The Forecasting Waste Generation Model based on Linked Open Data and the DPSIR Framework.

Case study concerning municipal waste in the Czech Republic.

Jiří Hřebíček[1], Jiří Kalina[1], Jana Soukopová[1], Jan Prášek[2] and Jiří Valta[2]

[1]Masaryk University, Brno, Czech Republic

[2]Czech Environmental Information Agency, Praha, Czech Republic

Introduction

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.
4. Construction of a *multi-linear regression model of waste streams generation* with predictors from the DPSIR framework analyses.
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

Waste stream number	Waste stream	ELW waste codes
1	All wastes	010101-200399
2	All waste of other category	Codes from 010101-200399 without *
3	All waste of hazardous category	Codes from 010101-200399 with *
4	Municipal solid waste	200101, 200102, 200108, 200110, 200111, 200113*, 200114*, 200115*, 200117*, 200119*, 200121*, 200123*, 200125, 200126*, 200127*, 200128, 200129*, 200130, 200131*, 200132*, 200133*, 200134, 200135*, 200136, 200137*, 200138, 200139, 200140, 200141, 200199, 200201, 200202, 200203, 200301, 200302, 200303, 200306, 200307, 200399, 150101, 150102, 150103, 150104, 150105, 150106, 150107, 150109, 150110*, 150111*
5	Mixed municipal waste	200301

Processing of the historical annual waste streams generation and treatment

- ISOH - Waste management information system

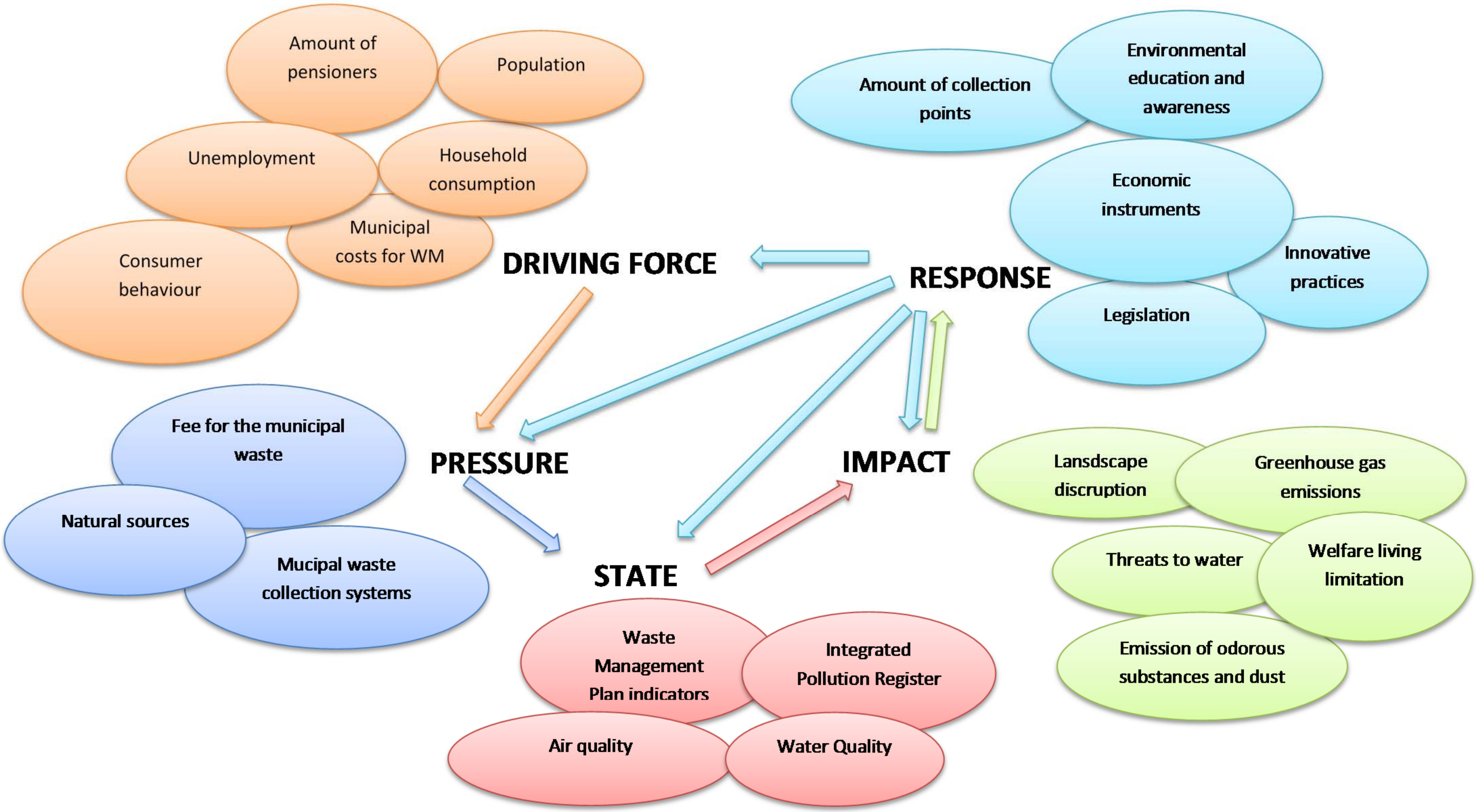
- ISOH contains data concerning waste generation and treatment by generators and data concerning facilities to treat, recover and dispose of waste. Every year, it records more than 70,000 different generators´ reports in all 6,500 municipalities of the Czech Republic and more than 3,000 facilities´ reports.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

Factors in the DPSIR framework for MSW



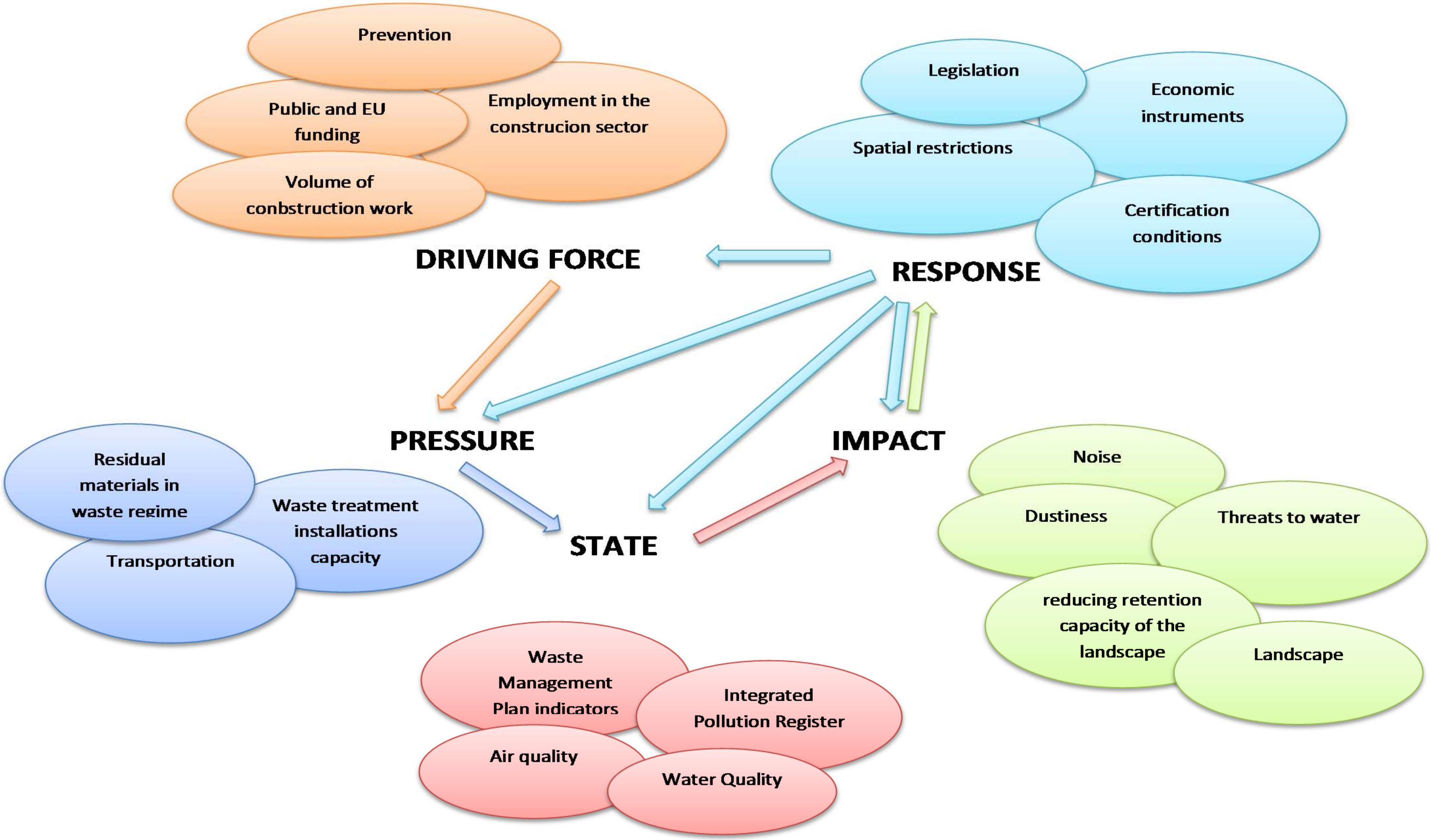
Main driving forces for MSW

- *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

Year	2009	2010	2011	2012	2013	2014
amount [tonnes]	5 325 179	5 361 883	5 388 058	5 192 784	5 167 805	5 323 947

Factors in the DPSIR framework for CDW



Main driving forces for CDW

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

Year	2009	2010	2011	2012	2013	2014
amount [tonnes]	18 520 614	18 480 355	17 387 158	17 318 625	17 904 590	19 124 592

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 $wf(t)$ at the given year t fulfils the equation

$$log(wf(t)) = a_{0f} + \sum_{i=1}^{Kf}\left(a_{if}\log(Aif(t))\right) + \varepsilon tf \qquad (1)$$

where

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

$\varepsilon tf = \log(wf(t) - \log(\hat{w}tf)$, 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 εtft=2009,…,2014, have the mean equal to 0 and the normal distribution.

- If we want to establish the confidence interval of predictors *Aift*, 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 ŵ*tf* and *Âi,tf* for the next years *t*=2015,…, the model (1) will be more accurate and an approximation error *εtf,ps* will be smaller.

- Furthermore, we assume that the predictors *Aif(t)*, *i*=1,…,K^f, for *t*=2015,…,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 $Aif(t)$:

$$aif \cdot wf(t)/Aif(t) \quad . \qquad (2)$$

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,…,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 pi. This is the smallest level of the F-test in which we would reject the hypothesis H0: $\{s^2 = si^2\}$. We set this value as pi = 1-H (Fi). This procedure is repeated gradually for other predictors Aif(t), for which we calculate Fi statistics and the test p-values pi, i = 2, ..., Kf.

Let us choose the level of significance α (values 0.05 or 0.1 are usually selected) of the predictors. We calculate p-values pi ,i = 1,…,Kf and compare them with this level of significance α:

- If pi>α => the null hypothesis H0: $s^2= si^2$ is rejected. Conclusion: the variances of different models are statistically significant and the ith predictor Aif(t) is significant.
- If pi<α => we cannot reject the hypothesis H0. 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 ith predictor is not significant.

.

Extrapolation of the predictors

For the calculation of the forecast waste stream generation wf(t) in the years t=2015,…,2024 it is necessary to know the values of predictors Aif(t), i=1,…, Kf , t=2015,…,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 Âitf , i=1,…,Kf; t=2009,…,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 Âitf , i=1,…,Kf; t=2009,…,2014 the values of predictors Aif(t),i=1,…, Kf , t=2015,…,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,…,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 prediciton

Step 1: Input data for MSW prediction model

Initial year:
(first year in which the generation will be modelled).

2015

Statistical significance:
(probability in % of not being wrong when estimating significant predictors).

95

Previous MSW generation:
(values in tons/year with decimal dot separator; split by comma separator, NA for not available value).

5325179,5361883,5388058,5192784,5167805,5323947

Previous household expenditures for food, clothes anfd shoes:
(values in bilions (10^9) CZK, with decimal dot separator; split by comma separator, NA for not available value).

501.546,508.128,533.015,550.119,565.34,585.939

Previous mid-year population:
(values in thousand of inhabitants, with decimal dot separator; split by comma separator, NA for not available value).

10491,10517,10497,10509,10510,10525

Previous number of retired:
(values in thousand of retired persons, with decimal dot separator; split by comma separator, NA for not available value).

2790,2881,2873,2866,2858,2863

Previous unemployment:
(values in %, with decimal dot separator; split by comma separator, NA for not available value).

6.7,7.3,6.7,7,7,6.1

Insert values into the model

Municipal solid waste forecast generation analysis

Step 2: Expectations of future development 2024

	2015:	2016:	2017:	2018:	2019:	2020:	2021:	2022:	2023:	2024:	
Household expenditures for food, clothes anfd shoes [bil. CZK]	589.9733	595.4104	600.8475	606.2846	611.7216	617.1550	622.5920	628.0291	633.4662	638.9033	Source: https://vdb.czso.cz/vdbvo2/faces/index.jsf?page=statistiky#katalog=30847
Mid-year population [thousand of inhabitants]	10538	10534	10531	10527	10524	10520	10507	10493	10480	10467	Source: https://www.czso.cz/csu/czso/obyvatelstvo_hu
Number of retired [thousand of inhabitants]	2884	2848	2843	2838	2833	2828	2823	2818	2813	2808	Source: http://www.cssz.cz/cz/o-cssz/informace/informacni-materialy/statisticke-rocenky.htm
Unemployment [%]	6.6	5.9	5.8	5.8	5.8	5.8	5.8	5.8	5.8	5.8	Source: https://vdb.czso.cz/vdbvo2/faces/cs/index.jsf?page=vystup-objekt&pvo=ZAM06&zo=N&z=T&f=TABULKA&verze=-1&nahled=N&sp=N&filtr=G~F_M~F_Z~F_R~F_P~_S~_null_null_&katalog=30853&str=v95&c=v3__RP2014
1st prevention measure (effect) [% of MSW generated]	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	
2nd prevention measure (effect) [% of MSW generated]	0	2	5	9	14	20	27	35	44	9	
3rd prevention measure (effect) [% of MSW generated]	0	3	6	9	12	15	18	21	24	27	

Insert values into the model

Municipal solid waste forecast generation analysis

Step 3: Final prediction and model equation

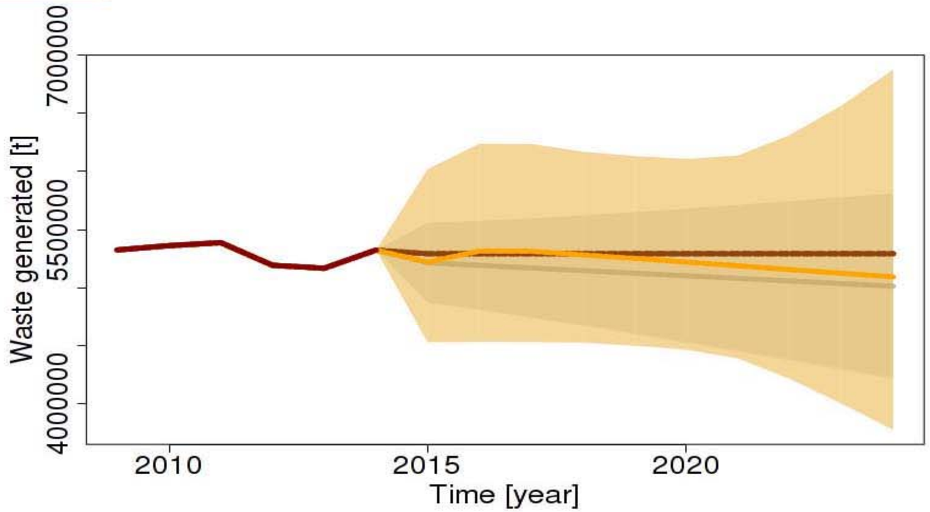
Show plot legend:

- ◉ none
- ○ left down
- ○ right down
- ○ left up
- ○ right up

Show confidence interval:

- ◉ yes
- ○ no

Refresh plot



generation = 50331795.5 + -22390.5 . year
generated = 235.5678 . household ^ -0.4101 . population ^ 0.3962 . retired ^ 1.2023 . unemployment ^ -0.3317

Download table

generation . 50331795.5 . -22390.5 . year
generation = 5292619.2353
Table of modelled values:

	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Modelled values	**5292619**	**5292619**	**5292619**	**5292619**	**5292619**	**5292619**	**5292619**	**5292619**	**5292619**	**5292619**
Lower threshold	5044674	5044674	5044674	5044674	5044674	5044674	5044674	5044674	5044674	5044674
Upper threshold	5552751	5552751	5552751	5552751	5552751	5552751	5552751	5552751	5552751	5552751
Linear extrapolation	**5325179**	**5361883**	**5388058**	**5192784**	**5167805**	**5323947**	**5214909**	**5192519**	**5170128**	**5147738**
Lower threshold	5325179	5361883	5388058	5192784	5167805	5323947	4871840	4809740	4742870	4672570
Upper threshold	5325179	5361883	5388058	5192784	5167805	5323947	5557978	5575298	5597386	5622906

Municipal solid waste forecast generation analysis

Step 4: Sensitivity analysis

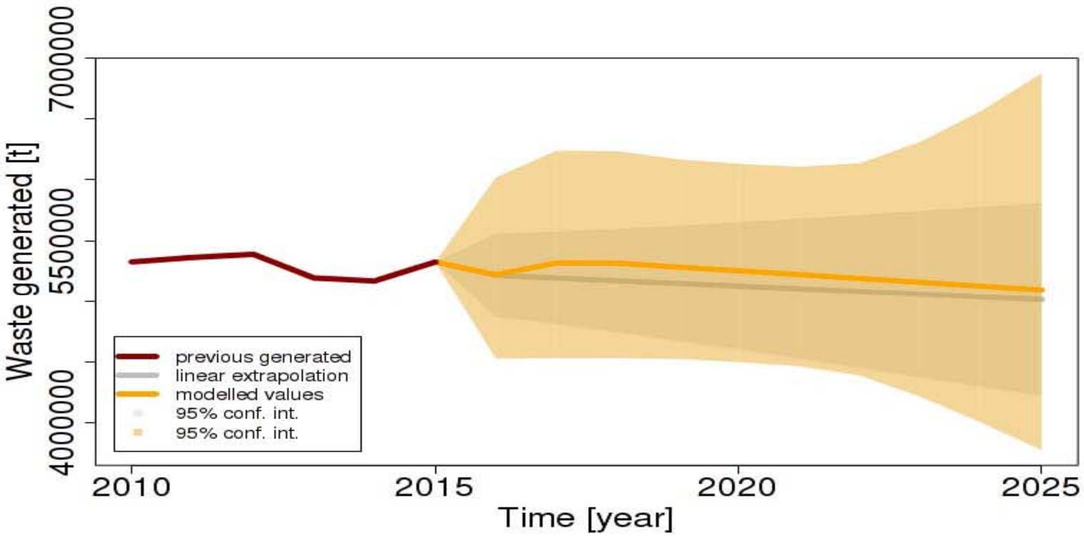
Show plot legend:

- none
- left down
- right down
- left up
- right up

Refresh plot

Show confidence interval:

- yes
- no



generation = 50331795.5 + -22390.5 . year
generation = 235.5678 . household ^ -0.4101 . population ^ 0.3962 . retired ^ 1.2023 . unemployment ^ -0.3317

Table of model outputs (in logarithmic form):

Predictor	Degrees of freedom	Sum of squared residuals	F statistics	p value
Variable: household	1	0.000324	6.3	0.241
Variable: population	1	5.3e-05	1	0.494
Variable: retired	1	2.6e-05	0.5	0.606
Variable: unemployment	1	0.001039	20.3	0.139
Residuals	1	5.1e-05		

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