3. In the third step, Figure, results are shown in a form of a table, mathematical expression of (1) and a time-plot, showing the future MSW generation development and effects of the prevention measures taken (the curves representing the prediction interval and shaded areas showing the confidence intervals).

4. In the fourth step, Figure, a sensitivity analysis is presented, showing decision makers an estimated effect of the individual predictors from the model (1) and the quality of forecasting.
Municipal solid waste forecast generation analysis

**MSW generation prediction**

**Step 1: Input data for MSW prediction model**

- **Initial year:** (first year in which the generation will be modelled).
- **Statistical significance:** (probability in % of not being wrong when estimating significant predictors).
- **Previous MSW generation:** (values in tons/year with decimal dot separator; split by comma separator, NA for not available value).
- **Previous household expenditures for food, clothes and shoes:** (values in billions (10^9) CZK, with decimal dot separator; split by comma separator, NA for not available value).
- **Previous mid-year population:** (values in thousand of inhabitants, with decimal dot separator; split by comma separator, NA for not available value).
- **Previous number of retired:** (values in thousand of retired persons, with decimal dot separator; split by comma separator, NA for not available value).
- **Previous unemployment:** (values in %, with decimal dot separator; split by comma separator, NA for not available value).

<table>
<thead>
<tr>
<th>Initial year</th>
<th>Statistical significance</th>
<th>Previous MSW generation</th>
<th>Previous household expenditures for food, clothes and shoes</th>
<th>Previous mid-year population</th>
<th>Previous number of retired</th>
<th>Previous unemployment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>95</td>
<td>5325179,5361893,5396083,5192784,5167805,5323947</td>
<td>501,546,598,129,533,015,550,119,565,34,585,939</td>
<td>10491,10617,10497,10609,10510,10526</td>
<td>2790,2881,2873,2866,2858,2863</td>
<td>6,7,7,3,6,7,7,6,1</td>
</tr>
</tbody>
</table>

[Insert values into the model]
# Municipal solid waste forecast generation analysis

## Step 2: Expectations of future development 2024

|---------------------------------------------------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|


Unemployment [%] | 2015: 6.6 | 2016: 5.9 | 2017: 5.6 | 2018: 5.8 | 2019: 5.8 | 2020: 5.8 | 2021: 5.8 | 2022: 5.8 | 2023: 5.8 | 2024: 5.8 |


Source: [https://www.czso.cz/cs/otbyvaleistvo.hu](https://www.czso.cz/cs/otbyvaleistvo.hu)


Municipal solid waste forecast generation analysis

Step 3: Final prediction and model equation

- Show plot legend: none
- Show confidence interval: yes
- left down
- right down
- left up
- right up

Waste generated (t)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Modelled values</td>
<td>5292619</td>
<td>5292619</td>
<td>5292619</td>
<td>5292619</td>
<td>5292619</td>
<td>5292619</td>
<td>5292619</td>
<td>5292619</td>
<td>5292619</td>
<td>5292619</td>
</tr>
<tr>
<td>Lower threshold</td>
<td>5044674</td>
<td>5044674</td>
<td>5044674</td>
<td>5044674</td>
<td>5044674</td>
<td>5044674</td>
<td>5044674</td>
<td>5044674</td>
<td>5044674</td>
<td>5044674</td>
</tr>
<tr>
<td>Upper threshold</td>
<td>5552761</td>
<td>5552761</td>
<td>5552761</td>
<td>5552761</td>
<td>5552761</td>
<td>5552761</td>
<td>5552761</td>
<td>5552761</td>
<td>5552761</td>
<td>5552761</td>
</tr>
<tr>
<td>Linear extrapolation</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
</tr>
<tr>
<td>Lower threshold</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
</tr>
<tr>
<td>Upper threshold</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
<td>5325179</td>
</tr>
</tbody>
</table>

generation = 50331795.5 - 22990.5 year

Download table

- generation = 235.3579, household = -0.4101, population = 0.3962, retired = 1.2023, unemployment = -0.3317
Municipal solid waste forecast generation analysis

Step 4: Sensitivity analysis

Show plot legend:
- none
- left down
- right down
- left up
- right up

Show confidence interval:
- yes
- no

Table of model outputs (in logarithmic form):

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Degrees of freedom</th>
<th>Sum of squared residuals</th>
<th>F statistics</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable: household</td>
<td>1</td>
<td>0.00324</td>
<td>6.3</td>
<td>0.241</td>
</tr>
<tr>
<td>Variable: population</td>
<td>1</td>
<td>5.3e-05</td>
<td>1</td>
<td>0.494</td>
</tr>
<tr>
<td>Variable: retired</td>
<td>1</td>
<td>2.6e-05</td>
<td>0.5</td>
<td>0.606</td>
</tr>
<tr>
<td>Variable: unemployment</td>
<td>1</td>
<td>0.001039</td>
<td>20.3</td>
<td>0.139</td>
</tr>
<tr>
<td>Residuals</td>
<td>1</td>
<td>5.1e-05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

- The predictors of the forecasting models (1) were chosen through a selection process that included opinions from experts, literature review based on relevance and applicability to different waste streams settings.
- Appropriate predictors were selected and fitted into the DPSIR framework, which is presented for MSW and CDW.
- The construction of the forecasting models consisted of construction and definition of the predictors based on the DPSIR framework, integration of the predictors into the forecasting model and analysis of sensitivity of the predictors.
- The forecasting model was implemented as open source software and it was verified using appropriate data. The outputs of the developed forecasting model are presented for MSW of the Czech Republic with scenarios to follow the EU action plan for the circular economy.
Thank you for your attention

Questions?

Prof. Dr. Jiří Hřebíček
Institute of Biostatistics and Analyses
Masaryk University
Brno, Czech Republic

hrebicek@iba.muni.cz