Modelling and forecasting waste generation – DECWASTE information system

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

[1] Masaryk University, Brno, Czechia

[2] Czech Environmental Information Agency, Praha, Czechia

Introduction – content of the presentation

- Aim of the Decwaste project
- DPSIR methodology and results
- Open linked data sources
  - Ministry of the Environment
  - Ministry of Finance
  - Ministry of Regional Development
  - Czech Statistical Office
  - etc.
- Multilinear model of waste generation
- R Shiny web interface
- Results and the future development

Aim of the Decwaste project

The project, financed by the MoE consist of a development of a decision support modelling tool for waste generation (prevention) following these conditions:

- forecast quantities for all waste streams defined by Waste Statistics Regulation (EC);
- forecast the generation for 10-years period (up to 2025 for the first year) using a suitable mathematical model;
- deploy a method of Driving Force-Pressure-State-Impact-Response (DPSIR) to select appropriate model predictors;
- use linked open data of the Czech eGovernment system (current waste data, industrial and economical indicators) to gain current values of the predictors.

Aim of the Decwaste project

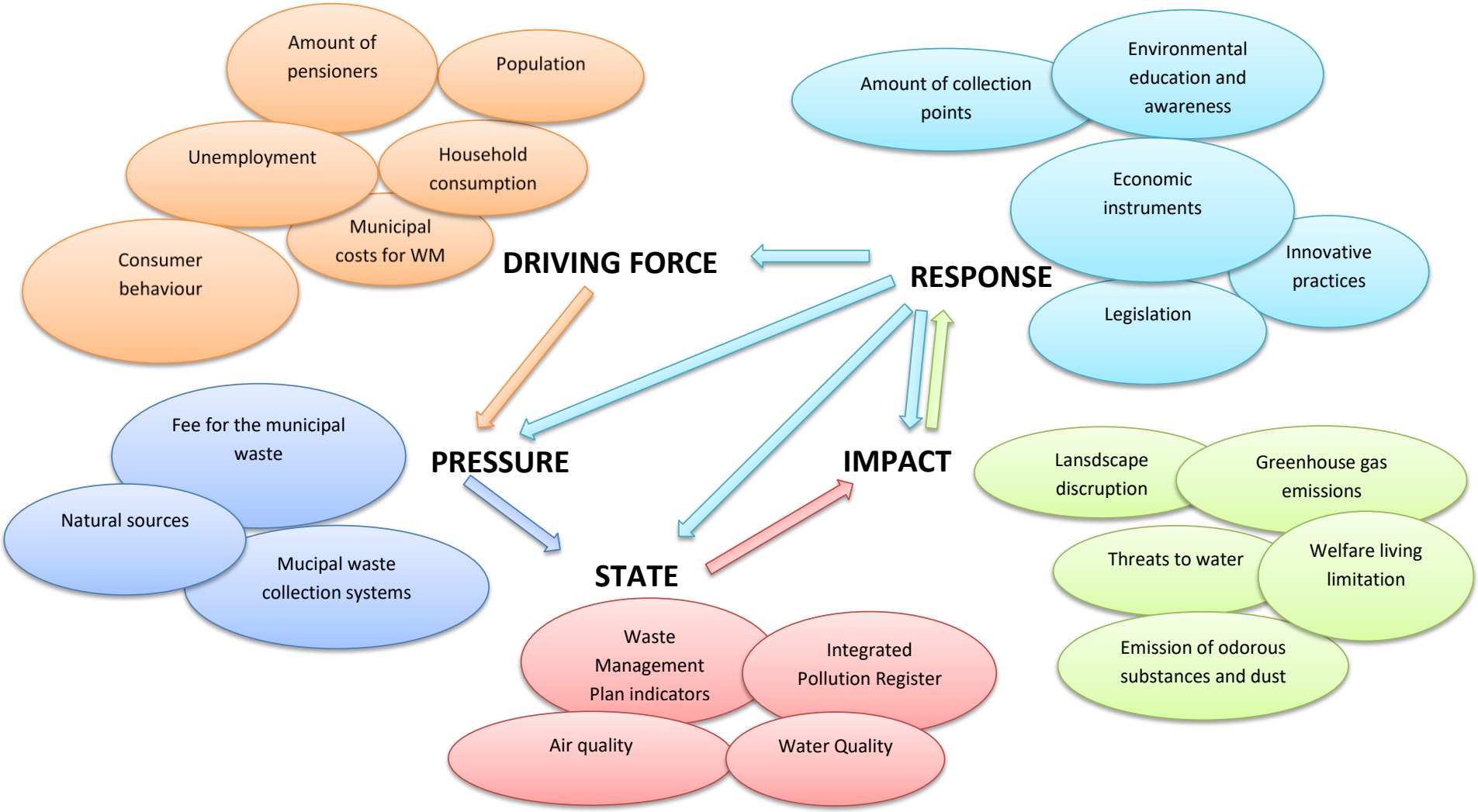
Development of new DECWASTE forecasting model of waste stream generation includes the following consequent steps:

1. *Identification of each waste stream listed in Section 2(1) (51 items) and waste generating activity listed in Section 8(1) (19 items) of Annex I of WSR and development of computation formulae for their amounts. This is done using aTable of equivalence of Annex III of WSR linking the European Waste Classification for Statistics and the European List of Waste.*
2. *Processing of the historical annual waste generation and treatment reports (2009–2015) based on the LoW provided by waste generators and facilities in Czech Republic to ISOH and creating their data sets of waste categories of WSR.*

Introduction

3. *Identification and development of socioeconomic and demographic predictors for waste categories and activities based on the DPSIR framework (which have influence on the generation of waste categories) using (linked open) government data (eGovernment systems) of Czech Republic.*
4. *Development of the multi-linear regression model of waste category generation with predictors from the DPSIR framework analyses.*
5. *Forecasting of predictors from the DPSIR framework and the calculation of waste category forecasts.*
6. *Analysis of sensitivity/significance of the predictors for each waste stream model.*
7. *Visualization scenarios forecasting the quantity of waste categories for the pre-scribed period 2016-2025.*

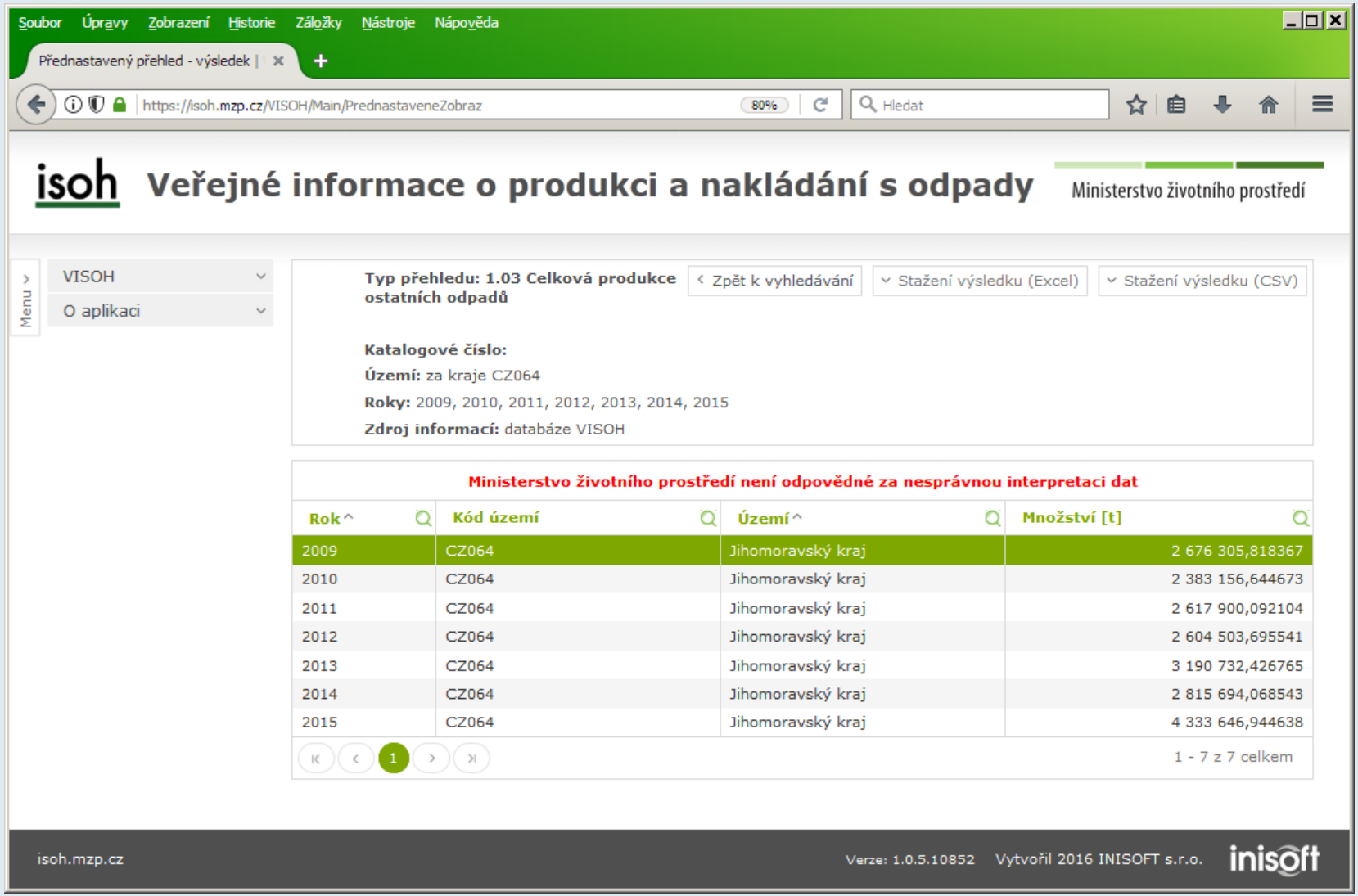
DPSIR framework – household waste option



Open linked data (eGovernment)

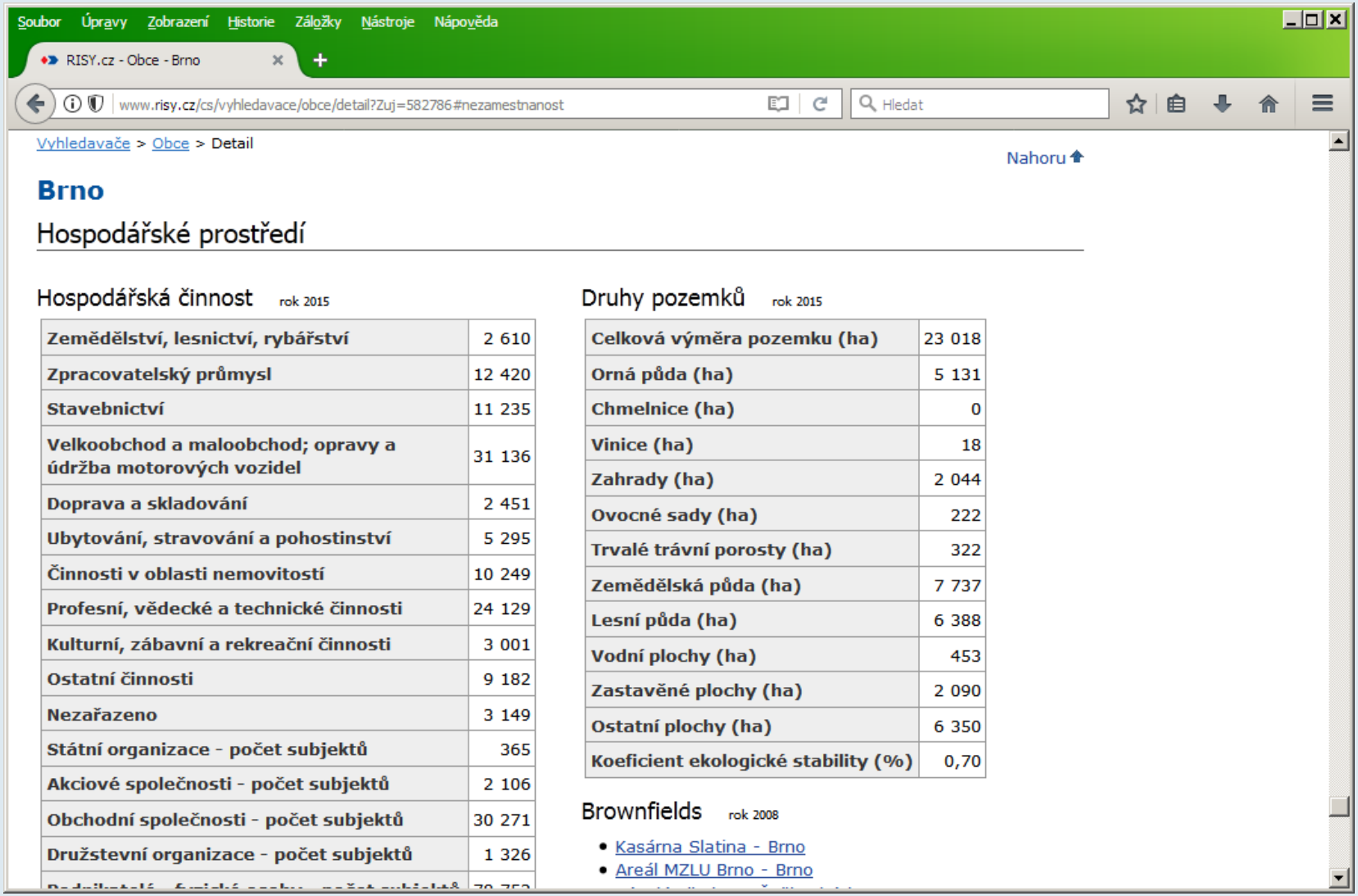
- eGovernment systems provide sources of necessary input data for the DPSIR framework predictors for different waste streams.
- The related linked open data for the DPSIR framework predictors are often available on web sites of Ministry of the Environment MoE, Czech Environmental Infomation Agency (CENIA), Ministry of Finance (MoF), Ministry of Regional Development (MoRD), Czech Statistical Office (CZSO) and Institute of Health Information and Statistics (ÚZIS).
- 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.

Open linked data (eGovernment)



Open linked data (eGovernment)

Open linked data (eGovernment)



Soubor Úpravy Zobrazení Historie Záložky Nástroje Nápověda

RISY.cz - Obce - Brno

www.risy.cz/cs/vyhledavace/obce/detail?Zuj=582786#nezamestnanost

Hledat

Vyhledavače > Obce > Detail

Nahoru

Brno

Hospodářské prostředí

Hospodářská činnost rok 2015

Zemědělství, lesnictví, rybářství	2 610
Zpracovatelský průmysl	12 420
Stavebnictví	11 235
Velkoobchod a maloobchod; opravy a údržba motorových vozidel	31 136
Doprava a skladování	2 451
Ubytování, stravování a pohostinství	5 295
Činnosti v oblasti nemovitostí	10 249
Profesní, vědecké a technické činnosti	24 129
Kulturní, zábavní a rekreační činnosti	3 001
Ostatní činnosti	9 182
Nezařazeno	3 149
Státní organizace - počet subjektů	365
Akciové společnosti - počet subjektů	2 106
Obchodní společnosti - počet subjektů	30 271
Družstevní organizace - počet subjektů	1 326

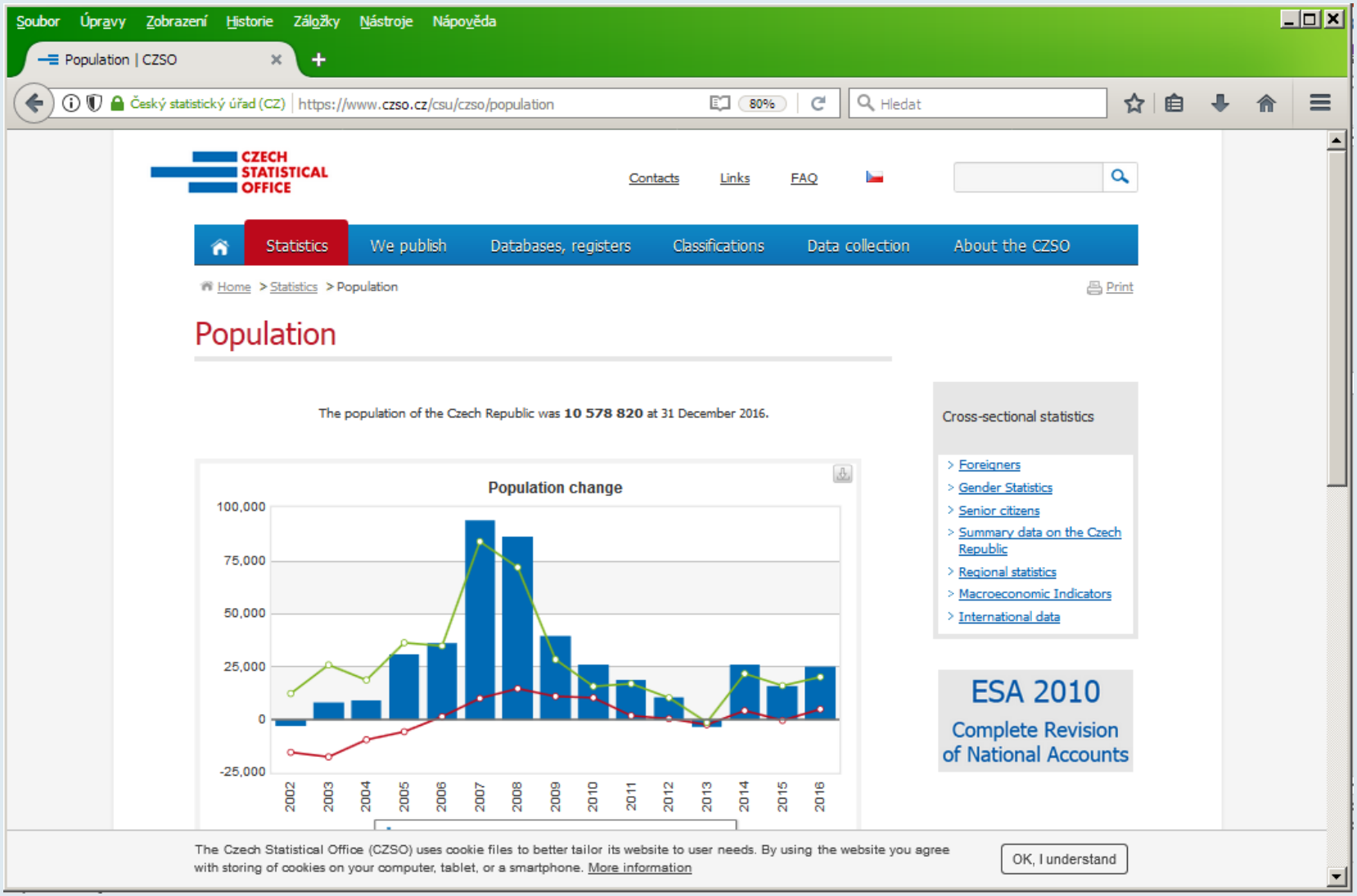
Druhy pozemků rok 2015

Celková výměra pozemku (ha)	23 018
Orná půda (ha)	5 131
Chmelnice (ha)	0
Vinice (ha)	18
Zahrady (ha)	2 044
Ovocné sady (ha)	222
Trvalé trávní porosty (ha)	322
Zemědělská půda (ha)	7 737
Lesní půda (ha)	6 388
Vodní plochy (ha)	453
Zastavěné plochy (ha)	2 090
Ostatní plochy (ha)	6 350
Koeficient ekologické stability (%)	0,70

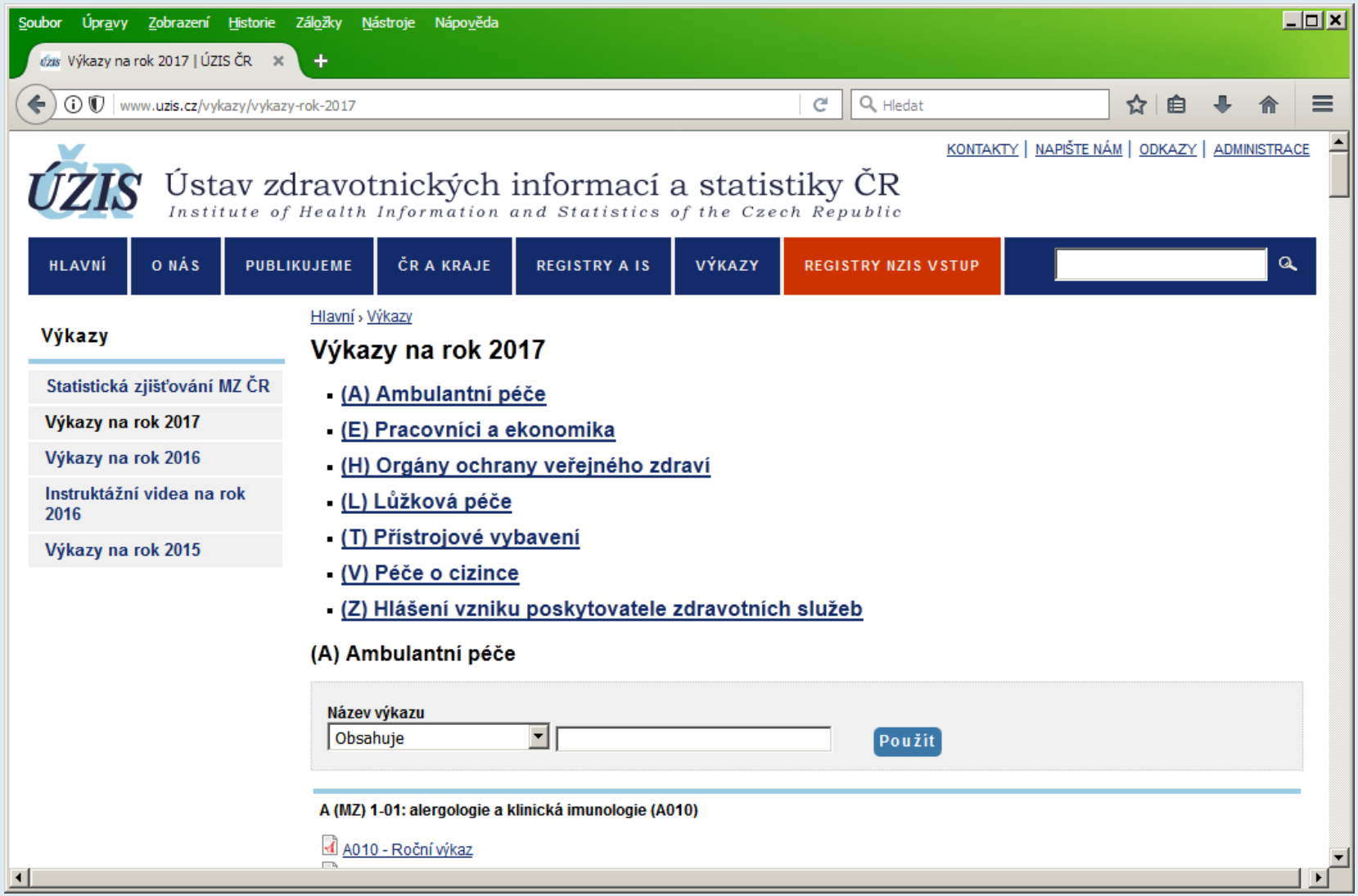
Brownfields rok 2008

- Kasárna Slatina - Brno
- Areál MZLU Brno - Brno

Open linked data (eGovernment)



Open linked data (eGovernment)



Open linked data – list of the predictors

abbreviation	description	unit per year	source
hdp	Gross domestic product	billion CZK	http://czso.cz
energie	Electric energy by steam power stations	GWh	http://www.eru.cz
prumysl	Index of an industrial production	%	http://czso.cz
obyvatel	Census (number of inhabitants)	person	http://risy.cz
stavbys	Type “S” construction works	million CZK	http://czso.cz
domacnosti	Household expenditures	billion CZK	http://czso.cz
duchodci	Share of retirees	%	http://risy.cz
nezamestnanost	Unemployment rate	%	http://risy.cz
eobyvatel	Share of inhabitants with water treatment plants	%	http://mvcr.cz
ochrana	Municipal expenditures for landscape and biodiversity protections	million CZK	http:// monitor.statnipokladna.cz/
urbanizace	Share of inhabitants living in towns and cities	%	http://risy.cz
vrak	Annual number of car wrecks	piece	http://autovraky.mzp.cz/
stari	Share of inhabitants > 65 years	%	http://risy.cz

Open linked data – list of the predictors

abbreviation	description	unit per year	source
osobys	Number of persons deployed in construction works	thousands of persons	
sfdi	Investments into the transport infrastructure	billion CZK	http://sfdi.cz
byty	Number of completed dwellings	piece	http://czso.cz
kontejnery	Number of containers for separated municipal waste	piece	http://ekokom.cz
mzda	Average wage	CZK	http://czso.cz
uvery	Amount of consumer credit granted	million CZK	http://cnb.cz
vozidla	Number of road vehicles	piece	http://autosap.cz
doprava	Freight transport performance	billion tkm	http://mdcr.cz
ovody	Waste water treatment expenditures	thousand CZK	http://czso.cz
zarizeni	Total number of health facilities	piece	http://www.uzis.cz/
hospital	Number of hospitalized patients	person	http://www.uzis.cz/
osetreni	Number of outpatient treatments	piece	http://www.uzis.cz/
zdravydaje	Healthcare expenditures	million CZK	http://www.uzis.cz/

Multivariate model of waste generation

Let us assume that for given waste category f, the amount of waste $\hat{w}_t^f$ and predictors $\hat{A}_{i,t}^f$, $i=1,\ldots, K^f$; in years $t=2009,\ldots,2015$ are known, where K^f is the number of predictors for the waste category f of WSR. Let waste category generation $w^f(t)$ at the given year t fulfil the equation

$$\boldsymbol{log}(\boldsymbol{w^f(t)}) = \boldsymbol{a_0^f} + \sum_{i=1}^{K^f} \boldsymbol{a_i^f} \cdot \mathbf{log}\left(\boldsymbol{A_i^f(t)}\right) + \boldsymbol{\varepsilon_t^f}, \qquad (1)$$

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 category f,
- $\varepsilon_t^f = \log\left(w^f(t)\right) - \log(\hat{w}_t^f)$, for $t=2009,\ldots,2015$ are approximation errors.
- Coefficients $a_0^f,\ldots,a_K^f$ in (1) for each waste category f are calculated using multiple regression on the basis of the values of waste generation $\hat{w}_t^f$ and predictors $\hat{A}_{i,t}^f$, $i=1,\ldots,K^f$; $t=2009,\ldots,2015$. Approximation errors ε_t^f, $t=2009,\ldots,2015$, have the mean equal to 0 and the normal distribution.

Model – parameters for HW identified by DPSIR

- *Population*: development of the population is a natural driving force of household waste generation, but a development of the census need not to be significant.
- As shown before, the number of *pensioners* and their change could play a role in household waste generation processes.
- *Unemployment* as well as families with small children, students and other low income groups of inhabitants could reveal lower household waste generation.
- *Household expenditures* are suspect to influence the household waste generation although there are several waste prevention measures taken.

Model – present HW generation

year	Number of inhabitants [person]	Total generation of household waste [t]	Total generation of household waste per capita [kg/person]
2010	10,517,247	3,732,838	355
2011	10,495,430	3,598,752	343
2012	10,509,286	3,424,688	323
2013	10,510,719	3,348,158	319
2014	10,524,783	3,398,917	323
2015	10,542,942	3,314,371	314

Multi-linear regression model of waste streams generation

If we want to establish the confidence interval of predictors $A_i^f(t)$, it is necessary to restrict their number to $K^f \leq 5$, since we only have a time series of six past known values. If we have the values of $\hat{w}_t^f$ and $\hat{A}_{i,t}^f$ for the next years t=2016,..., the model (1) will be more accurate and the approximation error $\varepsilon_t^{f,ps}$ will be smaller.

Furthermore, we assume that the predictors $A_i^f(t)$, i=1,...,K^f, for t=2016,...,2025 have either known values (e.g. GDP, population, household consumption, etc.) from the eGovernment systems [6, 7] or are determined by an appropriate extrapolation method or are chosen by decision makers using DECWASTE.

Multi-linear regression model of waste streams generation

These models are implemented in DECWASTE, written in language R and they use predefined predictors $\hat{A}_{i,t}^{f}$, which were parsed from linked open data. The outputs $w^{f}(t)$ are time series of the amount of waste generated for the years t=2016,...,2025 of waste category f.

An estimate of the waste category generation $w^{f}(t)$ can be expressed after treatment (1) as

$$\boldsymbol{w^{f}(t) = A_{0}^{f} \cdot \prod_{i=1}^{K^{f}} A_{i}^{f}(t)^{a_{i}^{f}}}, \quad (2)$$

where $A_{0}^{f} = \exp(a_{0}^{f})$, while we have neglected the error ε_{t}^{f}.

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 H0: $\{s^2 = s_i^2\}$. We set this value as $p_i = 1\text{-}H(F_i)$.

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

- 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 $A_i^f(t)$ 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.

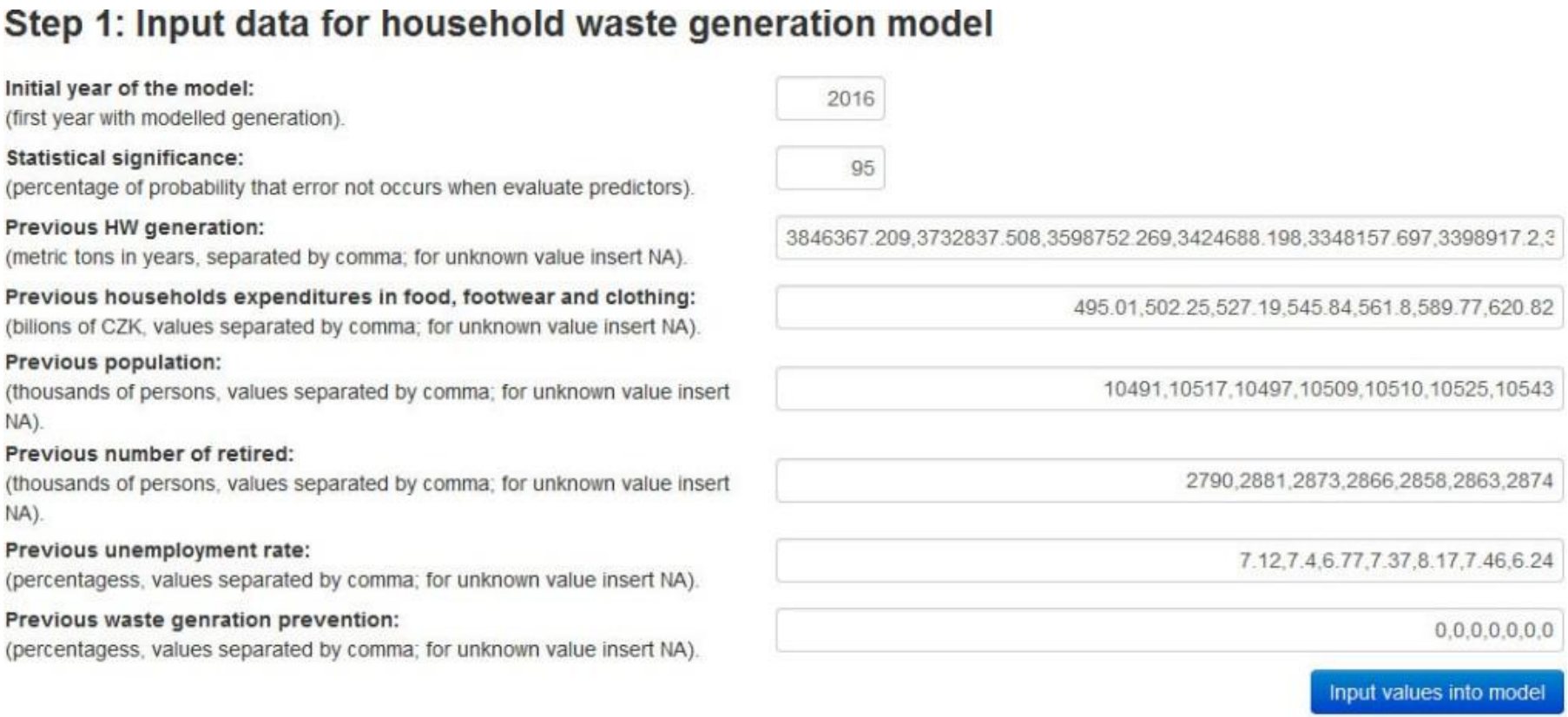
Extrapolation of the predictors

Let us suppose that for each waste category f the amount of waste category generation $\hat{w}_t^f$ and predictors $\hat{A}_{i,t}^f$, i=1,...,K^f; t=2009,...,2015 are known and we have calculated for each waste category f the coefficients a_0^f,..., a_K^f in the model (1) by using multiple regression.

For the calculation of the forecast waste category generation $w^f(t)$ in the years t=2016,...,2025 it is necessary to know the values of predictors $A_i^f(t)$, i=1,..., K^f, t=2016,...,2025. 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 $\hat{A}_{i,t}^f$, i=1,...,K^f; t=2009,...,2015. In this case, the procedure is as follows:

- Enter the values of the predictors based on experts' estimates or other appropriate sources;
- On the basis of the values of the predictors $\hat{A}_{i,t}^f$, i=1,...,K^f; t=2009,...,2015 the values of predictors $A_i^f(t)$,i=1,..., K^f, t=2016,...,2025 are calculated using either linear or exponential extrapolation.

R Shiny web interface



The model is available on following URL:
http://opencpu.iba.muni.cz:8080/shiny/tacr/household_waste/

R Shiny web interface

Step 2: Expected development up to 2025

	2016:	2017:	2018:	2019:	2020:	2021:	2022:	2023:	2024:	2025:	
Households expenditures in food, footwear and clothing: [bil. CZK]	632.82	653.79	674.76	695.73	716.69	737.66	758.63	779.59	800.56	821.53	Source: https://vdb.czso.cz/vdbvo2/faces/index.jsf?page=statistiky#katalog=30847
Population [thsds. persons]	10491	10439	10388	10336	10284	10271	10257	10244	10231	10217	Source: https://www.czso.cz/csu/czso/inhabitantsstvo_hu
Retired [thsds. persons]	2944	3011	3078	3148	3219	3291	3365	3441	3518	3597	Source: http://www.cssz.cz/cz/o-cssz/informace/informacni-materialy/statisticke-rocenky.htm
Unemployment [%]	5.19	6.08	5.89	5.69	5.5	5.3	5.11	4.91	4.71	4.52	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
Prevention [%]	2	4	5	7	8	10	11	12	14	15	

Input values into model

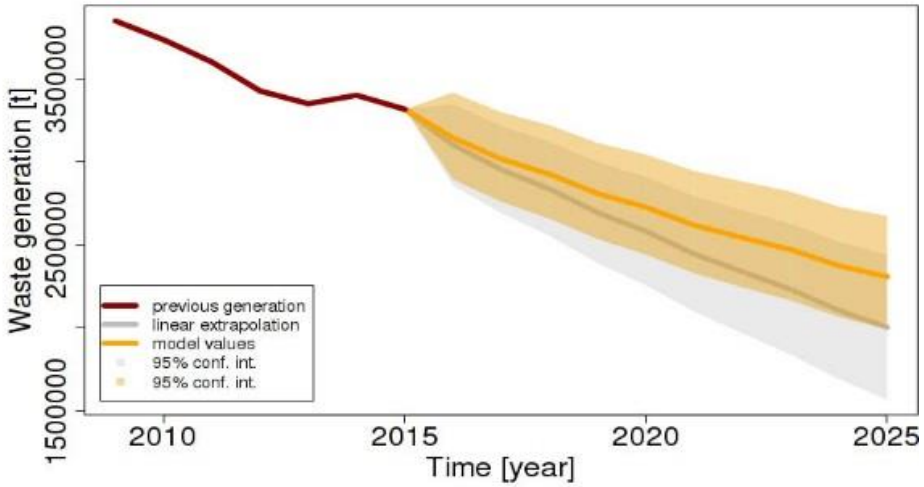
R Shiny web interface

Step 3: Prognosis and model equation

Show plot legend: Show confidence interval:

- nowhere / yes
- bottom left / no
- bottom right
- upper left
- upper right

Refresh plot



Extrapolation (time): generation = (184202618.9 + -89800.8 × year) × (1 - prevention)
Model: log(generation) = (19.102 + -0.6389 × log(households)) × (1 - prevention)

Download table

Extrapolation in time: generation = (184202618.9 + -89800.8 × year) × (1 - prevention)
Model generation: log(generation) = (19.102 + -0.6389 × log(households)) × (1 - prevention)
Table of model values:

	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
Model generation	**3142702**	**3015105**	**2924115**	**2807124**	**2724773**	**2616872**	**2541862**	**2469916**	**2373191**	**2307163**
Lower border	2892087	2757503	2656426	2532306	2440366	2326659	2243432	2164017	2064174	1992307
Upper border	3415033	3296770	3218778	3111766	3042326	2943285	2879990	2819056	2728469	2671778
Linear extrapolation	**3100954**	**2951460**	**2835405**	**2692198**	**2580633**	**2443711**	**2336636**	**2231357**	**2103416**	**2002627**
Lower border	2855472	2689412	2552168	2389578	2254724	2097809	1966937	1837807	1690913	1566925
Upper border	3346435	3213508	3118642	2994818	2906541	2789614	2706336	2624907	2515918	2438328

R Shiny web interface

Step 4: Sensitivity analysis

Show plot legend:

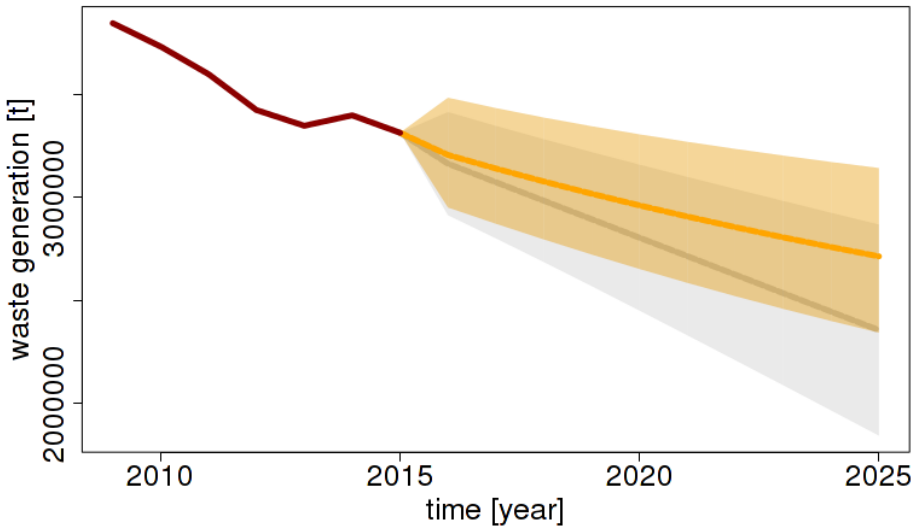
- (•) nowhere
- ( ) lower left
- ( ) lower right
- ( ) upper left
- ( ) upper right

Refresh plot

Show confidence interval:

- (•) yes
- ( ) no

Final model (significant predictors only)



generation = 184202618.9 + -89800.8 × year
generation = 197643651.6578 × households ^ -0.6389

Table of model results (in log form):

Predictor	Degrees of freedom	Sum of squared residuals	F statistics	p value
Variable: households	1	0.017	40.3	0.024
Variable: inhabitants	1	0	0.9	0.437
Variable: retired	1	0.001	2.6	0.25
Variable: unemployment	1	0.001	2.1	0.286
Residuals	2	0.001	NA	NA

Conclusion

- The system DECWASTE allows decision makers at the national level of Czech Republic to make sustainable waste management decisions and customize national strategies for waste data acquisition, management and processing.
- The most inetersting results show, that the relation between the identified driving forces is usually no so straightforward as expected and sometimes it is even really surpristing.
- In case of the household waste, the only significant predictor is household expenditures, but in a negative form:
- generation = 197,643,652 × households ^ -0.639

  this means, that expenditures of a billion of CZK (37 mil. €) decreases the generation by about 0,106%.

Thank you for your attention
Questions?

Jiří Kalina, Jiří Hřebíček

Masaryk University

Brno, Czech Republic

{kalina, hrebicek}@iba.muni.cz