

# Preliminary Assessment of Polycyclic Aromatic Hydrocarbons Associated to Airborne PM10 in Győr, Hungary

A. Szabó Nagy, Zs. Csanádi, J. Szabó

Széchenyi István University, Physics and Chemistry Department  
Egyetem square 1, 9026 Győr, Hungary  
e-mail: nszaboa@sze.hu

**Abstract:** The aim of this study was to assess the ambient concentrations of polycyclic aromatic hydrocarbons (PAHs) associated to the PM10 aerosol fraction in an urban site of Győr, Hungary in 2011. The levels of total PAHs and the individual compounds show large variability during the sampling periods. The total PAH concentrations ranged from 0.88 ng/m<sup>3</sup> to 95.80 ng/m<sup>3</sup> with the mean value of 20.06 ng/m<sup>3</sup>. Four and five-rings PAHs were the dominant species in the samples. The levels of carcinogenic PAH species (benzo[a]pyrene, benz[a]anthracene, sum of the three benzo[fluoranthene isomers, indeno[1,2,3-c,d]pyrene and dibenz[a,h]anthracene) determined were compared with published data of other Hungarian cities. The potential carcinogenic risks of PAHs were also estimated.

**Keywords:** *polycyclic aromatic hydrocarbons, PAHs, ambient air, air quality*

## 1. Introduction

Atmospheric aerosols are very important ambient components in point of air quality. Monitoring the particulate matter with an aerodynamic diameter smaller than 10 µm (PM10) and PM10-bound polycyclic aromatic hydrocarbons (PAHs) could have important environmental significance and health protection aspects [1].

PAHs are a large group of organic compounds included two or more coupled benzene rings arranged in various configurations. In addition to natural sources they generally derive from the incomplete combustion of organic materials and pyrolysis of hydrocarbons (e.g. coal, petrol, diesel and wood). The main PAH sources are traffic exhausts and industrial emissions, domestic heating and oil refining. Moreover, it is well known that road dust resuspension can contribute to fine aerosols especially during dry weather conditions. It is suggested that road dust may be a significant source of particle-bound PAHs in ambient air. Possible PAH sources in road dust include vehicle exhaust, tire, pavement (asphalt or bitumen), and oil spill. PAH concentrations are the highest in areas of traffic followed by urban sites, and lowest in rural sites [2-3]. There are thousands of PAH compounds in the environment but in practice PAH analysis is limited to the

determination of a few compounds [4-5]. Some main representatives of PAHs are shown in Fig. 1.

PAHs belong to the group of persistent organic pollutants with toxic, carcinogenic and mutagenic properties. Benzo[a]pyrene (BaP) has been identified as an indicator carcinogenic PAH compound [6]. The degree of toxicity and carcinogenicity of PAHs is dependent on the type of compound. PAHs can persist and accumulate in the environment with high risk of bioaccumulation in human and animal tissues, they are largely resistant to biodegradation and can remain in the environment for long periods [7]. The physicochemical properties of PAHs make them highly mobile in the environment, allowing them to spread in air, soil, and water. Atmospheric PAHs can be present in solid or gaseous substances, they are associated predominantly with particulate matters [8].

The aim of this work was to assess the ambient concentrations of PAHs in the PM10 aerosol fraction in an urban site of Győr, Hungary in 2011. This was the first time to comprehensively study the PAH pollution status in the Győr atmosphere. The potential carcinogenic risks of PAHs were also estimated. The levels of carcinogenic PAHs determined in our study were compared with published data of other Hungarian cities.

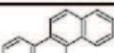
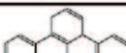
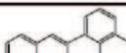
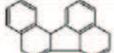
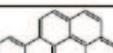
 <i>naphthalene*</i> C <sub>10</sub> H <sub>8</sub>	 <i>acenaphthylene (D)</i> C <sub>12</sub> H <sub>8</sub>	 <i>acenaphthene</i> C <sub>12</sub> H <sub>10</sub>
 <i>fluorene (D)</i> C <sub>13</sub> H <sub>10</sub>	 <i>phenanthrene (D)</i> C <sub>14</sub> H <sub>10</sub>	 <i>anthracene (D)</i> C <sub>14</sub> H <sub>10</sub>
 <i>fluoranthene (D)</i> C <sub>16</sub> H <sub>10</sub>	 <i>pyrene (D)</i> C <sub>16</sub> H <sub>10</sub>	 <i>benzo[a]anthracene (B2)</i> C <sub>18</sub> H <sub>12</sub>
 <i>chrysene (B2)</i> C <sub>18</sub> H <sub>12</sub>	 <i>benzo[b]fluoranthene (B2)</i> C <sub>20</sub> H <sub>12</sub>	 <i>benzo[k]fluoranthene</i> C <sub>20</sub> H <sub>12</sub>
 <i>benzo[j]fluoranthene</i> C <sub>20</sub> H <sub>12</sub>	 <i>benzo[a]pyrene (B2)</i> C <sub>20</sub> H <sub>12</sub>	 <i>benzo[e]pyrene</i> C <sub>20</sub> H <sub>12</sub>
 <i>dibenz[a,h]anthracene (B2)</i> C <sub>22</sub> H <sub>14</sub>	 <i>benzo[g,h,i]perylene (D)</i> C <sub>22</sub> H <sub>12</sub>	 <i>indeno[1,2,3-c,d]pyrene (B2)</i> C <sub>22</sub> H <sub>12</sub>

Figure 1. Structure of primary important PAHs [5] (\*: not included in priority list; D: not listed as human carcinogen; B2: probable human carcinogen)

## 2. Experimental

### 2.1. Study area

Győr (47°41'02"N, 17°38'06"E) is the most important city in the northwest area of Hungary halfway between Wien, Bratislava and Budapest situated on one of the important roads of Central Europe. The city is the sixth largest in Hungary, and one of the seven main regional centres of the country. The location of Győr is shown in Fig. 2. The number of inhabitants is about 128,500. Győr is a dynamically developing city due to its good geographic situation and as an emphasized centre in automotive industry. It has become one of the largest economic, industrial and traffic areas of Hungary. The monitoring site is located at the junction of Tihanyi Árpád Street and Ifjúság Boulevard, where the main pollution source is the traffic.

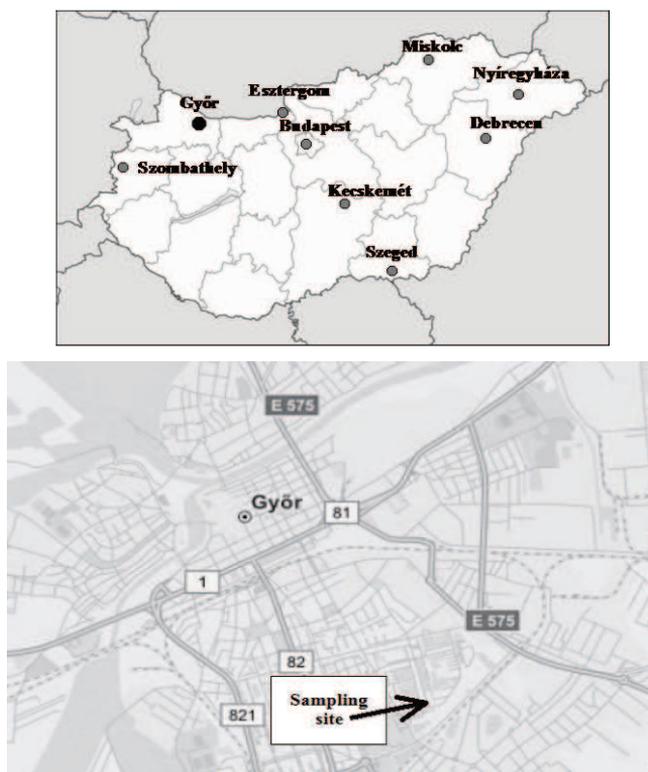


Figure 2. Schematic map of Hungary showing the location of Győr and the sampling site

### 2.2. Sampling and chemical analysis

The concentrations of PM<sub>10</sub> aerosol samples were collected in every third month in spring (I.), summer (II.), autumn (III.) and winter (IV.) at 14 day intervals, continuously for 24 hours in 2011 at the monitoring site of Győr. A Digital High Volume DHA80

(Digital Elektronik AG, Switzerland) sampler [9] was used for collection ambient aerosol particles, which were chemically analysed later. This equipment is considered to be equivalent to the requirements of the European Standard (EN 12341) for sampling PM<sub>10</sub> matter [10]. In our previous work [11] we described the PM<sub>10</sub> sampling, in details.

The ultrasonic liquid-solid extraction of the filter (Advantec QR-100 quartz fibre, d=150 mm) and the PAH analysis were conducted in accordance with the Hungarian standard method procedure [12]. A gas chromatograph-mass selective detector (GC/MSD) system consisting of an Agilent 6890 gas chromatograph with an Rtx-5MS Integra GC column (30 m long, 0.25 mm internal diameter, 0.25 µm coating, 5% diphenyl – 95% dimethyl polysiloxane eluent) and an Agilent 5973 MSD was used in the study. The total PAH concentration was regarded as the sum of the concentrations of 19 measured PAH species for each collected sample.

### 3. Results and discussion

#### 3.1. Concentration and distribution of PAHs

Table 1 gives an overview of the concentration ranges, mean values and standard deviations (SD) of the measured PAH compounds associated to PM<sub>10</sub> aerosol particles during different sampling periods in 2011, respectively. Fig. 3 shows the average concentrations and SD of total PAHs during the sampling periods.

The total PAH concentrations ranged from 0.88–95.80 ng/m<sup>3</sup> with the mean value of 20.06 ng/m<sup>3</sup>. The levels of total PAHs were relatively higher in heating season (periods I. and IV.) than in spring or summer. Similar to the concentration trend of total PAHs, concentrations of individual PAHs exhibit a large variability during the different sampling periods.

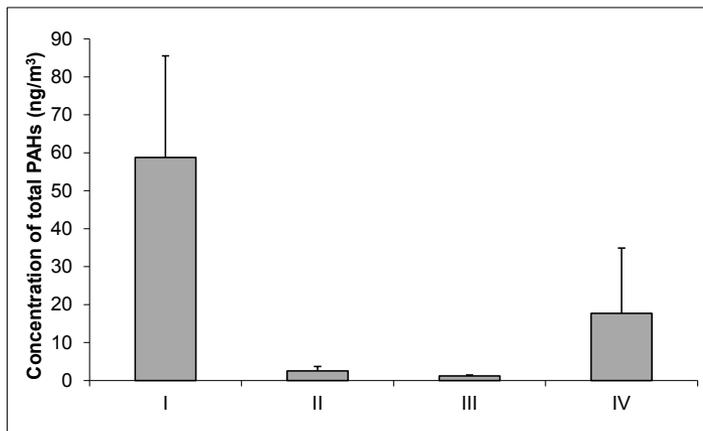


Figure 3. Total PAH concentrations at the urban site of Győr during the four sampling periods in 2011

Table 1. Concentrations of PAHs in PM10 at the urban site of Győr during different sampling periods in 2011 (ng/m<sup>3</sup>)

PAH compound	Sampling period			
	I. 16 February- 1 March	II. 4 May- 17 May	III. 1 August- 14 August	IV. 1 November- 15 November
Naphthalene	0.58–0.90 0.66 ± 0.09	0.11–0.64 0.17 ± 0.14	0.29–0.38 0.34 ± 0.03	0.18–0.31 0.22 ± 0.03
2-Methylnaphthalene	0.35–0.83 0.50 ± 0.13	0.12–0.32 0.16 ± 0.05	0.10–0.19 0.13 ± 0.02	0.04–0.09 0.05 ± 0.01
Acenaphthylene	0.03–0.99 0.33 ± 0.27	ND	ND	ND–0.02 0.00 ± 0.01
Acenaphthene	0.02–0.13 0.05 ± 0.03	ND–0.03 0.00 ± 0.01	ND	ND
Fluorene	0.07–0.91 0.34 ± 0.26	0.03–0.09 0.04 ± 0.02	0.03–0.04 0.03 ± 0.00	ND–0.02 0.02 ± 0.01
Phenanthrene	1.12–7.58 3.89 ± 2.32	0.10–0.28 0.18 ± 0.06	0.09–0.14 0.11 ± 0.02	0.02–0.37 0.13 ± 0.11
Anthracene	0.08–0.80 0.32 ± 0.23	ND–0.03 0.01 ± 0.01	ND–0.02 0.01 ± 0.01	ND–0.06 0.02 ± 0.02
Fluoranthene	2.04–17.70 9.11 ± 5.10	0.11–0.53 0.25 ± 0.11	0.03–0.09 0.06 ± 0.02	0.28–5.43 1.52 ± 1.68
Pyrene	1.59–12.12 6.51 ± 3.34	0.09–0.50 0.21 ± 0.11	0.04–0.09 0.06 ± 0.02	0.40–6.54 1.86 ± 2.02
Benz[a]anthracene	0.79–6.19 3.81 ± 1.86	0.03–0.14 0.07 ± 0.03	0.01–0.07 0.03 ± 0.01	0.19–6.99 1.82 ± 2.24
Chrysene	1.17–4.62 3.16 ± 1.22	0.08–0.26 0.15 ± 0.06	0.02–0.09 0.05 ± 0.02	0.27–6.72 1.93 ± 2.10
Benzo[b-k-j]fluoranthene	3.52–16.39 10.99 ± 4.40	0.25–1.51 0.57 ± 0.38	0.06–0.26 0.14 ± 0.06	0.96–10.41 3.76 ± 3.33
Benzo[e]pyrene	2.32–9.15 6.20 ± 2.35	0.11–0.53 0.22 ± 0.13	0.03–0.12 0.07 ± 0.03	0.25–3.06 1.01 ± 0.89
Benzo[a]pyrene	0.87–6.86 3.80 ± 1.90	0.03–0.43 0.15 ± 0.13	0.02–0.09 0.04 ± 0.02	0.41–6.92 2.14 ± 2.09
Indeno[1,2,3-c,d]pyrene	0.81–7.88 4.71 ± 2.40	0.06–0.45 0.15 ± 0.12	0.03–0.11 0.06 ± 0.03	0.43–5.36 1.77 ± 1.58
Dibenz[a,h]anthracene	0.09–0.96 0.51 ± 0.29	ND–0.05 0.02 ± 0.02	ND–0.07 0.02 ± 0.02	0.04–0.76 0.22 ± 0.24
Benzo[g,h,i]perylene	0.89–6.43 3.88 ± 1.79	0.07–0.52 0.20 ± 0.13	0.03–0.12 0.06 ± 0.03	0.32–3.61 1.21 ± 1.03

ND: Not detected

The compositional pattern of PAHs is shown in Fig. 4 and 5. Four and five-rings PAHs including fluoranthene (FLT), pyrene (PYR) and sum of the three benzofluoranthene isomers (BbkjF) were the dominant species in the samples. The carcinogenic PAH species (BaP, BbkjF, benz[a]anthracene (BaA), indeno[1,2,3-c,d]pyrene (IND) and dibenz[a,h]anthracene (DahA)) together contributed 43.36 % of the mass of the total PAHs on average.

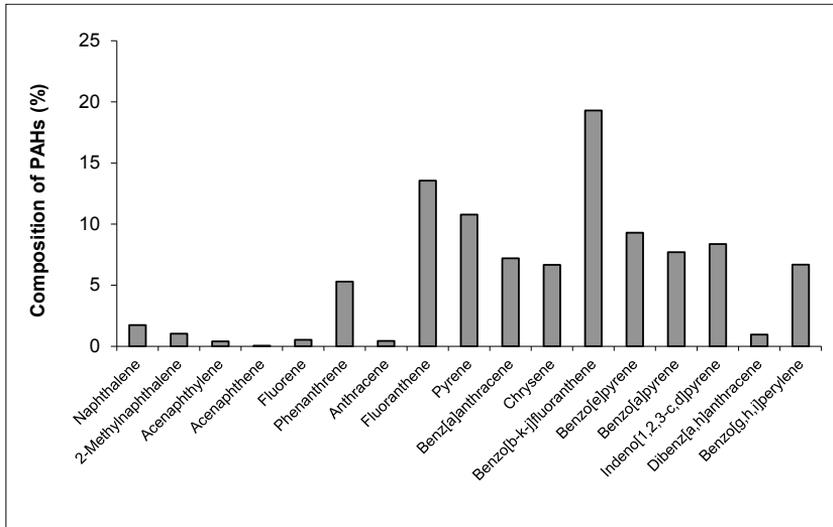


Figure 4. The composition of PAHs measured in Győr in 2011

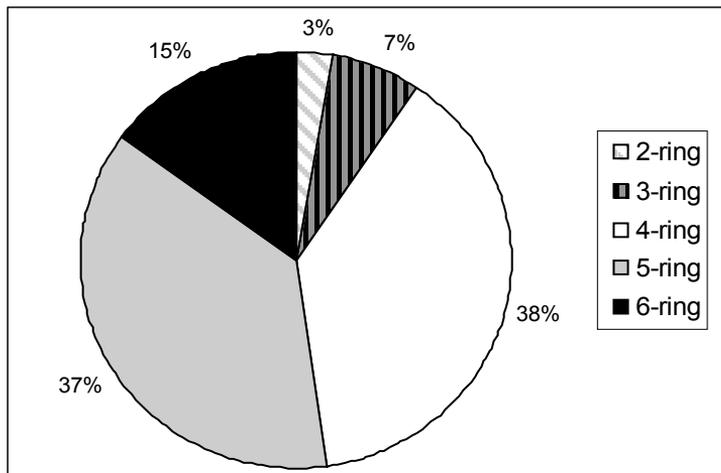


Figure 5. The composition of PAHs based on the number of their constituent benzene rings measured in Győr in 2011

### 3.2. Concentration of benzo[a]pyrene

Among the PAH compounds, only the BaP concentration is regulated in Hungary and also in EU. The Hungarian daily and annual mean limit values for health protection are 1 ng/m<sup>3</sup> and 0.12 ng/m<sup>3</sup>, respectively [13]. However, the annual mean target value in the EU legalisation (also in Hungary) is 1 ng/m<sup>3</sup> [14]. There is no relevant guideline value for other PAH compounds or total PAHs. The Hungarian air quality index (AQI) can also be used as a reference [15]. It defines air quality categories based on the ratio of pollution (Table 2).

The BaP concentrations ranged from 0.41 to 6.92 ng/m<sup>3</sup> with the mean of 2.97 ng/m<sup>3</sup> in heating season, and from 0.02 to 0.43 ng/m<sup>3</sup> with the mean of 0.09 ng/m<sup>3</sup> in the non-heating season. The annual mean concentration was 1.53 ng/m<sup>3</sup>, which exceeded the EU target value. However, it was 12.75 times higher than the Hungarian limit value for health protection. The BaP contamination exceeded the Hungarian daily limit value in 65.52 % of the samples collected in heating season. The comparison of the daily BaP concentrations with the AQI values show mainly acceptable, polluted or heavily polluted results in heating season, while indicate excellent or good air quality in the non-heating season.

Table 2. Hungarian air quality index for air pollutants as BaP

Category	1. Excellent	2. Good	3. Acceptable	4. Polluted	5. Heavily polluted
Related to the limit value in %	0–40	40–80	80–100	100–200	200–

### 3.3. Toxicity of PAHs

BaP believed to be the most toxic PAH and it has been well characterized toxicologically. However, less information is available for most of the other PAHs. Several approaches have been developed to obtain a more accurate assessment of potential risk of exposure to a complex mixture of PAHs using toxic equivalency factors (TEFs) based on BaP. The TEF methodology was developed by the U.S. Environmental Protection Agency (EPA) to evaluate the toxicity and assess the risks of a mixture of structurally related chemicals [16]. The TEF for each PAH compound is an estimate of the relative toxicity of the PAH compound compared to BaP. BaP equivalent (BaP-eq) concentration is a useful metric to quantitatively assess the carcinogenic health risk of PAHs [17-18].

The total BaP-eq concentrations based on a number of contributing compounds in the sample can be calculated with the following equation:

$$\text{Total BaP-eq} = \sum c_i \cdot \text{TEF}_i \quad (1)$$

where,  $c_i$  and  $\text{TEF}_i$  are the concentration and toxicity equivalency factor of individual PAH compounds, respectively.

In this study TEF values for 14 PAH compounds reported by Larsen and Larsen were used [17]. Table 3 shows the summarized data for the applied TEF approach. BaP-eq was calculated for each component by multiplying the average concentration by its TEF value for each sampling period. The results show that among the analysed PAHs, BaP is the main contributor to the carcinogenicity of the examined PAHs. DaA, IND, FLT and BbkjF also have significant effect in total BaP-eq. The calculated BaP-eq concentration in winter is 25 and 75 times higher than in spring and summer, respectively.

Table 3. Benzo[a]pyrene equivalent exposure profile in PM10 at the urban site of Győr during different sampling periods in 2011

PAH compound	TEF [17]	Sampling period							
		I.		II.		III.		IV.	
		BaP-eq (ng/m <sup>3</sup> )	BaP-eq (%)						
Phenanthrene	0.0005	0.0019	0.031	0.0001	0.036	0.00006	0.066	0.00007	0.002
Anthracene	0.0005	0.0002	0.002	0.0001	0.002	0.00001	0.006	0.00001	0.0001
Fluoranthene	0.05	0.456	7.308	0.0125	5.059	0.003	3.595	0.076	2.557
Pyrene	0.001	0.0065	0.104	0.0002	0.085	0.0001	0.072	0.0019	0.063
Benzo[a]anthracene	0.005	0.0191	0.306	0.0004	0.142	0.00015	0.180	0.0009	0.031
Chrysene	0.03	0.0948	1.521	0.0045	1.821	0.0015	1.798	0.0579	1.948
Benzo[b]fluoranthene*	0.1	0.3663	5.877	0.019	7.689	0.0047	5.597	0.1253	4.216
Benzo[k]fluoranthene*	0.05	0.1832	2.939	0.0095	3.845	0.0023	2.792	0.0627	2.108
Benzo[j]fluoranthene*	0.05	0.1832	2.939	0.0095	3.845	0.0023	2.792	0.0627	2.108
Benzo[e]pyrene	0.002	0.0124	0.199	0.0004	0.178	0.0001	0.168	0.0020	0.068
Benzo[a]pyrene	1.0	3.800	60.969	0.150	60.704	0.04	47.939	2.140	71.990
Indeno[1,2,3-c,d]pyrene	0.1	0.471	7.557	0.015	6.070	0.006	7.191	0.177	5.954
Dibenz[a,h]anthracene	1.1	0.561	9.001	0.022	8.903	0.022	26.366	0.242	8.141
Benzo[g,h,i]perylene	0.02	0.078	1.245	0.004	1.619	0.0012	1.438	0.0242	0.814
Total BaP-eq		6.233	100	0.247	100	0.083	100	2.973	100

\*: The individual benzofluoranthene isomer concentrations were calculated as the average of Bb<sub>k</sub>jF for the TEF approach

### 3.4. Source identification of PAHs

According to the formation mechanisms, PAHs can be classified as pyrogenic or petrogenic PAHs. Normally, petrogenic PAHs mainly derive from the leakage of crude oil and the refined products such as gasoline, diesel fuel and fuel oil from urban vehicle traffics, whereas the incomplete combustion of diesel/shale/crude oil/coal leads reflects to pyrogenic/combustion origin. Several diagnostic ratios are used to infer the possible sources in environmental samples [5, 19]. PAH isomer pairs have similar physical-chemical properties, thus they have similar dilution and distribution in particulate matter and other environmental phases even though the individual isomers may be characteristic of different sources. The PAH diagnostic ratios used in this study for source identification are summarized in Table 4.

Table 4. PAH diagnostic ratios used as source indicator

Diagnostic ratio* (References)	Sources	Ranges and mean values of diagnostic ratios in this study during different sampling periods			
		I.	II.	III.	IV.
FLT/(FLT+PYR) [19]	< 0.4 Petrogenic				
	0.4–0.5 Fossil fuel combustion	0.53–0.60	0.51–0.59	0.43–0.56	0.16–0.50
	> 0.5 Grass, wood, coal combustion	0.58	0.54	0.50	0.45
BaA/(BaA+CHR) [19]	< 0.2 Petrogenic				
	0.2–0.5 Petrogenic or combustion	0.40–0.58	0.25–0.36	0.25–0.44	0.20–0.51
	> 0.5 Combustion	0.52	0.31	0.37	0.44
IND/(IND+BghiP) [5]  [19]	0.18 Cars				
	0.37 Diesel				
	0.56 Coal				
	0.62 Wood burning				
0.35–0.70 Diesel emission	0.46–0.58	0.35–0.49	0.43–0.50	0.55–0.61	
< 0.2 Petrogenic	0.54	0.42	0.47	0.59	
0.2–0.5 Petrogenic or combustion					
> 0.5 Grass, wood, coal combustion					
FLU/(FLU+PYR) [5]	> 0.5 Diesel	0.02–0.07	0.06–0.28	0.25–0.43	0–0.05
	< 0.5 Gasoline	0.05	0.17	0.37	0.02
BaP/BghiP [5]	0.5–0.6 Traffic	0.57–1.36	0.32–1.21	0.40–1.0	1.23–2.03
		0.99	0.72	0.62	1.58
FLT/BeP [5]	3.5±0.5 Automobile exhaust	0.85–2.01	0.53–1.64	0.56–1.60	0.51–2.13
		1.37	1.25	0.92	1.52
PYR/BeP [5]	6±1 Automobile exhaust	0.66–1.38	0.47–1.44	0.50–1.60	0.62–2.41
		0.99	1.03	0.91	1.71

\*: FLT- fluoranthene, PYR- pyrene, BaA- benz[a]anthracene, CHR- chrysene, IND- indeno[1,2,3-c,d]pyrene, BghiP- benzo[g,h,i]perylene, FLU- fluorene, BaP- benzo[a]pyrene, BeP- benzo[e]pyrene

The results of concentration ratios of FLT/(FLT+PYR) and IND/(IND+benzo[g,h,i]perylene (BghiP)) suggest that the combustion of both liquid and solid fuels was the dominant source of PAH contamination in the Győr atmosphere. The ratio values of BaA/(BaA+chrysene (CHR)) show mainly mixed sources of pyrogenic combustion and petrogenic sources. However, over 80 % of all the samples collected in winter show combustion origins based on the results of BaA/(BaA+CHR) diagnostic ratio. The BaP/BghiP ratio obtained in this study indicate traffic source over 20 % of all the samples. Vehicular emissions can be derived from gasoline engines based on ratio values of fluorene (FLU)/(FLU+PYR). The IND/(IND+BghiP) ratio values indicate diesel emission. However, the FLT/benzo[e]pyrene (BeP) and PYR/BeP do not indicate automobile exhaust.

Although molecular ratios are often used for characterizing possible pollution sources, their ratios can be altered due to the reactivity of some PAH species with other atmospheric species, such as ozone and/or oxides of nitrogen. In addition to the atmospheric reactivity, degradation that may occur during the sampling process and can also modify the atmospheric PAH levels and thus the ratios between PAHs [5].

### 3.5. Comparison with other Hungarian cities

Fig. 6 illustrates that the annual mean concentration of BaP observed for Győr is comparable with published data of other Hungarian cities (see also Fig. 2) in 2011 [15]. The BaP levels almost in all presented cities exceeded the EU target value in 2011. However, the air quality for BaP in Hungary generally corresponds to the EU average [14]. The exposure value excess is due to the traffic and domestic heating, collectively [14-15].

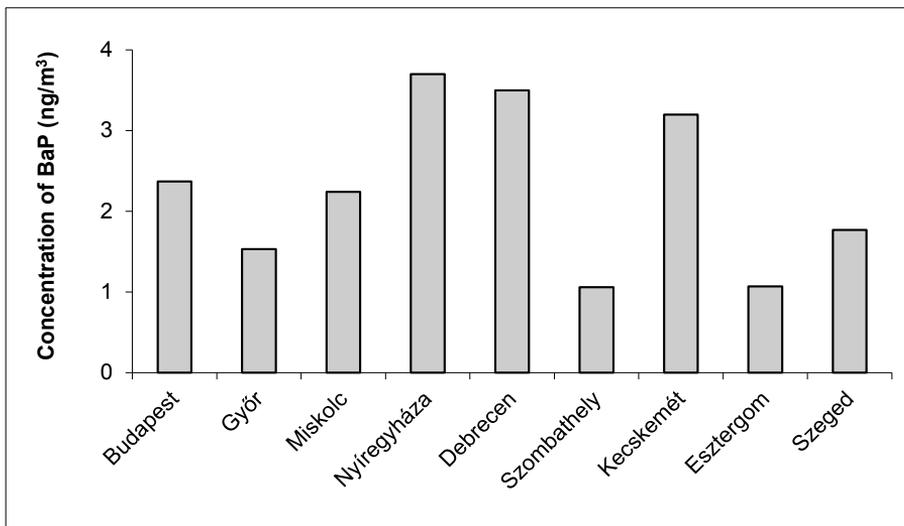


Figure 6. Annual mean concentration of benzo[a]pyrene measured in Győr and other Hungarian cities in 2011 [15]

The annual mean concentrations of some other carcinogenic PAH species (BaA, BbkjF, IND and DahA) observed for Győr were also compared with published data of other cities determined under the Hungarian PM10 Monitoring Programme [15]. The data illustrated in Fig. 7 show that the levels of carcinogenic PAHs in Győr PM10 samples were lower than measured in several other Hungarian cities. However, the national air quality database for carcinogenic PAH species is incomplete in some cities. Furthermore, there are no data for total PAHs or other individual PAH compounds. Based on TEF approach reported by Larsen and Larsen [17] on the mean concentrations of carcinogenic PAH species determined at the different Hungarian urban areas, it was found that the three highest carcinogenic exposure were in Nyíregyháza, Debrecen and Kecskemét (Table 5). Similar to the results of Győr, BaP has the highest carcinogenic potency followed by DahA.

In a previous study [20], 13 PAHs could be evaluated quantitatively in PM10 aerosol samples of Budapest (traffic-related site) in the period of 2004–2007. Similar to the results of Győr determined in our study, relatively higher concentrations of PAHs were detected in aerosol samples of Budapest during winter compared with other seasons. The trend is mainly caused by the large seasonal variation in ambient temperature and solar radiation. Additionally, the lower atmospheric mixing height in winter had significant impact on the concentrations of PAHs. The carcinogenic species together contributed 49 % of the mass of the total PAHs on average. The study has highlighted that the major source of organic compounds in the aerosol of Budapest during the study period was fossil fuel combustion from automobiles. Moreover, the PAH contribution from heating increased in winter.

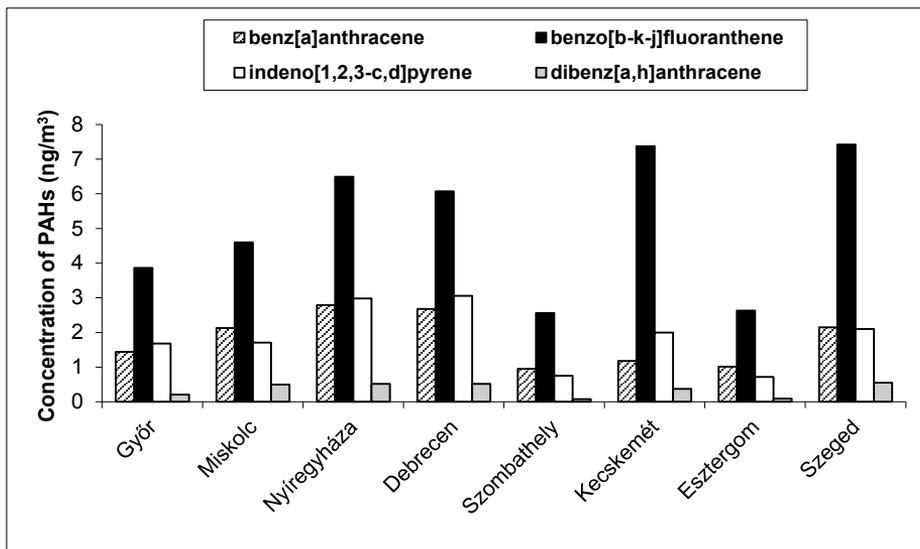


Figure 7. Annual mean concentration of some PAH compounds measured in Győr and other Hungarian cities in 2011

Table 5. BaP-eq values for the 7 carcinogenic PAH compounds in Győr and other Hungarian cities in 2011

Cities	BaP	BaA	BbF	BkF	BjF	IND	DahA	Total
	BaP-eq (ng/m <sup>3</sup> )							
Győr	1.53	0.072	0.129	0.065	0.065	0.168	0.231	2.259
Miskolc	2.24	0.107	0.153	0.077	0.077	0.171	0.550	3.374
Nyíregyháza	3.70	0.139	0.216	0.108	0.108	0.298	0.572	5.142
Debrecen	3.50	0.134	0.202	0.101	0.101	0.306	0.572	4.916
Szombathely	1.06	0.048	0.085	0.0425	0.043	0.075	0.088	1.441
Kecskemét	3.20	0.059	0.246	0.123	0.123	0.200	0.407	4.358
Esztergom	1.07	0.051	0.088	0.044	0.044	0.072	0.099	1.468
Szeged	1.77	0.108	0.247	0.1235	0.124	0.210	0.605	3.187

#### 4. Conclusions

PM10 aerosol samples were collected and concentrations of 19 individual PAH compounds were determined in an urban site of Győr during four sampling periods in 2011. The levels of total PAHs and the individual PAH compounds exhibit large variability during the sampling periods. The annual average concentration of BaP was almost 1.5 times higher than the EU target value. Relatively higher concentrations were observed in heating season. Four and five-rings PAHs including FLT, PYR and BbkjF were the dominant species in the samples. Based on calculated BaP-eq concentrations, the results show that different primary individual PAH components have significant environmental impact to BaP-eq. The most important is BaP, which component has the main effect on BaP-eq concentrations.

However, the comparison of the annual mean concentrations of BaP and other carcinogenic PAHs determined in Győr with other Hungarian cities it was found that the levels of PAHs are relatively low in the Győr atmosphere. Moreover, our results has highlighted that the future Hungarian air quality studies should be focused on source apportionment of PAHs, particularly BaP. Studies on the effect of the meteorological conditions on the PAH concentrations would also be important.

#### Acknowledgement

This work was funded by the Széchenyi István University, Hungary. We are obliged to József Erdős, István Vass, Bálint Kauker, Zsuzsanna Károly Némethné, Tünde Takács Kovácsné, Lajosné Bakódy and Péter Lautner (North Transdanubian Regional Environmental Protection and Nature Conservation Inspectorate Laboratory, Hungary) for chemical analyses, data and site information. We also thank József Erdős for field support.

## References

- [1] Kameda T: Atmospheric chemistry of polycyclic aromatic hydrocarbons and related compounds. *Journal of Health Science*, Vol. 57, No. 6, pp. 504–511, 2011.
- [2] Maliszewska-Kordybach, B: Sources, Concentrations, Fate and Effects of Polycyclic Aromatic Hydrocarbons (PAHs) in the Environment. Part A: PAHs in Air, *Polish Journal of Environmental Studies*, Vol. 8, No. 3, pp. 131–136, 1999.
- [3] Hellén H, Hakola H, Haaparanta S, Pietarila H, Kauhaniemi M: Influence of residential wood combustion on local air quality. *Science of The Total Environment*, Vol. 393, pp. 283–290, 2008.  
DOI: [10.1016/j.scitotenv.2008.01.019](https://doi.org/10.1016/j.scitotenv.2008.01.019)
- [4] Kim KH, Jahan SA, Kabir E, Brown RJC: A review of airborne polycyclic aromatic hydrocarbons (PAHs) and their human health effects. *Environment International*, Vol. 60, pp. 71–80, 2013.  
DOI: [10.1016/j.envint.2013.07.019](https://doi.org/10.1016/j.envint.2013.07.019)
- [5] Ravindra K, Sokhi R, van Grieken R: Atmospheric polycyclic aromatic hydrocarbons: Source attribution, emission factors and regulation. *Atmospheric Environment*, Vol. 42, No. 13, pp. 2895–2921, 2008.  
DOI: [10.1016/j.atmosenv.2007.12.010](https://doi.org/10.1016/j.atmosenv.2007.12.010)
- [6] Nielsen T, Jorgensen HA, Larsen JC, Poulsen M: City air pollution of polycyclic aromatic hydrocarbons and other mutagens: occurrence, sources and health effects. *The Science of the Total Environment*, Vol. 189, pp. 41–49, 1996.  
DOI: [10.1016/0048-9697\(96\)05189-3](https://doi.org/10.1016/0048-9697(96)05189-3)
- [7] Wania F, Mackay D: Tracking the distribution of persistent organic pollutants, *Environmental Science and Technology*. Vol. 30, No. 9, pp. 390–396, 1996.  
DOI: [10.1021/es962399q](https://doi.org/10.1021/es962399q)
- [8] Chang KF, Fang GC, Chen JC, Wu YS: Atmospheric polycyclic aromatic hydrocarbons (PAHs) in Asia: A review from 1999 to 2004. *Environmental Pollution*, Vol. 142, No. 3, pp. 388–396, 2006.  
DOI: [10.1016/j.envpol.2005.09.025](https://doi.org/10.1016/j.envpol.2005.09.025)
- [9] Digital Electronic AG: Digital high volume aerosol sampler. Manual, Hagenau, Switzerland, 2010.
- [10] EN 12341: Air quality - Determination of the PM10 fraction of suspended particulate matter – Reference method and field test procedure to demonstrate reference equivalence of measurement methods. 1998.
- [11] Szabó Nagy A, Csanádi Zs, Szabó J: Levels of Selected Metals in Ambient Air PM10 in an Urban Site of Győr, Hungary. *Acta Technica Jaurinensis*, Vol. 7, No. 2, pp. 146–152, 2014.  
DOI: [10.14513/actatechjaur.v7.n2.271](https://doi.org/10.14513/actatechjaur.v7.n2.271)
- [12] MSZ ISO 12884:2003: Ambient air. Determination of total (gas and particle-phase) polycyclic aromatic hydrocarbons. Collection on sorbent-baked filters with gas chromatographic/mass spectrometric analyses (in Hungarian).
- [13] 4/2011 (I.14.): Guidelines for the air load levels and the stationary point source emissions (in Hungarian).
- [14] EEA: Air quality in Europe - 2013 report, European Environment Agency, Luxembourg, 2013.

- [15] OMSZ ÉLFO: Summary of the OLM PM<sub>10</sub> sampling program in 2011, Reference Centre for Air Quality Protection, 2012 (in Hungarian).
- [16] U.S. EPA: Supplementary Guidance for Conducting Health Risk Assessment of Chemical Mixtures, U.S. Environmental Protection Agency, Washington, DC. EPA/630/R-00/002, August, 2000.
- [17] Larsen JC, Larsen PB: Chemical carcinogens, in: Air pollution and health, issues in environmental science and technology. Editors: Hester RE, Harrison RM, The Royal Society of Chemistry, Cambridge, pp. 35–65, 1998.
- [18] Boström CE, Gerde P, Hanberg A, Jernström B, Johansson C, Kyrklund T, Rannug A, Törnqvist M, Victorin K, Westerholm R: Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. *Environmental Health Perspectives*, Vol. 110, No. 3, pp. 451–488, 2002.
- [19] Yunker MB, Macdonald RW, Vingarzan R, Mitchell HR, Goyette D, Sylvestre S: PAHs in the Fraser River basin: a critical appraisal of PAH ratios as indicators of PAH source and composition. *Organic Geochemistry*, Vol. 33, No. 4, pp. 489–515, 2002.  
DOI: [10.1016/S0146-6380\(02\)00002-5](https://doi.org/10.1016/S0146-6380(02)00002-5)
- [20] Muránszky G, Óvári M, Virág I, Csiba P, Dobai R, Záray Gy: Chemical characterization of PM<sub>10</sub> fractions of urban aerosol. *Microchemical Journal*, Vol. 98, No. 1, pp. 1–10, 2011.  
DOI: [10.1016/j.microc.2010.10.002](https://doi.org/10.1016/j.microc.2010.10.002)

# Distinction of Road Categories by Road Users

G. Kosztolányi-Iván, Cs. Koren, A. Borsos

Széchenyi István University, Department of Transport Infrastructure  
Egyetem tér 1, 9026 Győr, Hungary  
e-mail: ivang@sze.hu

**Abstract:** In order to create self-explaining roads, a remarkable difference should exist between road categories, whereas within a given road category the layout should be homogenous. The paper analyses, how many and which road categories are identified and distinguished by road users. A picture sorting task was completed to find out how road users group 45 different road scenes, and how these groups correspond to road categories according current standards.

*Keywords:* road safety, road category, road scene, survey participants, driving speed

## 1. Introduction

A relatively new approach to safe road infrastructure is the self-explaining road. A self-explaining road is a road designed and built in a way that it induces adequate behaviour and thereby less driving errors are expected. Therefore, road design parameters have to be used that promote the correct behaviour of road users.

The results of our previous surveys on Hungarian roads have shown that on the usual road categories like motorways and normal two-lane primary roads the speed choice is clear for road users, i.e. these roads are self-explaining. On the other hand, there are also road categories, which are not self-explaining and therefore road users have difficulties to choose the appropriate speeds [1]. To assess the degree of uncertainty of the drivers another survey was completed. This survey of requested speeds at various road scenes has shown that in unclear situations the standard deviation of chosen speeds is higher than in unambiguous situations and the inhomogeneous distribution of driving speeds can increase the risk of accidents [2].

In the road transport system the human factor holds the central role as far as the accident causation factors are concerned [3], [4]. Cross-sectional layout is very important from the driver's point of view regarding the proper behaviour during driving (for example choosing the appropriate driving speed). Therefore, cross-sectional layout of rural roads was investigated with different mathematical and statistical tools. The researches proved that the Hungarian rural road network design should be simplified; four types of cross-sectional design would be necessary and sufficient in Hungary. With this simplification safe behaviour in traffic can be generated by the easier identification of behaviour forms required [5], [6].

## 2. Classification of roads by road users

### 2.1. Picture sorting task

Applying a method used by Weller et al. [7], a picture sorting task was performed with 104 university students. Respondents received road scene photographs in printed form with the description, which was similar to Weller’s but with changes applied. The following description was used [8] [9]:

„You are about to see 45 pictures of roads; your task is to make a useful classification of these pictures (3, 4, 5, 6, 7, 8, or 9 groups). Try to imagine yourself driving on the road and ask yourself how you would behave or which behaviour you would expect from other drivers on the same road. Sort pictures in such a way that the behaviour on the roads in a group is the same, and different from behaviour in other groups. There is no good or wrong sorting; make groups that you find useful yourself. Do this quickly, without thinking too long. You are free in choosing the number of pictures within each group and the total number of groups (between 3 and 9).

When your groups are ready, write each group (pictures marked with number on the back side) in separate columns on the other side of this paper. Write at least one - possibly more – key word for each group, which is typical for that group.”

The aim of the survey was to explore how road users classify various road types and whether this distinction corresponds to road categories from our current design guidelines.

The 45 photographs depicted traditional road types, like motorways, expressway scenes and normal primary road scenes outside urban areas. A less well-known category is a main road with elevated speed limit of 100 or 110 kmph. Some of them had physical separation between traffic directions others did not. There were also urban roads and on the border of settlements often appearing transition zone also. Road types were sorted into nine categories according to the type of separation, number of traffic lanes and speed limit. Road types and their main characteristics are shown in Fig. 1.

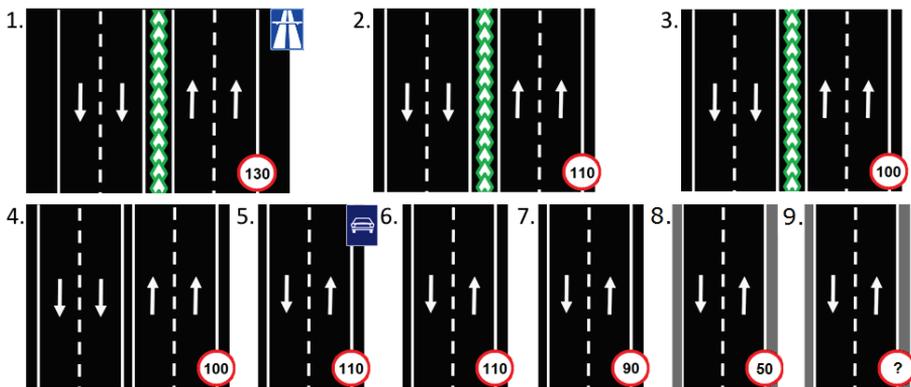


Figure 1. Schemes of the nine road categories and their posted speed limits

Table 1. shows that respondents most often formed 4 or 5 groups from road scenes.

Table 1. Number of respondents depending on the number of groups created

<i>Number of groups</i>	<i>Respondents</i>	<i>Percentage</i>
3	14	13%
4	27	<b>26%</b>
5	28	<b>27%</b>
6	20	19%
7	10	10%
8	4	4%
9	2	2%
Σ	104	100%

The most common key words used for the description regardless to the number of groups chosen, were urban area or inside built-up area, as well as the motorway or high speed.

To the group marked with the word motorway, respondents often added scenes of 2x2 lanes main roads with elevated speed limit in addition to normal motorway pictures. Reference to speed, low, medium or high speed, acceleration, reducing speed, braking or speed limit was often among the key words. Some respondents referred to safety or accident risk.

Respondents with four groups typically created two clearly separated groups: motorways and roads inside built-up area. The other two groups were variously formed; there were some who divided into good and poor pavement quality roads, while others described groups as expressway and main road. Some respondents referred to the number of traffic lanes or roadside trees as group features.

When choosing five groups, traffic volume was mentioned as key word, which was not typical for lower group numbers. Here visibility and presence of curves has also appeared as an influencing factor. Typical descriptions were: motorway - expressway - main road - minor road - built-up area, and motorway - high traffic volume - medium traffic volume - low traffic volume - built-up area. Some referred to the number of traffic lanes and distinguished transition zones from urban roads: motorway - 2x2 lane road - 2x1 lane road - road towards and leading out of city - built-up area. There were also who referred only to the driving speed: very high speed - high speed - medium speed - low speed - very low speed, and sometimes pedestrians and cyclist were also mentioned.

For persons sorting in six groups, categories according to the number of traffic lanes were typically further divided by presence or absence of physical separation between traffic directions.

Those respondents, who made seven or even more groups, often mentioned overpasses or presence or absence of emergency lane, in addition to the features mentioned above.

## 2.2. Cluster-analysis of road scene pictures

The results of the picture sorting task were summarised in a symmetrical 45x45 matrix. The elements of this similarity matrix show, how many persons have put pictures  $i$  and  $j$  in the same group. The elements in the main diagonal of this matrix are equal to the number of respondents. From this matrix, a normalised one was generated, in which all the elements in the main diagonal are equal to 1 and all other elements are between 0 and 1, showing the frequency of getting into the same group.

Figures 2-4 show details from the summarized matrix with pairs of pictures. Fig. 2 shows that 86% of the respondents put pictures 4 and 7 into the same group. Both pictures show primary rural road scenes, this grouping fits well to the real classification.

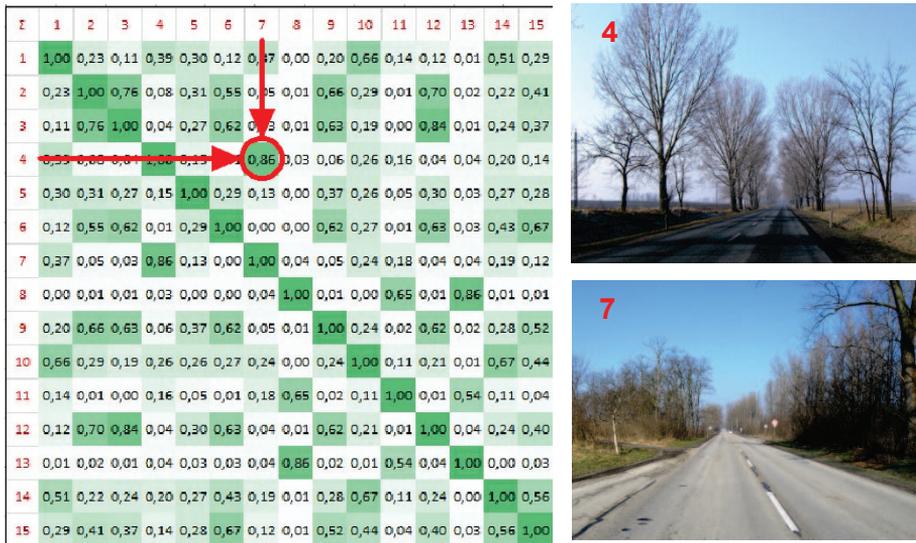


Figure 2. Detail A from the similarity matrix

In Figure 3, pictures 6 and 12 belong to different road categories: picture 6 shows a dual carriageway road with elevated speed limit of 100 kmph, while picture 12 is a motorway. Despite this difference, 63% of respondents felt that they belong together. The similarity between these categories can mislead road users.



Figure 3. Detail B from the similarity matrix

Fig. 4 shows a pair of pictures, where the cross-sections are quite different with 2x2 lanes in picture 9 and 2x1 lanes in picture 14. Despite this crucial difference, 28 percent of respondents linked these two pictures, thinking that they belong to the same category.



Figure 4. Detail C from the similarity matrix

For the further data analysis, the cluster analysis module of SPSS software was used. The method of hierarchical clustering seemed to be most appropriate.

The agglomerative algorithm is a ‘bottom up’ approach, each observation (here each picture) starts in its own cluster, and pairs of clusters are merged as one moves up the

hierarchy. At the end we get one cluster, which contains all elements. The results of hierarchical clustering are usually presented in a dendrogram, showing the merging process. If the tree is cut at a certain height, at that point the results of clustering can be interpreted.

There are various agglomerative clustering methods, this program applies the ‘average linkage clustering’ method, where the distance of two clusters is determined based on pair-wise average distance of all the elements, where one element of the pair belongs to one cluster and the other element to another cluster.

The more participants placed the given pair of pictures into the same group, the earlier these road scenes were linked in the dendrogram (Fig. 5). Therefore, the further away two pictures are from each other, the fewer participants put them into the same group.

The dendrogram showed in Fig. 5 was cut at five branches. On the horizontal axis each picture is marked with a number, on the vertical axis the rescaled distances of clusters are shown. Vertical lines show joined clusters. The position of the line on the scale indicates the distance at which clusters are joined. The observed distances are rescaled to fall into the range of 1 to 25 therefore the actual distances are not shown. However, the ratio of the rescaled distances within the dendrogram is the same as the ratio of the original distances.

From left to right, the first group contains 2x1 lane roads with elevated speed limit and 2x1 lane expressways. Roads with 2x2 lanes and elevated speed limit, without physical separation are in the second group. All 2x1 lane main rural roads belong to the third group. The fourth cluster contains roads with physical separation between traffic lanes: motorways and 2x2 lane roads with elevated speed limit. All urban roads and roads of transition zones are collected in the fifth group. This is a clear classification system.

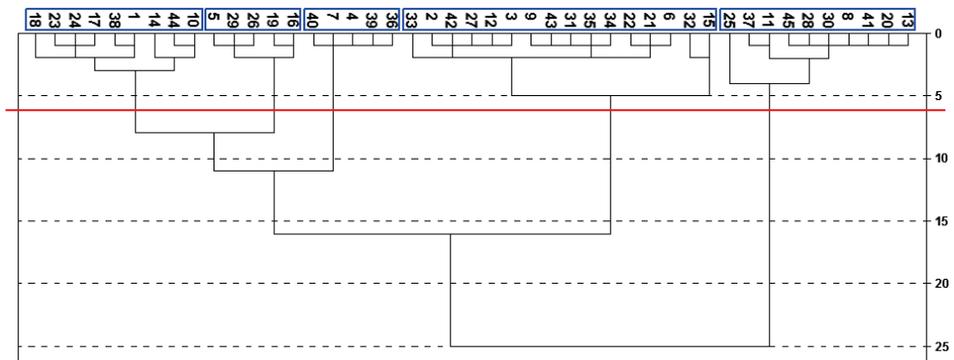


Figure 5. Dendrogram for all groups

If one stops the process at more clusters (moving the red line upwards one or two steps), no such clear classification can be found. Consequently, the picture sorting exercise shows that only 4-5 road types can be clearly distinguished by road users.

Next, some detailed results will be shown using the dendrogram, highlighting some parts of it.

### 2.3. Detailed results of the cluster-analysis

From the picture series 8-41-20-13, shown in Fig. 6, it is clear that all of these were taken on urban road sections. The classification made by the respondents is corresponding to the reality, as these images were classified into the same group by the majority of respondents and linked together in the first phase of clustering.

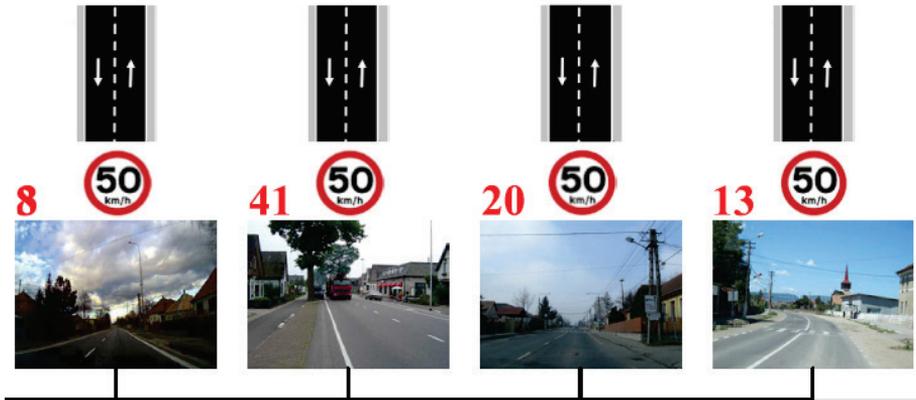


Figure 6. A detail from the dendrogram – urban road scenes correctly linked at first step

The same results can be observed for main roads outside built-up areas and for motorways. Pictures 40-7-4 (Fig. 7) were connected at the first step, as well as pictures 27-12-3 that were taken on motorways (Fig. 8).

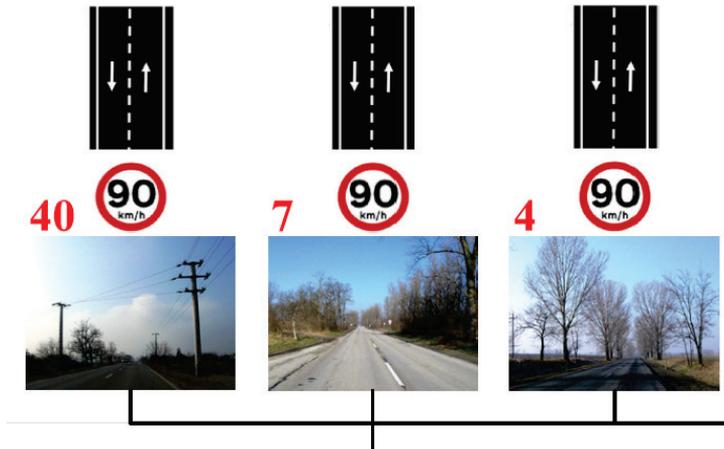


Figure 7. A detail from the dendrogram – rural main road scenes correctly linked at first step

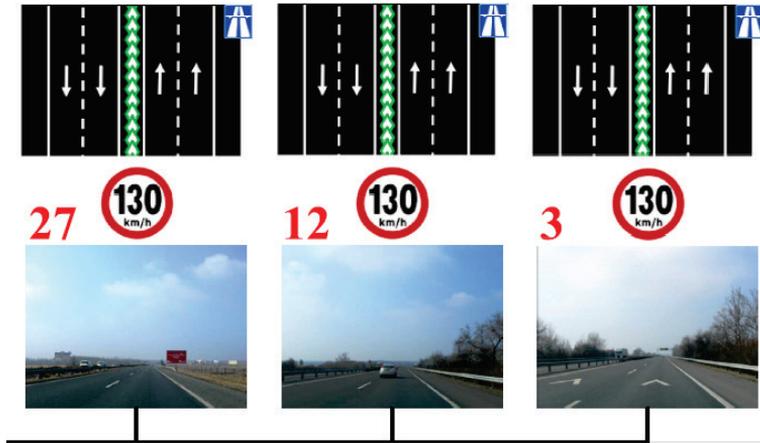


Figure 8. A detail from the dendrogram – motorway scenes correctly linked at first step

Picture group 2-42-27 (Fig. 9) was also linked at first step, that means according to the respondents they clearly belong together. Apparently this is a false classification, since the pictures 2 and 27 were actually taken on a motorway, while image 42 represents a 2x2 lane dual carriageway road with elevated speed limit of 100 kmph. So there should be noticeably 30 kmph speed difference between these two road types.

According to the similarity matrix, the ratio of placing pictures 2-27 into one group is 0.70, for pictures 2-42 it is 0.66 and for scenes 27-42 is 0.79. For these reasons, we conclude that 2/3 – 3/4 of respondents cannot distinguish 2x2 lane dual carriageway roads with elevated speed limit from motorways.

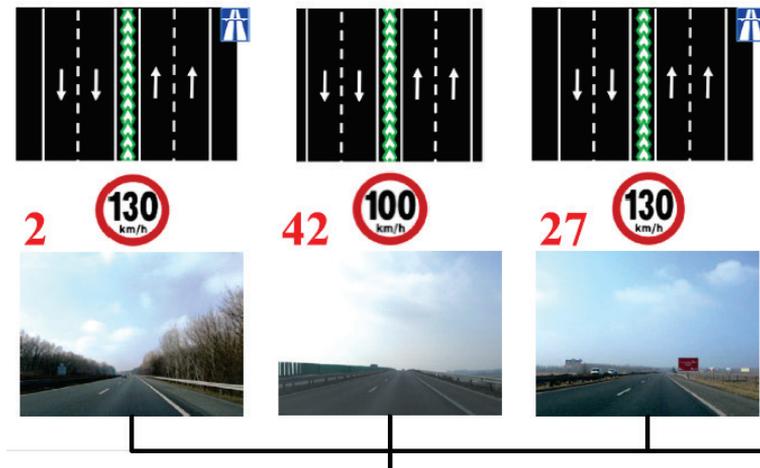


Figure 9. A detail from the dendrogram –incorrectly linked scenes at first step

The early linking of pictures 22-21-6 (Fig. 10) shows that the grouping is influenced by individual road elements. These three pictures show three different road types, but each contains a bridge over the road. These pictures were linked at the first step and merged with the group of other dual carriageway roads at the second step only. According to the similarity matrix, the ratio of placing pictures 22-21 into one group is 0.74, for pictures 22-6 is 0.76 and for scenes 21-6 is 0.81.

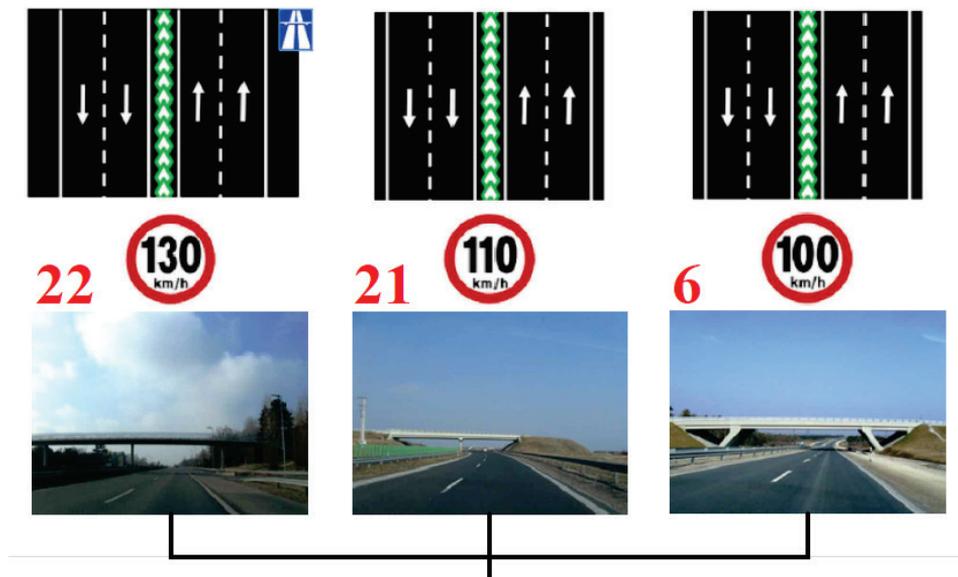


Figure 10. A detail from the dendrogram –scenes linked at first step incorrectly, due to similar road elements

The respondents get into difficult situation when grouping the scenes of transition zones; they were not able to connect these pictures with each other, as it is visible in Fig. 11. Picture 45 was placed into the group of urban roads at first step. Pictures 37 and 11 were linked with each other in the first step and then, in the second step they were merged to the urban roads group. Picture 25 was added to this mix of urban and transition road's group only at fourth step (Fig. 12.).

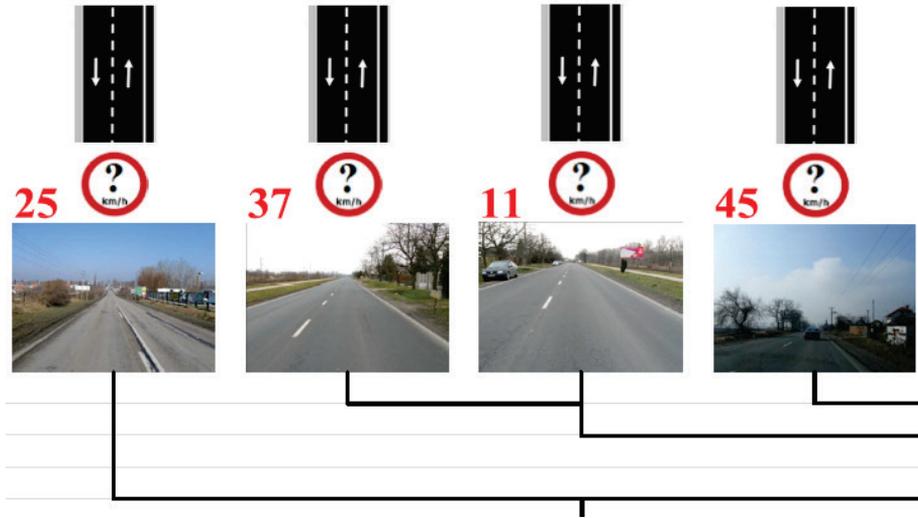


Figure 11. A detail from the dendrogram –related scenes linked at later steps

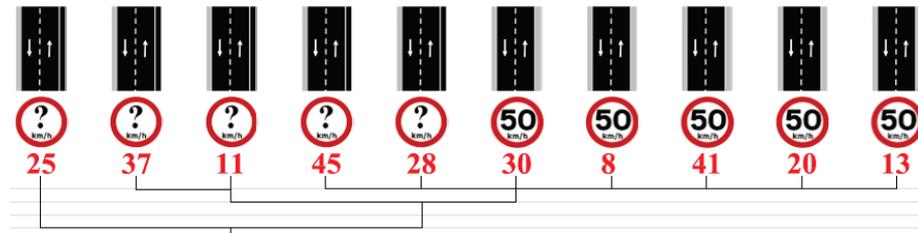


Figure 12. A detail from the dendrogram –scenes linked at later steps

### 3. Classification of roads according to speed choice

The choice of speed by drivers mostly depends on the layout and conditions of the environment of the road and the current traffic conditions on it [10], [11], [12]. Different geometric parameters of roads have different effects on vehicle speeds [13], [14], [15], [16], [17]. There are a number of researches dealing with how driving speeds affect the safety of road infrastructure [18], [19], [20], [21].

The speed choice of road users has traditionally been measured by speed cameras on the roads. Some studies apply another method: respondents had to choose driving speed according to road scene photographs, which were shown them [22], [23], [24]. Similarly to the studies mentioned above, driving speeds were studied by a questionnaire survey [25], [26].

In our questionnaire survey respondents had to review photographs of road scenes. Participants were asked to state what speed they preferred for each road scene, they were

not informed about the actual speed limit. The aim of the study was to explore how road users can recognize various types of roads.

According to the average of chosen driving speeds and the standard deviation clearly and unclearly identified roads were reported earlier [27], [28]. After that a further analysis was made [9]. Similarly to the picture-sorting task presented above, according to the speed choice exercise, road users can only distinguish 4-5 road types clearly. Additional clusters cannot be clearly linked to road types.

#### **4. Conclusions**

The results confirmed previous investigations that some road types are recognizable for road users while there are also roads that cause uncertainty; these roads are not self-explaining. As uncertainty can cause risky situations, in these sections road users should be informed with special care about their expected behaviour.

Traditional road classification distinguishes a high number of road categories. The Hungarian road design guidelines outside built-up areas define 8 different design categories and within each category there are also additional 2-3 'subcategories' distinguished [29]. Therefore about 15 different types of rural roads can be designed.

According to the cluster analysis based on chosen driving speeds, as well as in the picture sorting task, the result is that road users can clearly distinguish only 4-5 road categories. These numbers are in harmony with the new German and Dutch classifications, which are also based on this observation [30], [31], [32]. It is proposed to upgrade the Hungarian technical specifications, guidelines according to these principles.

#### **References**

- [1] Iván G, Koren Cs: Survey of free speeds on roads outside built-up areas with elevated speed limits in Hungary. *Journal of Society for Transportation and Traffic Studies*, Vol. 4, No. 2, pp. 8-17, 2013.
- [2] Iván G, Koren Cs: Recognition of built-up and non-built-up areas from road scenes, *Transport Research Arena 2014. Innovate Mobility*, Paris, France, 2014.04.14–17., Paper 18069, p. 9, 2014.
- [3] Berta T: The human as a part of the transport system (in Hungarian). *Közúti és Mélyépítési Szemle*, Vol. 57, No. 12, pp. 20–24, 2007.
- [4] Holló P: Thoughts on the role of human factors in road safety (in Hungarian). *Közúti és Mélyépítési Szemle*, Vol. 57, No. 12, pp. 25–26, 2007.
- [5] Török Á: Simplification of road transport infrastructure layout for better self-explanation. *American Journal of Vehicle Design*, Vol. 1, No. 1, pp. 16–20, 2013.
- [6] Hlédik E, Lógó E, Török Á: The assessment of cross-sectional arrangement of road traffic infrastructure with mathematical methods (in Hungarian). *Közlekedéstudományi Szemle*, Vol. 62, No. 6, pp. 15–20, 2012.
- [7] Weller G, Schlag B, Friedel T, Rammin C: Behaviourally relevant road categorisation: A step towards self-explaining rural roads. *Accident Analysis and Prevention*, Vol. 40, pp. 1581–1588, 2008.  
DOI: [10.1016/j.aap.2008.04.009](https://doi.org/10.1016/j.aap.2008.04.009)

- [8] Iván G, Koren Cs: Distinction of road classes in design guidelines and from the viewpoint of road users (in Hungarian). XVII. International Conference on Civil Engineering and Architecture, Hungarian Technical Scientific Society of Transylvania, Csíksomlyó, Romania, pp. 130–137, 13-16 June 2013.
- [9] Iván G: Distinction of road categories by road users compared to road classification in design guidelines. Pollack Periodica, Vol. 9, No. 3, pp. 23–34, 2014.  
DOI: [10.1556/Pollack.9.2014.3.3](https://doi.org/10.1556/Pollack.9.2014.3.3)
- [10] Berta T, Török Á: Layout effect of roadway on road vehicle speeds. Pollack Periodica, Vol. 4, No. 1, pp. 115–120, 2009.  
DOI: [10.1556/Pollack.4.2009.1.12](https://doi.org/10.1556/Pollack.4.2009.1.12)
- [11] Edquist J, Rudin-Brown Ch. M, Lenné M: Road design factors and their interactions with speed and speed limits. Monash University Accident Research Centre, Report No. 298, 2009.
- [12] Bella F: Driver perception of roadside configurations on two-lane rural roads: Effects on speed and lateral placement. Accident Analysis and Prevention, Vol. 50, pp. 251–262, 2013.  
DOI: [10.1016/j.aap.2012.04.015](https://doi.org/10.1016/j.aap.2012.04.015)
- [13] Ivan JN, Thomas J, Borsos A: Motor vehicle speeds: Recommendations for urban sustainability. Transportation Research Record, Vol. 2301, pp. 1–8, 2012.  
DOI: [10.3141/2301-01](https://doi.org/10.3141/2301-01)
- [14] Vörös A: Examining the driver's practice of choosing speed on some peripheral higher-speed road sections with mixed traffic (in Hungarian). Közlekedéstudományi Szemle, Vol. 59, No. 1., pp. 37–48, 2009.
- [15] Ivan JN, Garrick NW, Hanson G: Designing roads that guide drivers to choose safer speeds, Connecticut Transportation Institute of the University of Connecticut. Technical Rport, No. JHR 09-321, 2009.
- [16] Brewer J, German J, Krammes R, Movassaghi K, Okamoto J, Otto S, Ruff W, Sillan S, Stamatiadis N, Walters R: Geometric design practices for European roads, American Trade Initiatives, Technical Report, No. FHWA-PL-01-026, 2001.
- [17] Findley DJ, Hummer JE, Rasdorf W, Zegeerd Ch. V, Fowler TJ: Modeling the impact of spatial relationships on horizontal curve safety. Accident Analysis and Prevention, Vol. 45, pp. 296–304, 2012.  
DOI: [10.1016/j.aap.2011.07.018](https://doi.org/10.1016/j.aap.2011.07.018)
- [18] Mocsári T: Speed is the source of all trouble (in Hungarian). Közúti és Mélyépítési Szemle, Vol. 54, No. 9, pp. 14–18, 2004.
- [19] Mocsári T: Effect of vehicle speed on the safety of road infrastructure (in Hungarian). PhD Thesis, Multidisciplinary Doctoral School of Engineering Sciences, Széchenyi István University 2012.
- [20] Borsos A: Optimizing and modeling the safety effects of road infrastructure measures (in Hungarian). PhD Thesis, Multidisciplinary Doctoral School of Engineering Sciences, Széchenyi István University, 2010.
- [21] Gaca S: Speeds, sight distances, sight fields in road design - their impact on traffic safety (in Hungarian). Közúti és Mélyépítési Szemle, Vol. 54, No. 9, pp. 9–13, 2004.

- [22] Garrick NW: Speeds and street design results UConn and UCD. Highway design class, University Lecture, University of Connecticut, 2011.
- [23] Goldenbeld Ch., van Schagen I: The credibility of speed limits on 80 km/h rural roads: The effects of road and person(ality) characteristics. *Accident Analysis and Prevention*, Vol. 39, pp. 1121–1130, 2007.  
DOI: [10.1016/j.aap.2007.02.012](https://doi.org/10.1016/j.aap.2007.02.012)
- [24] Lahaussé JA, van Nes N, Fildes BN, Keall MD: Attitudes towards current and lowered speed limits in Australia. *Accident Analysis and Prevention*, Vol. 42, pp. 2108–2116, 2010.  
DOI: [10.1016/j.aap.2010.06.024](https://doi.org/10.1016/j.aap.2010.06.024)
- [25] Iván G: Safe speed choice from the viewpoint of road users. 8th European conference of young researchers and scientists - Transcom 2011, University of Žilina, Slovakia, 27-29 June 2011, pp. 67–70.
- [26] Iván G: Survey of free speeds on rural roads based on road scene photographs. *Pollack Periodica*, Vol. 7, No. 1, pp. 65–74, 2012.  
DOI: [10.1556/Pollack.7.2012.1.6](https://doi.org/10.1556/Pollack.7.2012.1.6)
- [27] Koren Cs, Iván G: Survey of speed choices based on photos (in Hungarian), XV. International Conference on Civil Engineering and Architecture. Hungarian Technical Scientific Society of Transylvania, Csíksomlyó, Romania, 2-5 June 2011, pp. 303–309.
- [28] Iván G, Koren Cs: Are roads with elevated speed limits self-explaining? (in Hungarian). *Közlekedésépítési Szemle*, Vol. 61. No. 5, pp. 30–36, 2011.
- [29] HRS, Road Planning Guideline (in Hungarian) No. e-UT 03.01.11, Hungarian Road Society, Ministry of Transport, Telecommunication and Energy, 2008.
- [30] Koren Cs: Classification of roads, Safety of road infrastructure (in Hungarian). Ed. by Koren Cs, Győr, Universitas-Győr Non-profit Kft, pp. 87–96, 2012.
- [31] SWOV, Fact sheet sustainable safety: principles, misconceptions, and relations with other visions. SWOV, Leidschendam, the Netherlands, 2010.  
[http://www.swov.nl/rapport/Factsheets/UK/FS\\_Sustainable\\_Safety\\_principles.pdf](http://www.swov.nl/rapport/Factsheets/UK/FS_Sustainable_Safety_principles.pdf),  
(last visited 20 December 2013)
- [32] FGSV, Richtlinien für die Anlage von Landstraßen RAL. Forschungsgesellschaft für Strassen- und Verkehrswesen, Köln, 2013.

# Toughness Improvement in Ternary HDPE/PS/PET Polymer Blends with Compatibilizer

K. Dobrowszky, F. Ronkay

Department of Polymer Engineering, Faculty of Mechanical Engineering,  
Budapest University of Technology and Economics  
Műegyetem rkp. 3, 1111 Budapest, Hungary  
e-mail: dobrowszky@pt.bme.hu

**Abstract:** Preparing polymer blends is an effective way to develop new raw materials. In the present study, the morphological and mechanical properties were investigated of ternary polymer blends. Immiscible plastics: polyethylene terephthalate (PET), high density polyethylene (HDPE) and polystyrene (PS) were blended, which can be found in post-consumer waste in large quantities and compatibilizer agents were also added to the blend. Three different processing methods, internal mixing, extrusion and injection moulding were used during the production. The results show that the best mechanical properties can be achieved if the ternary polymer blends are produced by extrusion followed by injection moulding or direct injection moulding without prior homogenization, and styrene/ethylene-butylene/styrene copolymer grafted with maleic anhydride (SEBS-g-MA) was also added to the polymer blend.

**Keywords:** ternary blend, compatibilization, toughness, scanning electron microscopy

## 1. Introduction

Preparing polymer blend – a mixture of at least two macromolecular substances – is an effective way to develop new raw materials, where the favourable properties of the polymers can be tailored for industry [1]. However, in most cases polymers are not compatible with each other, because of their high molecular weight and for thermodynamic reasons [2]. The miscibility between two phases can be described by Flory–Huggins solution theory (Eq. (1)):

$$\Delta G_m/RT = \Phi_1/V_1 \cdot \ln \Phi_1 + \Phi_2/V_2 \cdot \ln \Phi_2 + \chi'_{12} \cdot \Phi_1 \Phi_2 \quad (1)$$

where  $\Delta G_m$  is the change of Gibbs free energy,  $R$  is the universal gas constant,  $T$  is the temperature,  $\Phi_1$  and  $\Phi_2$  are the volume fractions,  $V_1$  and  $V_2$  are the molar volume of components “1” and “2”,  $\chi'_{12}$  is the dimensionless Flory-Huggins binary interaction parameter [3].

It can be also difficult to choose the suitable processing parameters for preparing blends. The improper selection of processing temperature can easily cause degradation in one of the components of blend. Polyolefins, like polyethylene (PE) and polypropylene (PP) require generally lower melting temperature during the processing than engineering plastics, like polyethylene terephthalate (PET). When two or more plastics are mixed together, usually a heterogeneous morphological structure is formed, that induces low mechanical properties due to the poor interfacial interaction between the components. To achieve good dispersion between the immiscible phases, copolymer or compatibilizer can be added to the polymer blends, in order to obtain a stable and finer morphology structure leading better impact strength properties [4]. Nowadays, nanocomposites have opened a new way in the development of polymer blends and composites. Thus, the tensile strength and Young's modulus can be further increased by using nanoclay like montmorillonite [5].

Ternary polymer blends have gained exceptional attention lately [6-8]. Razavi et al. [6] investigated binary and ternary blends, using recycled HDPE, recycled PS and recycled PET. They found the phases were incompatible with each other, but ternary blend showed a better performance in morphological characteristics and mechanical properties. However, their research did not cover the opportunities of compatibilization. Omonov et al. [7] found a great correlation between the morphology development of blends and the applied reactive compatibilization process. Yen et al. [8] compared two different processing methods to obtain the change of shell-core structure of the dispersed phases. They found, the impact property of blends strongly depends on the thickness of the additive between the phases, which can be controlled by the correct processing.

To achieve the desired mechanical properties in ternary blends, the morphology and the size of the dispersed phases have to be controlled [9]. The most influential factors, during the formation of the microstructure, are the compositional ratio, the viscosity and the interfacial bond between the phases [10]. By combining the effects of these factors, different complex morphological structures can be formed, e.g. tri-continuous phases, bi-continuous structures with one dispersed phase, or a matrix with two dispersed phases, where sometimes one of the components is encapsulated by the other ones [11]. The different morphological structures of ternary blends are presented on Fig. 1.

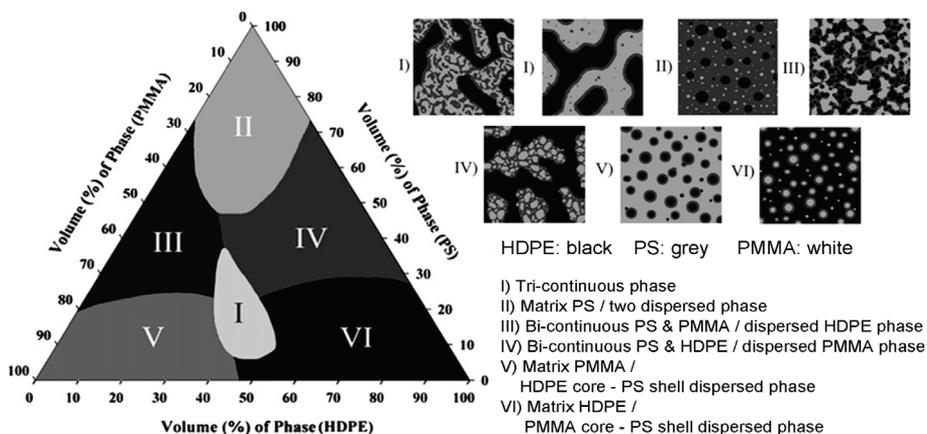


Figure 1. Different morphological states of various ternary blends [11]

Recycling is important because of social demands and regulatory requirements. Recycling of the polymers can be realized in two different ways: separating the waste stream into pure fractions [12], or using additives in order to improve the compatibility between the components [13]. Consequently, this examination of ternary blends can also promote the recycling process of polymers, particularly the area of packaging [14].

The aim of this paper is to present the effects of different blending methods and the changes in mechanical and morphological properties. It would be also important to know the effects of the prior homogenisation on morphology, tensile and impact properties of blends. In the first step, styrene/ethylene-butylene/styrene copolymer grafted with maleic anhydride (SEBS-g-MA), which additive is recommended by previous studies [15-17], was added to the ternary 70/15/15 wt% HDPE/PS/PET blends in order to improve the impact strength and the elongation at break. Another additive polypropylene grafted with maleic anhydride (PP-g-MA), which is not a recommended additive, was also added to the blends in order to investigate the potential negative effects of an improperly selected additive on tensile and impact properties of blends.

## 2. Experimental

### 2.1. Materials

Three different types of plastics – which can be found in large quantities in post-consumer waste – PET, PS and HDPE were used during the study, where (i) HDPE was Tipelin BA 550-13 (density 0.955 g/cm<sup>3</sup>; MFI=0.35 g/10 min, 190°C/2.16 kg) produced by TVK, (ii) PS was Edistir N 1840 (density 1.05 g/cm<sup>3</sup>; MFI=10.0 g/10 min, 200°C/5.0 kg) produced by Polimeri Europa and (iii) PET was NeoPET 80 (density 1.34 g/cm<sup>3</sup>; T<sub>m</sub>=248°C) produced by neogroup. Two kinds of compatibilizers were added to the ternary polymer blends in 4 wt% ratio: (i) the maleic anhydride grafted polypropylene (PP-g-MA) was Orevac CA 100 (MFI=10 g/10 min, 190°C/320 g; T<sub>m</sub>=167°C), produced by Arkema, (ii) the maleic anhydride grafted styrene/ethylene-butylene/styrene copolymer (SEBS-g-MA) was Kraton FG1901X (density 0.91 g/cm<sup>3</sup>; MFI=22 g/10 min, 230°C/5 kg).

## **2.2. Compoundation and test sample preparation**

70/15/15 wt% HDPE/PS/PET blends were prepared added with different types of compatibilizers in 4 wt%. The samples were made by three different methods: (i) premixing polymer blends with internal mixer, or (ii) using a twin screw extruder, which are followed by injection moulding after prior homogenization or (iii) making samples directly from the granules by injection moulding without prior homogenization.

PET was dried 6 hours in an air drying oven at 160°C. A Brabender Plasti-Corder PL-2000 was used to prepare the internal mixed HDPE/PS/PET ternary blends (temperature 275°C, melt-blending time 10 min, mixing speed 20 rpm), followed by a grinding stage to reach the suitable shape of blends for injection moulding.

The extrusion occurred in a Labtech Scientific LTE 26-44 twin screw extruder with a double circular cross section die with a diameter of 3 mm (zone temperature from 250°C to 275°C, rotation speed 40 rpm). The extrudates were cooled down in a water bath, finally a Labtech LZ-120/VS granulator was used to make 3 mm granules.

To prepare the injection moulded specimens with a cross section of 10x4 mm an Arburg Allrounder Advance 370S 700-290 was used, with the following parameters: the nozzle temperature was 275°C, injection flow was 20 cm<sup>3</sup>/s, injection volume was 46 cm<sup>3</sup>, holding pressure was 400±50 bar - depending on the content of the mixture, the cycle time was 50 s.

## **2.3. Test methods**

CEAST Modular Melt Flow Model 7027.000 was used to measure the viscosity in melt volume rate (MVR) of the polymers (275°C, 2.16 kg), according to ISO 1133.

The tensile tests were performed on a Zwick Z020 Universal Testing equipment with a 10 kN tensile head (test speed 20 mm/min, clamping distance 100 mm). 5 repetitions were done for each composition (according to ISO 527 standard).

Charpy impact strength was measured by a Ceast Resil Impactor Junior impact test machine, with a 15 J hammer and with an impact speed of 3.4 m/s, where the distance between the supports was 62 mm. 8 unnotched samples were measured for each composition, with a size of 80x10x4 mm (according to ISO 179-1).

JEOL JSM 6380LA scanning electron microscopy (SEM) was used at an acceleration voltage of 15 kV in secondary electron imaging mode, to study the morphological structures of the cryogenic fractured surface of blends (gold coating time was 45 s).

## **3. Results and discussion**

### **3.1. Rheology**

The viscosity of polymers can be characterized with MVR in the melt phase. As shown in Table 1. PS had the highest MVR rate, while HDPE can be characterized by low flowability during processing, because this plastic is recommended for blow technologies. This can be confirmed by the fact that the measured MVR value at 275°C is very close to the specified value by the manufacturer under the same load (but at

190°C). The exact MVR rate of PP-g-MA was difficult to determine by the applied parameters, because the MVR value was more than 100 cm<sup>3</sup>/10 min at 275°C.

Table 1. The measured MVR rates of applied plastics and SEBS-g-MA (275°C, 2.16 kg)

	HDPE	PS	PET	SEBS-g-MA
MVR [cm <sup>3</sup> /10 min]	0.95	79.01	55.09	33.07

### 3.2. Morphology investigation by SEM

The morphologies of the ternary 70/15/15 wt% HDPE/PS/PET blends, made by different processing methods, are shown on Fig. 2-4. The coarsest morphological structure occurred after internal mixing (Fig. 2.); the finest morphology formed after injection moulding because of the higher shear stress. It can be seen, that extrusion (Fig. 3.) and injection moulding (Fig. 4.) resulted in a fibrous morphology, while internal mixing resulted in an encapsulated structure between the dispersed PS and PET phases. It should be noticed, that due to repeated melting during injection moulding the previous morphology after prior homogenization was decomposed, resulting in the same morphology as presented on Fig. 4.

A finer dispersion occurred in blends containing 4 wt% SEBS-g-MA, than blends compatibilized with PP-g-MA. This phenomenon can be explained by the chemical interaction between the phases: the styrene block of SEBS-g-MA could miscible with PS, the hydrogenated ethylene-butadiene blocks react well with HDPE and the maleic anhydride groups of SEBS-g-MA could interact with hydroxyl end-groups of PET; resulting in a finer and stable morphology in HDPE/PS/PET blends.

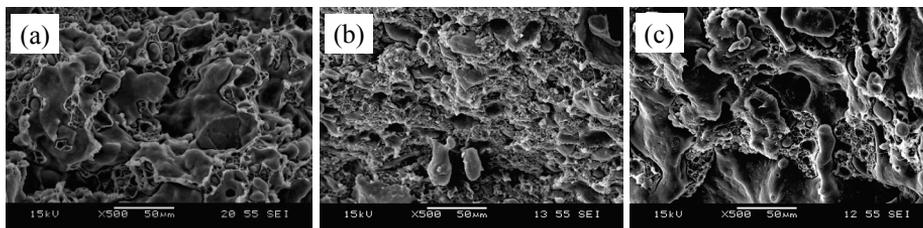


Figure 2. Morphology of internal mixing of 70/15/15 wt% HDPE/PS/PET ternary blends: (a) without compatibilizer, (b) 4 wt% SEBS-g-MA, (c) 4 wt% PP-g-MA

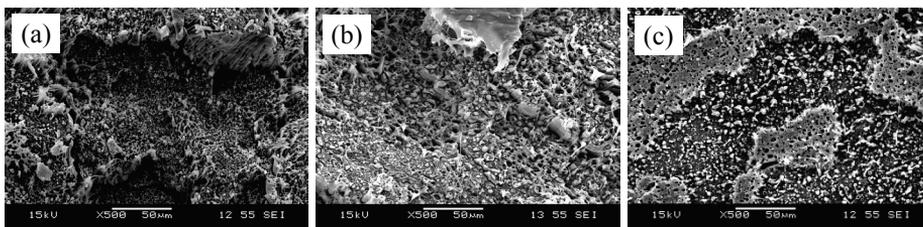


Figure 3. Morphology after extrusion of 70/15/15 wt% HDPE/PS/PET ternary blends: (a) without compatibilizer, (b) 4 wt% SEBS-g-MA, (c) 4 wt% PP-g-MA

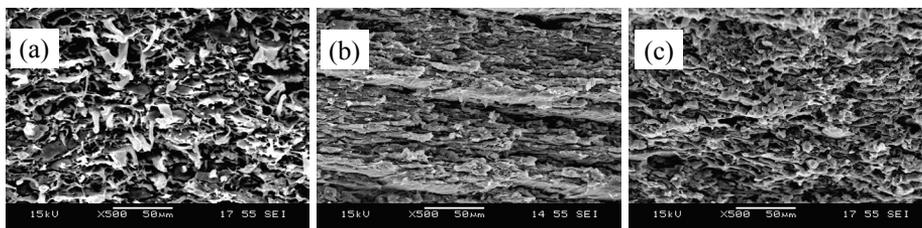


Figure 4. Morphology after injection moulding of 70/15/15 wt% HDPE/PS/PET ternary blends: (a) without compatibilizer, (b) 4 wt% SEBS-g-MA, (c) 4 wt% PP-g-MA

### 3.3. Tensile properties

Fig. 5. represents the stress-strain curves of the reference materials and the 70/15/15 wt% HDPE/PS/PET blends, using various production methods. The plastics applied in the study can be characterized by the following parameter: HDPE has the highest elongation at break, but the smallest tensile strength (25.7 MPa) and Young's modulus (0.92 GPa). The tensile strength of PET (56.5 MPa) is the highest of the applied plastics. The value of tensile strength of PS (36.7 MPa) was between PET and HDPE. PS is a rigid polymer with a significantly higher Young's modulus (1.98 GPa) than HDPE, which is comparable with the Young's modulus of PET (2.06 GPa).

Comparing the three different blend preparing methods, it can be stated, that the way of prior homogenization with internal mixer followed by injection moulding shows the smallest elongation and the lowest mechanical properties (e.g. 24-26 MPa tensile strength), regardless of whether the blends contained compatibilizer or not (see Table 2.). Presumably, this way of production (10 minutes internal mixing, followed by injection moulding) caused the highest shear forces overall, leading to a slight degradation in 70/15/15 HDPE/PS/PET blend.

Blends without compatibilizer, produced by prior extrusion followed by injection moulding, or directly injection moulded have the highest tensile strength (30.03 MPa and 30.55 MPa, respectively). There was no difference in tensile strength values among the different production methods, when SEBS-g-MA was added to blends. Nevertheless, the Young's modulus of blends was significantly higher in case of direct injection moulding. The Young's moduli of blends containing SEBS-g-MA have slightly decreased, because of the softening effect of the additive.

The elongation at break occurred between 5-10% of strain in blends without compatibilizer. The break occurred every time at higher strain, when 4 wt% SEBS-g-MA was added to the blends, while introducing PP-g-MA to the blends decreased the value of elongation at break. This suggests that PP-g-MA was not able to make a chemical interaction with the applied plastics, and when this additive is located in the boundary of the phases, the interfacial tension may be greater due to the presence of the new component.

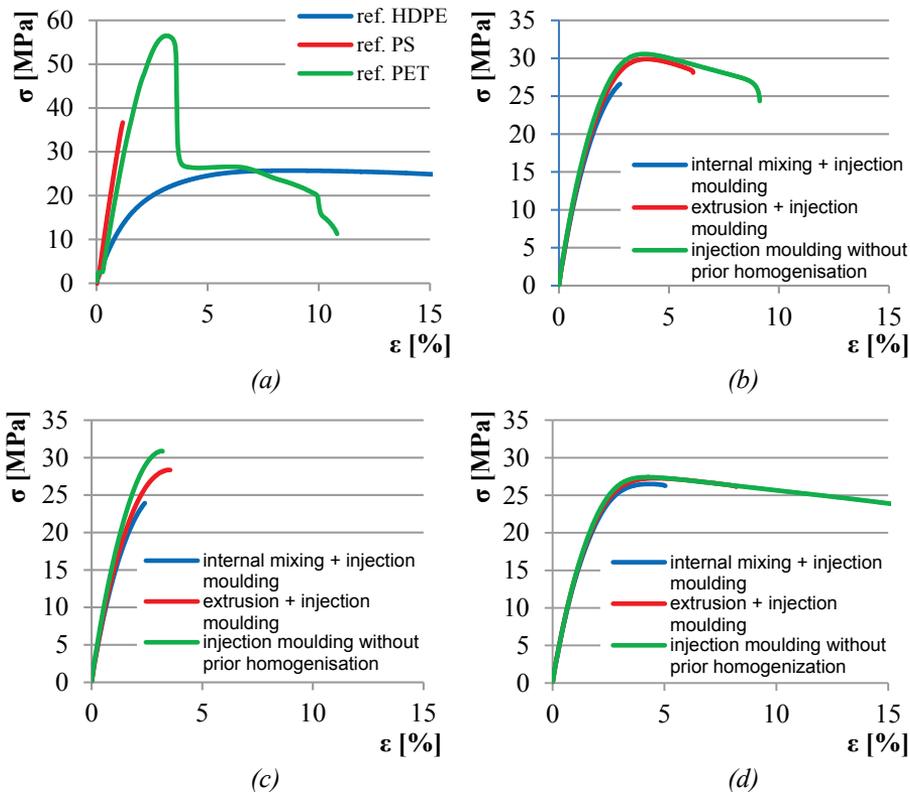


Figure 5. Stress-strain curves of reference materials and 70/15/15 wt% HDPE/PS/PET blends - produced by internal mixing followed by injection moulding, or extrusion followed by injection moulding, or direct injection moulding without previous homogenisation: (a) reference materials, (b) blends without compatibilizer, (c) blends with 4 wt% PP-g-MA, (d) blends with 4 wt% SEBS-g-MA

Despite blends consisted of 70 wt% HDPE, the tensile strength was slightly and the Young's modulus was significantly improved of blends than the values of HDPE reference, when prior extrusion followed by injection moulding or direct injection moulding were used (see Table 2.). Adding PS and PET to HDPE and forming a blend of these plastics can improve the strength and the modulus of HDPE. However, the elongation at break of the blends without compatibilizer has decreased. But in the case when 4 wt% SEBS-g-MA was added to the blends, an increase was observed in elongation at break of blends.

Table 2. Mechanical properties of 70/15/15 HDPE/PS/PET blend with two different types of compatibilizers, prepared by three different method: prior internal mixing, followed by injection moulding; prior extrusion, followed by injection moulding or injection moulding without previous homogenisation

	Tensile strength [MPa]			Young's modulus [GPa]		
	internal mixing + injection moulding	extrusion + injection moulding	injection moulding	internal mixing + injection moulding	extrusion + injection moulding	injection moulding
<b>70/15/15</b>	26.47 ± 0.27	30.03 ± 0.12	30.55 ± 0.09	1.13 ± 0.009	1.19 ± 0.010	1.44 ± 0.002
<b>70/15/15 + 4% SEBS-g-MA</b>	26.25 ± 0.35	27.54 ± 0.24	27.43 ± 0.15	1.06 ± 0.013	1.10 ± 0.009	1.23 ± 0.014
<b>70/15/15 + 4% PP-g-MA</b>	23.97 ± 0.31	28.36 ± 0.11	30.28 ± 0.66	1.12 ± 0.004	1.17 ± 0.007	1.24 ± 0.019

### 3.4. Impact strength

The Charpy impact strength of the 70/15/15 wt% HDPE/PS/PET blend is shown on Fig. 6. Comparing the mixtures, it can be stated, that blends containing PP-g-MA were more rigid, than blends without compatibilizer, and easily broke during the tests, regardless of the three ways of production of the blend. This behaviour is in good agreement with the results of tensile tests. By investigating the various processing methods, it can be stated that blends, produced by internal mixer followed by injection moulding have smaller impact strength in every case, than the samples which were produced by the two other methods.

Comparing the two different types of compatibilizers, it is visible, that the impact strength of the blends which contain SEBS-g-MA is more than three times higher, if they were produced by prior extrusion followed by injection moulding or only injection moulding, which suggests that SEBS-g-MA was able to connect with the phases of the blend. In contrast, as it was described above, the PP-g-MA might increase the interfacial tension when it is located in boundary of the blend phases. With the additive a coarse morphological structure was formed in blends – against SEBS-g-MA –, which resulted in a slight decrease of the impact strength of 70/15/15 HDPE/PS/PET blend.

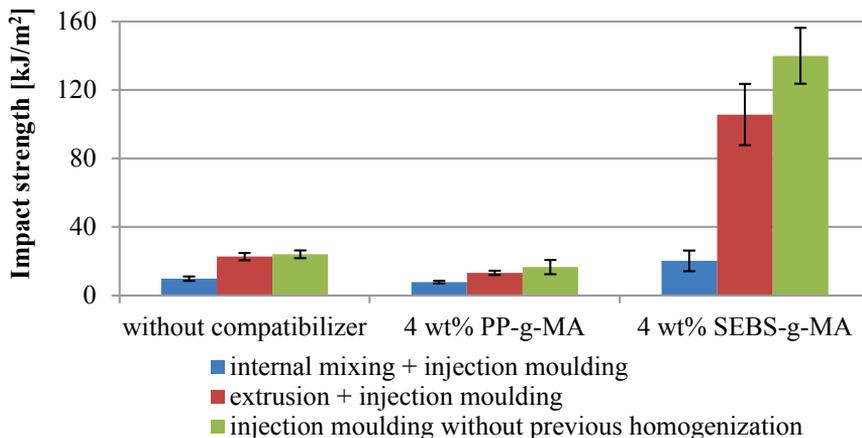


Figure 6. Charpy impact strength of 70/15/15 wt% HDPE/PS/PET ternary blend without compatibilizer, with 4 wt% PP-g-MA and 4 wt% SEBS-g-MA compatibilizer; produced by different methods: internal mixing followed by injection moulding, or extrusion followed by injection moulding, or injection moulding without previous homogenisation

#### 4. Conclusion

In this research the mechanical and morphological properties of the 70/15/15 wt% HDPE/PS/PET blend were investigated, when various processing methods and compatibilizers were applied. The tensile strength and Young's modulus of HDPE can be improved, if PS and PET were added to it. Based on the research, the best mechanical properties of the blend can be achieved, when the blends were prepared with extrusion followed by injection moulding or direct injection moulding, and 4 wt% SEBS-g-MA had been added to the blend. In case of SEBS-g-MA contained HDPE/PS/PET blend the fracture occurred at higher elongation, and the impact strength was also increased in a large extent, without significant decrease of tensile strength and Young Modulus. It has been also demonstrated, that improper selection of additives could reduce the elongation at break and impact properties of blends, without improving other mechanical properties. The results can provide a good guidance for industrial utilisation.

#### Acknowledgement

This research was realized in the frames of TÁMOP 4.2.4. A/1-11-1-2012-0001 "National Excellence Program - Elaborating and operating an inland student and researcher personal support system". The project was subsidized by the European Union and co-financed by the European Social Fund. The infrastructure of the research project was supported by the Hungarian Scientific Research Fund (OTKA K109224).

## References

- [1] Al-Salem SM, Lettieri P, Baeyens J: Recycling and recovery routes of plastic solid waste (PSW): A review. *Waste Management*, Vol. 29, No. 10, pp. 2625-2643, 2009.  
DOI: [10.1016/j.wasman.2009.06.004](https://doi.org/10.1016/j.wasman.2009.06.004)
- [2] Sperling LH: *Introduction to physical polymer science*. 4th ed. John Wiley & Sons Inc., Hoboken, 2006.
- [3] Manias E, Utracki LA: Chapter 2 - Thermodynamics of polymer blends. in *Polymer blends handbook*. Eds: Utracki LA, Wilkie CA, Springer, Dordrecht, Vol 1 (2nd ed), 171-289, 2014.
- [4] Li Y-Y, Hu S-W, Sheng J: Evolution of phase dimensions and interfacial morphology of polypropylene/polystyrene compatibilized blends during mixing. *European Polymer Journal*, Vol. 43, No. 2, pp. 561-572, 2007.  
DOI: [10.1016/j.eurpolymj.2006.10.018](https://doi.org/10.1016/j.eurpolymj.2006.10.018)
- [5] Hargitai H, Ibriksz T, Stifter J, Andersen E: Development of PA6/HDPE nanocomposite blends. *Materials Science Forum*, Vol. 729, pp. 216-221, 2013.  
DOI: [10.4028/www.scientific.net/MSF.729.216](https://doi.org/10.4028/www.scientific.net/MSF.729.216)
- [6] Razavi S, Shojaei A, Bagheri R: Binary and ternary blends of high-density polyethylene with poly(ethylene terephthalate) and polystyrene based on recycled materials. *Polymers for Advanced Technologies*, Vol. 22, No. 5, pp. 690-702, 2011.  
DOI: [10.1002/pat.1567](https://doi.org/10.1002/pat.1567)
- [7] Omonov TS, Harrats C, Groeninckx G: Co-continuous and encapsulated three phase morphologies in uncompatibilized and reactively compatibilized polyamide 6/polypropylene/polystyrene ternary blends using two reactive precursors. *Polymer*, Vol. 46, No. 26, 12322-12336, 2005.  
DOI: [10.1016/j.polymer.2005.10.022](https://doi.org/10.1016/j.polymer.2005.10.022)
- [8] Yin B, Li L-P, Yan Z, Gong L, Yang M-B, Xie B-H: Largely improved impact toughness of PA6/EPDM-g-MA/HDPE ternary blends: The role of core-shell particles formed in melt processing on preventing micro-crack propagation. *Polymer*, Vol. 54, No. 7, pp. 1938-1947, 2013.  
DOI: [10.1016/j.polymer.2013.02.001](https://doi.org/10.1016/j.polymer.2013.02.001)
- [9] Fayt R, Jerome R, Teyssié P: Molecular design of multicomponent polymer systems, 13. Control of the morphology of polyethylene/polystyrene blends by block copolymers. *Die Makromolekulare Chemie*, Vol. 187, No. 4, pp. 837-852, 1986.  
DOI: [10.1002/macp.1988.021890726](https://doi.org/10.1002/macp.1988.021890726)
- [10] Imre B, Renner K, Pukánszky B: Interactions, structure and properties in poly(lactic acid)/thermoplastic polymer blends. *eXPRESS Polymer Letters*, Vol. 8, No. 1, pp. 2-14, 2014.  
DOI: [10.3144/expresspolymlett.2014.2](https://doi.org/10.3144/expresspolymlett.2014.2)
- [11] Ravati S, Favis BD: Morphological states for a ternary polymer blend demonstrating complete wetting. *Polymer*, Vol. 51, No. 20, pp. 4547-4561, 2010.  
DOI: [10.1016/j.polymer.2010.07.014](https://doi.org/10.1016/j.polymer.2010.07.014)

- [12] Dobrowszky K, Ronkay F: Alternative polymer separation technology by centrifugal force in a melted state. *Waste Management*, Vol. 34, No. 11, pp. 2104-2112, 2014.  
DOI: [10.1016/j.wasman.2014.05.006](https://doi.org/10.1016/j.wasman.2014.05.006)
- [13] Pötschke P, Paul DR: Formation of Co-continuous Structures in Melt-Mixed Immiscible Polymer Blends. *Journal of Macromolecular Science, Part C: Polymer Reviews*, Vol. 43, No. 1, pp. 87-141, 2003.  
DOI: [10.1081/MC-120018022](https://doi.org/10.1081/MC-120018022)
- [14] Wang D, Li Y, Xie X-M, Guo B-H: Compatibilization and morphology development of immiscible ternary polymer blends. *Polymer*, Vol. 52, No. 1, pp. 191-200, 2011.  
DOI: [10.1016/j.polymer.2010.11.019](https://doi.org/10.1016/j.polymer.2010.11.019)
- [15] Pracella M, Rolla L, Chionna D, Galeski A: Compatibilization and properties of poly(ethylene terephthalate)/polyethylene blends based on recycled materials. *Macromolecular Chemistry and Physics*, Vol. 203, No. 10-11, pp. 1473-1485, 2002.  
DOI: [10.1002/1521-3935\(200207\)203:10/11<1473::AID-MACP1473>3.0.CO;2-4](https://doi.org/10.1002/1521-3935(200207)203:10/11<1473::AID-MACP1473>3.0.CO;2-4)
- [16] Sahnoune F, Lopez Cuesta JM, Crespy A: Improvement of the mechanical properties of an HDPE/PS blend by compatibilization and incorporation of CaCO<sub>3</sub>. *Polymer Engineering & Science*, Vol. 43, No. 3, pp. 647-660, 2003.  
DOI: [10.1002/pen.10053](https://doi.org/10.1002/pen.10053)
- [17] Carté TL, Moet A: Morphological origin of super toughness in poly(ethylene terephthalate)/polyethylene blends. *Journal of Applied Polymer Science*, Vol. 48, No. 4, pp. 611-624, 1993.  
DOI: [10.1002/app.1993.070480405](https://doi.org/10.1002/app.1993.070480405)

# On Wavelet Based Modeling of the Nitrogen Oxides Emission and Concentration due to Road Traffic in Urban Environment

Sz. Nagy

Széchenyi István University, Department of Telecommunications,  
Egyetem tér 1., H-9026 Győr, Hungary  
e-mail: nagysz@sze.hu

**Abstract:** Differential equations can be solved wavelet-based by representing the continuous functions by their wavelet expansion coefficients and thus the corresponding differential equations are transformed to matrix equations. The wavelet basis functions are organized into resolution levels of different frequency terms at different locations, and the main advantage of the wavelet expansion representation is that the wavelet based differential equation solving methods can be adaptive, it is possible to refine the solution locally, if the precision is not sufficient at some regions.

In case of the nitrogen oxides convection-advection equation, the urban environment should be taken as special material parameter in the differential equation's operator, and the matrix elements of the differential operator has to be calculated in a non-continuous environment, and the obstacles are placed so, that they are not at the boundaries of the support of the wavelets.

**Keywords:** *nitrogen oxides, wavelets, matrix elements, convection-advection, differential equation*

## 1. Introduction

Multiresolution analysis (MRA) or the wavelet analysis (the wavelet theory's basics are precisely summarized e.g., in [1, 2]) is a widely used tool of data and image processing, and it also has increasing share within differential equation solving methods. However, in modeling pollutant's convection, advection and dispersion they are only sporadically used [3] even though they have the natural ability to use different scales simultaneously.

The passengers and other participants of the urban traffic are exposed to the pollutants exhausted by the vehicles, and several of these pollutants are toxic, carcinogenic, or at least maleficent for living beings. The pollutants in the air are monitored regularly, but usually on only a few points per city and few times per day. Of course, these data are also valuable

but higher spatial and temporal resolution would be more desirable, especially because of epidemiological reasons, like monitoring the effect of the pollutant concentration on respiratory organ, or cardiovascular problems [4, 5].

The combination of the measured data with computational modeling is a recent trend in approximating the concentration of the pollutants [8, 9] – like one of the most dangerous group, the nitrogen oxides –, and the used models can be grouped into two subgroups, the large-scale models tend to apply grids with grid distance of tens of kilometers [6], while the small-scale ones can not treat larger scale tendencies [7]. Crooks and Isakov use wavelets to combine the two scales [3] based on the scalability of the wavelets: that the resolution level can be different at different locations, if the solution function contains higher frequency terms in one place and restricted to lower frequencies at other places, like the urban and rural environment. This property is also the main reason of their applications in image processing and compressing, like in the JPEG2000 standard [10], or the NASA's ICER [11]. However, none of the models in [3] calculate on wavelet basis, just the combination scheme is wavelet based.

This article shortly introduces the wavelet analysis in the next section, then derives the advection-diffusion equation's discretized form, and as the last step, gives calculation rules for determining the discretized form in urban environment. The goal is not to solve the diffusion equation in wavelet basis, it can be found in many applications, like [12–14], but to give a solution for modeling the urban environment – mainly the position of the buildings and roads, the various wind speed, – in formulating the discretized matrix equations and calculating the matrix elements. For a proper modeling of the obstacles locally different diffusion coefficients would be necessary, and these coefficients can be well approximated as step functions of the position. The matrix elements generated from step functions are calculated in this contribution.

## **2. On wavelet analysis**

Wavelets, the basis functions of MRA, are localized functions, they arise as dilations and translations of one common mother wavelet, moreover they can be used as simple building elements in expansion of continuous functions and they can describe the different features of different scales and positions well. Although wavelet-based methods, similarly to the finite elements solving schemes, are also members of the Galerkin-type solver family [15], the discretization technique in this case is more flexible, it can be systematically adapted to the problem during the solution itself, no previous knowledge or guess is necessary about the possible solution and its detailed and roughly representable regions. Wavelets are the basis functions of the system, they form an infinite series of refining resolution levels, each doubling the frequency resolution.

The effectiveness of the MRA technique lies in the fact, that in higher resolution levels most of the wavelets are unnecessary for sufficiently precise resolution, most of the

expansion coefficients are zero in these levels and thus they should not be stored and should not be included into the calculations [13, 16–21]. Calculating these coefficients and discarding them would be the first approximation, however, it is a waste of time and capacity if a method can be derived that predicts, which coefficients are the most important ones, and which are almost zero. In order to ensure the adaptivity and economy, a clear, quick, easily calculable prediction algorithm is necessary to decide which wavelets of the next resolution level are needed and which can be neglected, and such algorithms exist both for elliptical differential equations with sources [22] and for eigenvalue type differential equations [23]. In these applications, starting from an already calculated resolution level, the authors produce an economic prediction algorithm for determining whether a given wavelet in a given position is necessary for the sufficiently precise next level solution or it can be neglected. The method can be used for estimating the error of the given resolution solution.

## 2.1. Multiresolution analysis

In multiresolution analysis (MRA) or wavelet analysis the studied Hilbert space is decomposed into infinite resolution level subspaces, each one embedded to the higher resolution level subspaces,  $V_m \subset V_{m+1}$   $\{V_m, m \in \mathbb{Z}\}$ . The basic resolution level  $V_0$  is spanned by a single function, a so called *mother scaling function*  $\phi(x)$ , all the basis functions in the subspace are shifted versions of  $\phi(x)$  on an equidistant grid as  $\phi_\ell(x) = \phi(x - \ell), \ell \in \mathbb{Z}$ .

The higher resolution level mother scaling functions are generated from the basic resolution level one as  $\phi(2^m x)$ , and the basis functions are  $\phi_{m,\ell}(x) = \phi(2^m x - \ell)$ , which means that the grid is shrunk, as well. A very important property of the scaling functions is that the finer resolution level subspace  $V_{m+1}$  contains the lower resolution level subspaces, like  $V_m$ , thus all the functions in the rougher resolution level subspace – as the mother scaling function itself – can be expanded at the finer resolution subspace. This expression of the mother scaling function as a linear combination of the next level basis functions

$$\phi(x) = 2^{1/2} \sum_{i=0}^{N_s} h_i \phi(2x - i), \quad (1)$$

is called the *refinement equation*, and it is one of the basic equations of wavelet analysis. The expansion coefficients  $h_i$  determine the scaling function's shape, with the normalization condition  $\sum_{i=0}^{N_s} h_i = 1$ , and the number  $N_s$  gives the support of the mother scaling function, i.e., the support is  $[0, N_s)$ . In most of the cases, the larger support means smoother scaling functions, e.g., in the Daubechies scaling function family, the first one has  $N_s = 1$  and it is a step function, the second one with  $N_s = 3$  is everywhere continuous, but not everywhere differentiable, the third one is differentiable ( $N_s = 5$ ), the fifth is twice differentiable ( $N_s = 9$ ), etc.

A function, like the concentration of the nitrogen oxides in the air can be expressed – or usually just approximated – at resolution level  $M$  as

$$C^{[M]}(x) = \sum_{\ell \in \Omega_M} c_{M\ell} \phi_{M\ell}(x), \quad (2)$$

where  $c_{M\ell}$  are the expansion coefficients, i.e. the discretized version of the concentration, and all the indices, variables can be in one two or three dimensions – in higher dimensions, a scaling function can be e.g.,  $\phi_{m_1, m_2, \ell_1, \ell_2}(x_1, x_2) = \phi_{m_1, \ell_1}(x_1) \cdot \phi_{m_2, \ell_2}(x_2)$ , but other, non diadic product higher dimensions scaling functions exist, too, though very rarely used. The domain  $\Omega_M$  where the indices run is connected to the support of the function  $C(x)$  and the scaling functions and the number of its elements approximately doubles at each resolution level increment.

## 2.2. Wavelets and details

Wavelets are basis functions of the subspaces given as a difference between two neighboring resolution level scaling function subsets,

$$V_{m+1} = V_m \oplus W_m \quad (3)$$

These subsets, the so called *detail spaces*, are also generated from one single *mother wavelet*  $\psi(x)$ , similarly to the scaling functions, as  $\psi_{m,\ell}(x) = \psi(2^m x - \ell)$ .

The wavelets can also be expanded at the higher resolution level subspaces, like

$$\psi(x) = 2^{1/2} \sum_{i=0}^{N_s} (-1)^i h_{-i+1}^* \phi(2x - i), \quad (4)$$

with the same coefficients  $h_i$  as in (1). Here, the sign \* means complex conjugation. As a result of introducing the wavelets, a function  $C(x)$  can be approximated at resolution level  $M$  not only as (2), but also as

$$C^{[M]}(x) = \sum_{\ell \in \Omega_0} c_{0\ell} \phi_{0\ell}(x) + \sum_{m=0}^{M-1} \sum_{\ell \in \Omega_m} d_{m\ell} \psi_{m\ell}(x). \quad (5)$$

Here  $\Omega_m$  contains all the wavelets of resolution level  $m$  that overlap with the support of the function  $C(x)$ .

Theoretically the number of coefficients to be treated in (2) and in (5) are the same, but most of the coefficients  $d_{m,\ell}$  are very small and can be neglected. Smooth functions can be approximated very precisely in low resolution levels, only those parts of the functions need higher resolution, where the function varies rapidly, or has derivative discontinuities.

Of course, the sets of coefficients  $c_{M,\ell}$  and  $c_{0,\ell}$  with  $d_{m,\ell}$  can be transformed into each other using equations (1), (4) and their inverse.

### 3. The advection diffusion equation for nitrogen oxides

Nitrogen oxides are usually from high temperature combustion, in case of urban environment, the traffic and the combustion engines are the main source of the nitrogen monoxide, NO. As NO is a free radical, having one unpaired electron, it is easily oxidized in the air to nitrogen dioxide, NO<sub>2</sub>, that forms the infamous brownish dome above larger cities, and it can be further oxidized to acids, or other compounds depending on the other pollutants.

Air quality modeling, especially the modeling of the nitrogen oxides concentration, where several gases transform to one another can be carried out using weather conditions – like the moisture and the wind speed –, average vehicle count, etc. for describing these phenomena convection–advection equation in 2D along the coordinate  $x$  [24,25]

$$\left( \frac{\partial}{\partial t} - \mu \frac{\partial^2}{\partial x^2} + u \frac{\partial}{\partial x} - \sigma \right) C(x,y,t) = F(x,y,t), \quad (6)$$

is a good approximation. Here  $\mu$  is the diffusion coefficient from Fick's law,  $u$  is the velocity of the wind in the direction of  $x$ ,  $\sigma$  is the constant that covers the concentration changes due to chemical reactions, and  $F(x,y,t)$  describes the sources of the pollutants.

It is possible to use the three dimensional version of the above equations, with proper height and 3D wind and drift velocities, but as a first step, for demonstration, Eq. (6) is sufficient.

#### 3.1. Discretization of the equation

As a first step, Eq. (6) is discretized in time [22], i.e., it is approximated with a Crank-Nicholson finite difference equation. This scheme can be used in wavelet-based partial differential equation solvers, if not an eigensolution, but a time variation is needed, and there is a set of initial conditions. The boundary conditions are usually given in Dirichlet scheme, but in case of drifts – like the pollutants in the wind – the Neumann boundary conditions can be useful as well. The wavelet based solutions usually work well periodic boundary conditions.

Then both the known source and the unknown concentration should be expanded at the basic resolution level. As the basic scale can be chosen arbitrarily, without the loss of generality, we can select 0 as the basic level, thus

$$C^{[0]}(x) = \sum_{\ell \in \Omega_0} c_{0\ell}^{[0]} \phi_{0\ell}(x), \quad (7)$$

$$F^{[0]}(x) = \sum_{\ell \in \Omega_0} f_{0\ell}^{[0]} \phi_{0\ell}(x). \quad (8)$$

The notation  $c_{0\ell}^{[0]}$  and  $f_{0\ell}^{[0]}$  can be introduced for the vector of the expansion coefficients of the concentration and the source, respectively. The discretization step goes as follows. If

in the finite time differences counterpart of Eq. (6) the above formulae are applied and the equation is multiplied by an arbitrary scaling function, or by an arbitrary linear combination of the scaling function, it should remain valid, according to the weak formulation of the discretized equation. According to the previous statement, if the inner product  $\langle \cdot, \cdot \rangle$  is introduced, and the differential operator is denoted by  $\hat{D}$ , the time-discretized differential equation can be written as

$$\langle \phi_{0,k}, \hat{D}c^{[0]} \rangle = \langle \phi_{0,k}, f^{[0]} \rangle \quad (9)$$

for all basis functions  $\phi_{0,k}$  of the subspace  $V_0$  – of course, only those scaling functions are to be taken into account that overlap with the solution, i.e., the elements of  $\Omega_0$ . The result is a matrix equation

$$D_{k,\ell}^{[0]} c_\ell^{[0]} = f_k^{[0]} \quad (10)$$

with the source vector

$$f_k^{[0]} = \langle \phi_{0,k}, f^{[0]} \rangle = \int \phi_{0,k}^*(x) f^{[0]}(x) dx \quad (11)$$

and the stiffness matrix

$$D_{k,\ell}^{[0]} = \langle \phi_{0,k}, \hat{D}\phi_{0,\ell} \rangle = \int \phi_{0,k}^*(x) \cdot \hat{D}\phi_{0,\ell}(x) dx. \quad (12)$$

If higher resolution terms are also necessary for the precision, not only scaling function matrix elements, but wavelet matrix elements are also necessary. The expansion of the concentration and the source function at maximum resolution level  $M$  is given as

$$C^{[M]}(x) = \sum_{\ell \in \Omega_0} c_{0\ell}^{[M]} \phi_{0\ell}(x) + \sum_{m=0}^{M-1} \sum_{\ell \in \Omega_m} d_{m\ell}^{[M]} \psi_{m\ell}(x), \quad (13)$$

$$F^{[M]}(x) = \sum_{\ell \in \Omega_0} f_{0\ell}^{[M]} \phi_{0\ell}(x) + \sum_{m=0}^{M-1} \sum_{\ell \in \Omega_m} g_{m\ell}^{[M]} \psi_{m\ell}(x), \quad (14)$$

resulting in the  $M$ th level matrix equation

$$\begin{pmatrix} \langle \phi_{0,k}, \hat{D}\phi_{0,\ell} \rangle & \langle \phi_{0,k}, \hat{D}\psi_{n,\ell} \rangle \\ \langle \psi_{m,k}, \hat{D}\phi_{0,\ell} \rangle & \langle \psi_{m,k}, \hat{D}\psi_{n,\ell} \rangle \end{pmatrix} \cdot \begin{pmatrix} c_{0,\ell}^{[M]} \\ d_{n,\ell}^{[M]} \end{pmatrix} = \begin{pmatrix} f_{0,k}^{[M]} \\ g_{m,k}^{[M]} \end{pmatrix}. \quad (15)$$

Here the notation

$$D_{\mu,\nu}^{[M]} = \langle \xi_{m,k,t}, \hat{D}\xi_{n,\ell,s} \rangle \quad (16)$$

can also be introduced with the three-element indices  $\mu = \{m,k,t\}$  and  $\nu = \{n,\ell,s\}$ , where  $t$  and  $s$  denotes the type of the basis function  $\xi$  which can be either wavelet, or scaling function.

### 3.2. Matrix elements of resolution level $M = 0$

The above matrix elements  $D_{k,\ell}^{[0]}$ , and  $D_{\mu,\nu}^{[M]}$  can be calculated using the refinement equation (1) and the wavelet's expansion equation (4). No numerical integration is necessary, if the operator  $\hat{D}$  contains only the following types of terms: differentiations according to the space variable – as long as the scaling functions can be differentiated as many times as necessary –, multiplications with any positive power of the space variable, or any combination of the two [23, 26].

Substituting the refinement equation into (12), straightforward, but lengthy calculations lead to an eigenvalue equation in case of  $\hat{D}$  being either a derivative operator or a product with a power of the space variable. As an example, for the first derivative's matrix element the refinement equation for the differentiated scaling function is necessary, which differs from (1) only by a factor of 2, i.e.,

$$\frac{\partial}{\partial x}\phi(x) = 2 \cdot 2^{1/2} \sum_{i=0}^{N_s} h_i \frac{\partial}{\partial x}\phi(2x - i). \quad (17)$$

After substituting (1) and (4) into the formula

$$\langle \phi_{0,0}, \frac{\partial}{\partial x}\phi_{0,\ell} \rangle = \int (\phi(x))^* \frac{\partial}{\partial x}\phi(x - \ell) dx, \quad (18)$$

changing the integral variable then the summation index, the matrix element turns into

$$\langle \phi_{0,0}, \frac{\partial}{\partial x}\phi_{0,\ell} \rangle = 2 \sum_{k=0}^{N_s} \sum_{i=2\ell-k}^{N_s+2\ell-k} h_k^* h_{i+k-2\ell} \langle \phi_{0,0}, \frac{\partial}{\partial x}\phi_{0,k} \rangle, \quad (19)$$

which is clearly an eigenvalue equation for the eigenvalue 0.5 of the matrix

$${}^1M_{\ell k} = \sum_i h_i^* h_{k+i-2\ell}. \quad (20)$$

For the second derivative the eigenvectors corresponding to eigenvalue 0.25, for the third derivative, the eigenvectors corresponding to eigenvalue 0.125, etc. are giving the matrix elements, if they exist.

Note, that the matrix element (18) depends only on the difference of the indices, i.e.,

$$\langle \phi_{0,j}, \frac{\partial}{\partial x}\phi_{0,j+\ell} \rangle = \langle \phi_{0,0}, \frac{\partial}{\partial x}\phi_{0,\ell} \rangle. \quad (21)$$

For the operators that contain products with  $x^p$ , similar considerations lead to a set of iterative matrix equations, where the product with  $x^p$  depends on the product with

$x^{p-1}, x^{p-2}, \dots$  [23]

$$\langle \phi_{0,0}, x^p \cdot \phi_{0,\ell} \rangle = 2^{-p} \sum_{q=0}^p \binom{p}{q} \sum_{k=0}^{N_s} \sum_{i=2\ell-k}^{N_s+2\ell-k} k^q \cdot h_k^* h_{i+k-2\ell} \langle \phi_{0,0}, x^q \cdot \phi_{0,i} \rangle. \quad (22)$$

For the 0th order polynomial, as well as for the 0th derivative  $\langle \phi_{0,0}, \phi_{0,\ell} \rangle = \delta_{0\ell}$  is valid. Here  $\delta_{ab}$  is the usual Kronecker delta distribution: it is 1 if  $a = b$  and 0 in all other cases.

In the usual advection-diffusion equation no polynomials of the spatial coordinates are present, so the discretization of Eq. (6)

$$\left( T - \mu \langle \phi_{0,k}, \frac{\partial^2}{\partial x^2} \phi_{0,\ell} \rangle + u \langle \phi_{0,k}, \frac{\partial}{\partial x} \phi_{0,\ell} \rangle - \sigma \langle \phi_{0,k}, \phi_{0,\ell} \rangle \right) c_\ell^{[0]} = f_k^{[0]}, \quad (23)$$

where  $T$  summarizes the finite difference terms arising from the time discretization, depending on the method of deriving the finite differences, the number of previous time steps, the length of the time steps and the initial conditions.

### 3.3. Matrix elements for higher resolutions and wavelets

In case of (16) the matrix elements can be calculated from those of  $D_{k,\ell}^{[0]}$  using (1), (17) and (4) – and its derivative counterparts – respectively [23,26]. As a first step, if one of the resolution indices are non-zero in (18) or (21) the refinement equation results in

$$\langle \phi_{00}, \frac{\partial}{\partial x} \phi_{m\ell} \rangle = 2 \sum_{k=0}^{N_s} h_k^* \langle \phi_{00}, \frac{\partial}{\partial x} \phi_{m-1 \ell-2^{m-1}k} \rangle. \quad (24)$$

If the other index is also larger than 0, then the smaller of the two resolution levels can be compensated by refinement equations, and the problem is led back to (24) as

$$\langle \phi_{nk}, \frac{\partial}{\partial x} \phi_{m\ell} \rangle = \begin{cases} 2^n \langle \phi_{00}, \frac{\partial}{\partial x} \phi_{m-n \ell-2^{m-n}k} \rangle, & \text{if } m > n, \\ 2^m \langle \phi_{00}, \frac{\partial}{\partial x} \phi_{n-m k-2^{n-m}\ell} \rangle, & \text{if } n > m. \end{cases} \quad (25)$$

Using the scaling function expansion of the wavelet's derivative – which is very similar to (4), just a factor of 2 is introduced –, straightforwardly results in the wavelet matrix elements

$$\langle \phi_{nk}, \frac{\partial}{\partial x} \psi_{m\ell} \rangle = \sum_{i=0}^{N_s} h_{-i+1}^* \langle \phi_{nk}, \frac{\partial}{\partial x} \phi_{m+1 i+2\ell} \rangle, \quad (26)$$

$$\langle \psi_{nk}, \frac{\partial}{\partial x} \phi_{m\ell} \rangle = \sum_{j=0}^{N_s} h_{-j+1} \langle \phi_{n+1 j+2k}, \frac{\partial}{\partial x} \phi_{m\ell} \rangle, \quad (27)$$

$$\langle \psi_{nk}, \frac{\partial}{\partial x} \psi_{m\ell} \rangle = \sum_{i=0}^{N_s} \sum_{j=0}^{N_s} h_{-j+1} h_{-i+1}^* \langle \phi_{n+1 j+2k}, \frac{\partial}{\partial x} \phi_{m+1 i+2\ell} \rangle. \quad (28)$$

### 3.4. Matrix elements for discontinuous functions

In wavelet-based calculations the borders and the discontinuities cause always a problem, as the wavelets have finite support and can usually approximate continuous functions. Having a step-function in the system, like in case of the urban traffic the buildings, streets, and other obstacles necessitates numerical integration, which takes long time and gives not very high precision. For example, Table 1 contains the matrix element of a simple rectangular function

$$R_k(x) = \begin{cases} 1 & \text{if } x \in [k, k + 1), \\ 0 & \text{if } x \in (-\infty, k) \cup [k + 1, \infty). \end{cases} \quad (29)$$

for  $2^{10}$ ,  $2^{15}$ ,  $2^{20}$  and  $2^{23}$  points in the interval. The last row contains also the duration of the calculations, as an average of more runs on a desktop computer with dual core 2.2 GHz processor and 4 GB RAM, in Matlab.

Only the matrix elements  ${}^k Y_\ell = \langle \phi_{0,0}, R_k \phi_{0,\ell} \rangle$  are calculated, with  $k \in \{0, 1, \dots, N_s - 1\}$  as these rectangular functions overlap with the support of the scaling function  $\phi_{0,0}$ . Also those matrix elements are zero, where the two basis functions do not overlap – i.e., where  $|\ell| > N_s - 1$  is valid –, and where the rectangular function  $R_k$  does not overlap with  $\phi_{0,\ell}$  – i.e., where  $\ell < k - N_s + 2$  or  $\ell > k$ . All the other matrix elements can be derived by simple shifts, as  $\langle \phi_{0,i}, R_{k+i} \phi_{0,\ell+i} \rangle = \langle \phi_{0,0}, R_k \phi_{0,\ell} \rangle$ .

As the sum of the above matrix elements give the  $\langle \phi_{0,0}, \phi_{0,\ell} \rangle$  integrals, the following sum rules have to be fulfilled

$$\begin{aligned} \langle \phi_{0,0}, \phi_{0,-N_s+1} \rangle &= {}^0 Y_{-N_s+1}, \\ \langle \phi_{0,0}, \phi_{0,-N_s+2} \rangle &= {}^0 Y_{-N_s+2} + {}^1 Y_{-N_s+2}, \\ \langle \phi_{0,0}, \phi_{0,-N_s+3} \rangle &= {}^0 Y_{-N_s+3} + {}^1 Y_{-N_s+3} + {}^2 Y_{-N_s+3}, \\ &\vdots \\ \langle \phi_{0,0}, \phi_{0,0} \rangle &= {}^0 Y_0 + {}^1 Y_0 + {}^2 Y_0 + \dots + {}^{N_s-2} Y_0 + {}^{N_s-1} Y_0, \\ \langle \phi_{0,0}, \phi_{0,1} \rangle &= {}^1 Y_1 + {}^2 Y_1 + \dots + {}^{N_s-2} Y_1 + {}^{N_s-1} Y_1, \\ &\vdots \\ \langle \phi_{0,0}, \phi_{0,N_s-2} \rangle &= {}^{N_s-2} Y_{N_s-2} + {}^{N_s-1} Y_{N_s-2}, \\ \langle \phi_{0,0}, \phi_{0,N_s-1} \rangle &= {}^{N_s-1} Y_{N_s-1}, \end{aligned} \quad (30)$$

Note, that all of the values on the left hand side are zero, except for  $\langle \phi_{0,0}, \phi_{0,0} \rangle$ , which is 1.

It can be seen from Table 1, that the precision increases slowly and the duration of the calculation increases very rapidly. Alternatively to the numerical integration I suggest

Table 1. Matrix elements  ${}^k Y_\ell$  with their calculation time  $t$  of the rectangular functions  $R_k(x)$  as a result of numerical integration with various precisions  $\varepsilon$ . Daubechies-6 scaling functions are used with  $N_s = 5$ .

$\varepsilon$	$2^{10}$	$2^{15}$	$2^{20}$	$2^{23}$
${}^0 Y_{-4}$	-0.00000000015124	-0.00000000000001	0	0
${}^0 Y_{-3}$	-0.000443983002980	-0.000441411389722	-0.000441330880502	-0.000441328607930
${}^0 Y_{-2}$	-0.018852903241268	-0.018795005543038	-0.018793194141192	-0.018793143011087
${}^0 Y_{-1}$	0.123570297966252	0.123335808619037	0.123328472377250	0.123328265298497
${}^0 Y_0$	0.495840194948971	0.496622550356504	0.496647009103327	0.496647699480092
${}^1 Y_{-3}$	0.000443982821493	0.000441411389712	0.000441330880502	0.000441328607930
${}^1 Y_{-2}$	0.018988089017918	0.018929419189328	0.018927583639190	0.018927531827427
${}^1 Y_{-1}$	-0.117900092490012	-0.117682953517957	-0.117676160594197	-0.117675968851946
${}^1 Y_0$	0.471683782870471	0.470971706656950	0.470949448367932	0.470948820104503
${}^1 Y_1$	0.123570297966252	0.123335808619037	0.123328472377250	0.123328265298497
${}^2 Y_{-2}$	-0.000135184385099	-0.000134413646213	-0.000134389497999	-0.000134388816340
${}^2 Y_{-1}$	-0.005702899144805	-0.005685354773212	-0.005684805492510	-0.005684789987707
${}^2 Y_0$	0.031131664896956	0.031065665678285	0.031063599371681	0.031063541045095
${}^2 Y_1$	-0.117900092490012	-0.117682953517957	-0.117676160594197	-0.117675968851946
${}^2 Y_2$	-0.018852903241268	-0.018795005543038	-0.018793194141192	-0.018793143011087
${}^3 Y_{-1}$	0.000032689977738	0.000032499671929	0.000032493709459	0.000032493541149
${}^3 Y_0$	0.001342875637914	0.001338599123643	0.001338465237171	0.001338461457897
${}^3 Y_1$	-0.005702899144805	-0.005685354773212	-0.005684805492510	-0.005684789987707
${}^3 Y_2$	0.018988089017918	0.018929419189328	0.018927583639190	0.018927531827427
${}^3 Y_3$	-0.000443983002980	-0.000441411389722	-0.000441330880502	-0.000441328607930
${}^4 Y_0$	0.000001486637458	0.000001478184885	0.000001477919879	0.000001477912398
${}^4 Y_1$	0.000032689977738	0.000032499671929	0.000032493709459	0.000032493541149
${}^4 Y_2$	-0.000135184385099	-0.000134413646213	-0.000134389497999	-0.000134388816340
${}^4 Y_3$	0.000443982821493	0.000441411389712	0.000441330880502	0.000441328607930
${}^4 Y_4$	-0.00000000015124	-0.00000000000001	0	0
$t$ (s)	0.1038	2.278	73.59	2717

another calculation method similar to the ones presented in the previous two subsections. If the matrix element is calculated according to its definition

$${}^k Y_\ell = \langle \phi_{0,0}, R_k(x)\phi_{0,\ell} \rangle = \int_k^{k^1} \phi^*(x)\phi(x-\ell)dx \quad (31)$$

and the refinement equation (1) is substituted into both  $\phi_{0,0}^*$  and  $\phi_{0,\ell}$  after changing of the integration variable one arrives at

$${}^k Y_\ell = 2 \sum_{i_1=0}^{N_s} \sum_{i_2=1}^{N_s} h_{i_1}^* h_{i_2} ({}^{2\ell+i_2-i_1} Y_{2k-i_1} + {}^{2\ell+i_2-i_1} Y_{2k-i_1+1}). \quad (32)$$

The above equation also leads on an eigenvalue equation of the matrix

$$M_{k,\ell}^{[0]} = \begin{pmatrix} B_1 & B_0 & 0 & 0 & \dots & 0 \\ B_3 & B_2 & B_1 & B_0 & \dots & 0 \\ B_5 & B_4 & B_3 & B_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & B_{N_s} & B_{N_s-1} \end{pmatrix}, \quad (33)$$

with the blocks  $B_z =$

$$\begin{pmatrix} h_{N_s} h_z^* + & h_{N_s-1} h_{z-1}^* & 0 & 0 & 0 & \dots & 0 \\ h_{N_s} h_{z-1}^* & & & & & & \\ h_{N_s-2} h_z^* + & h_{N_s-1} h_z^* + & h_{N_s} h_z^* + & h_{N_s} h_{z-1}^* & 0 & \dots & 0 \\ h_{N_s-3} h_{z-1}^* & h_{N_s-2} h_{z-1}^* & h_{N_s-1} h_{z-1}^* & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & h_0 h_z^* & h_1 h_z^* + & h_2 h_z^* + & h_2 h_z^* + \\ & & & & h_0 h_{z-1}^* & h_1 h_{z-1}^* & h_1 h_{z-1}^* \\ 0 & \dots & 0 & 0 & 0 & h_0 h_z^* & h_1 h_z^* + \\ & & & & & & h_0 h_{z-1}^* \end{pmatrix}. \quad (34)$$

Matrix  $M_{k,\ell}^{[0]}$  has an eigenvector corresponding to eigenvalue 1, and this eigenvector can be normalized due to (30), as  $\langle \phi_{0,0}, \phi_{0,i} \rangle = \delta_{i0}$ . Numerical checks were carried out for various scaling functions. In case of Daubeches-6 basis set ( $N_s = 5$ ), the normalization condition is

$$1 = \langle \phi_{0,0}, \phi_{0,0} \rangle = {}^0 Y_0 + {}^1 Y_0 + {}^2 Y_0 + {}^3 Y_0 + {}^4 Y_0, \quad (35)$$

and the resulting matrix elements are listed in Table 2 The differences between the quantities  ${}^k Y_\ell$  numerically integrated and calculated with the eigenvalue method (32) are plotted

Table 2. Matrix elements  ${}^k Y_\ell$  with their calculation time  $t$  of the rectangular functions  $R_k(x)$  as a result of the eigenvalue-based calculation. Daubechies-6 scaling functions are used with  $N_s = 5$ .

${}^0 Y_{-4}$	-0.0000000000000000
${}^0 Y_{-3}$	-0.000441328283277
${}^0 Y_{-2}$	-0.018793135706778
${}^0 Y_{-1}$	0.123328235715790
${}^0 Y_0$	0.496647798105380
${}^1 Y_{-3}$	0.000441328283277
${}^1 Y_{-2}$	0.018927524425738
${}^1 Y_{-1}$	-0.117675941460163
${}^1 Y_0$	0.470948730352582
${}^1 Y_1$	0.123328235715790
${}^2 Y_{-2}$	-0.000134388718960
${}^2 Y_{-1}$	-0.005684787772732
${}^2 Y_0$	0.031063532712708
${}^2 Y_1$	-0.117675941460163
${}^2 Y_2$	-0.018793135706778
${}^3 Y_{-1}$	0.000032493517105
${}^3 Y_0$	0.001338460918000
${}^3 Y_1$	-0.005684787772732
${}^3 Y_2$	0.018927524425738
${}^3 Y_3$	-0.000441328283277
${}^4 Y_0$	0.000001477911329
${}^4 Y_1$	0.000032493517105
${}^4 Y_2$	-0.000134388718960
${}^4 Y_3$	0.000441328283277
${}^4 Y_4$	-0.000000000000000
$t$ (s)	0.002132

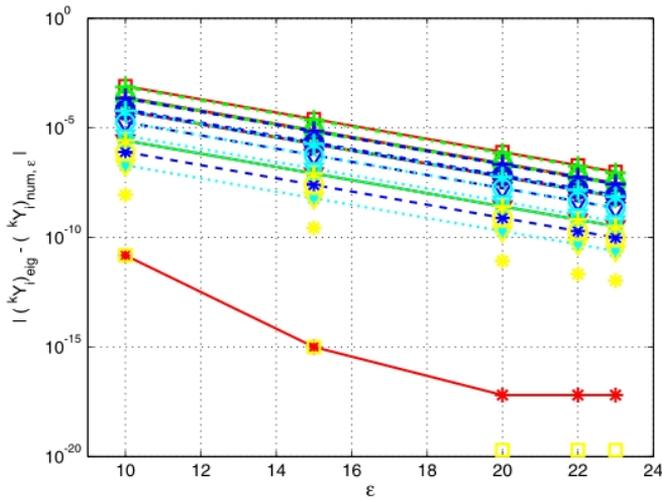


Figure 1. Difference of the numerically integrated values  ${}^k Y_\ell$  from their counterparts calculated with an eigenvalue equation of matrix  $M_{k,\ell}^{[0]}$  as a function of the grid points in a unit interval  $2^\varepsilon$ . Red color with continuous line means  $k = 0$ , green with dash-dot line mean  $k = 1$ , blue with dashed line  $k = 2$ , cyan with dotted line  $k = 3$  and the color yellow and markers without line means  $k = 4$ . The markers  $*$ ,  $\nabla$ ,  $\circ$ ,  $+$ , and  $\square$  stand for the indices  $\ell = k - 4$ ,  $k - 3$ ,  $k - 2$ ,  $k - 2$  and  $k$  respectively.

in Figure 1. Also the calculation time is shown on Figure 2, where an approximate power law behavior can be seen as expected for the numerical integrations.

In case of derivatives, similarly to the continuous case described in (19), each differentiation introduces only a factor of 2 to the matrix  $M_{k,\ell}^{[0]}$ , thus the eigenvector corresponding to the eigenvalues 0.5, 0.25, etc. should be found for the matrix elements with the first, second, etc. derivatives respectively.

The transformation to higher resolution levels goes similarly to (24–28), except, that with each use of the refinement equation, the number of the elements  $\langle \phi_{m,i}, R_k(x) \phi_{n,\ell} \rangle$  doubles, so for higher resolution levels an exponentially increasing number of lower resolution level integrals should be taken into account.

#### 4. Summary

In wavelet-based modeling the concentration changes of nitrogen oxides and other air pollutants in urban environment discontinuities are arising around the obstacles that the

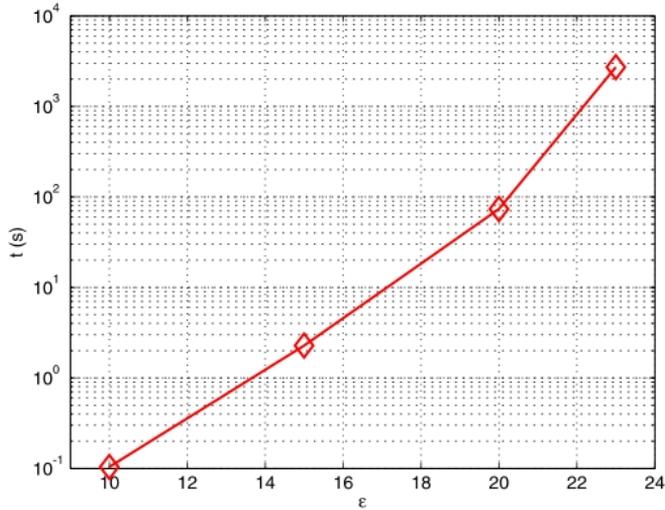


Figure 2. Computation time of the numerical integrations as a function of of the grid points in a unit interval  $2^\varepsilon$ .

gases can not penetrate. The discretization of these discontinuities can be carried out not only by numerical integration, but also by a much quicker and more economic calculation – derived in this paper – based on the eigenvalue equation of a matrix generated from the coefficients of the refinement equation, the basic equation of the wavelet analysis. The size of the matrix is  $N_s^2$ , with  $N_s$  being the length of the support of the wavelets.

The calculated matrix elements are system independent, they depend only on the type of the wavelets used, thus in later calculations they can be loaded from a database, but for calculating this database as precisely as possible, the method developed in this article, based on the results of [23, 26], is necessary.

## Acknowledgement

The publishing of this paper was supported by Hungary’s New Széchenyi Plan, co-financed by the European Social Fund, grant ID. TÁMOP-4.2.2.C-11/1/KONV-2012-0012 “Smarter Transport” Infocommunications support of cooperative transport systems.

## References

- [1] Daubechies I: Ten Lectures on Wavelets. CBMS-NSF regional conference series in applied mathematics 61, SIAM, Philadelphia, 1992.

- [2] Chui CK: *An Introduction to Wavelets*. Academic Press, San Diego, 1992.
- [3] Crooks J, Isakov V: A wavelet-based approach to blending observations with deterministic computer models to resolve the intraurban air pollution field. *J. Air & Waste Management Assoc.*, Vol. 63, p. 1369-1385, 2013. DOI:10.1080/10962247.2012.758061
- [4] Lobdell DT, Isakov V, Baxter L, Touma JS, Smuts MB, Özkaynak H: Feasibility of assessing public health impacts of air pollution reduction programs on a local scale: New Haven case study. *Environ. Health Perspect.*, Vol. 119, pp. 487-493, 2011. DOI:10.1289/ehp.1002636
- [5] Metzger K, Klein M, Flanders W, Peel J, Mulholland J, Langberg J, Tolbert P: Ambient air pollution and cardiac arrhythmias in patients with implantable defibrillators. *Epidemiology* Vol. 18, pp. 585-592, 2007. DOI:10.1097/EDE.0b013e318124ff0e
- [6] Appel K, Bhave P, Gilliland A, Sarwar G, Roselle S: Evaluation of the community multiscale air quality (CMAQ) model version 4.5: Sensitivities impacting model performance; part II-particulate matter. *Atmos. Environ.* Vol. 42, pp. 6057-6066, 2008. DOI:10.1016/j.atmosenv.2008.03.036
- [7] Cimorelli A, Perry S, Venkatram A, Weil J, Paine R, Wilson R, Lee R, Peters W, Brode WR: AERMOD: A dispersion model for industrial source applications. Part I: General model formulation and boundary layer characterization. *J. Appl. Meteorol.* Vol. 44, pp. 682-693, 2004. DOI:10.1175/JAM2227.1
- [8] Gelfand A, Sahu S: *Combining Monitoring Data and Computer Model Output in Assessing Environmental Exposure*. Oxford University Press, Oxford 2009.
- [9] Zidek J, Le N, Liu Z: Combining data and simulated data for spacetime fields: Application to ozone. *Environ. Ecol. Stat.*, Vol. 19, pp. 37-56, 2012. DOI:10.1007/s10651-011-0172-1
- [10] Christopoulos Ch, Skodras A, Ebrahimi T: The JPEG2000 Still Image Coding System: An Overview. *IEEE Trans. Consumer Electronics*, Vol. 46, pp. 1103-1127, 2000. DOI:10.1109/30.920468
- [11] Kiely A, Klimesh M: *The ICER Progressive Wavelet Image Compressor*. IPN Progress Report 42-155, November 15, 2003.
- [12] Urban K: *Wavelet Methods for Elliptic Partial Differential Equations*. Oxford University Press, Oxford, 2009.
- [13] Dahmen W: *Wavelets and Multiscale Methods for Operator Equations*. *Acta Numerica* Vol. 6, p. 55, 1997. DOI:10.1017/s0962492900002713

- [14] Nasif H, Omori R, Suzuki A, Naguib M, Nagy M: Wavelet-Based Algorithms for Solving Neutron Diffusion Equations. *J. Nuclear Sci. Techn.*, Vol. 38, pp. 161-173, 2001. DOI:10.1080/18811248.2001.9715018
- [15] Galerkin BG: On electrical circuits for the approximate solution of the Laplace equation (in Russian). *Vestnik Inzh.* Vol. 1, p. 897, 1915.
- [16] Nagy Sz, Pipek J: Multiresolution Analysis of Density Operators, Electron Density, and Energy Functionals. *Int. J. Quantum Chem.*, Vol. 84, p. 523, 2001. DOI:10.1002/qua.1406
- [17] Pipek J, Nagy Sz: Refinement trajectory and determination of eigenstates by a wavelet based adaptive method. *J. Chem. Phys.* Vol. 125, 174107, 2006. DOI:10.1063/1.2363368
- [18] Goedecker S, Ivanov O: Linear scaling solution of the Coulomb problem using wavelets. *Solid State Commun.*, Vol. 105, p. 665-669, 1998.
- [19] Yang Sh, Ni G, Ho SL, Machado JM, Rahman MA, Wong HC: Wavelet-Galerkin Method for Computations of Electromagnetic Fields – Computation of Connection Coefficients. *IEEE Trans. Magn.*, Vol. 36, p. 644, 2000. DOI:10.1109/20.877532
- [20] Rickard Y: An efficient wavelet-based solution of electromagnetic field problems. *Appl. Numer. Math.*, Vol. 58, p. 472, 2008. DOI:10.1016/j.apnum.2007.01.020
- [21] Nagy Sz, Pipek J: A wavelet-based adaptive method for determining eigenstates of electronic systems. *Theor. Chim. Acta*, Vol. 125, p. 471-479, 2010. DOI:10.1007/s00214-009-0653-6
- [22] Dahmen W: Wavelet Methods for PDEs – Some Recent Developments. *J. Comput. Appl. Math.*, Vol. 128, p. 123, 2001. DOI:10.1016/S0377-0427(00)00511-2
- [23] Pipek J, Nagy Sz: An Economic Prediction of Refinement Coefficients in Wavelet-Based Adaptive Methods for Electron Structure Calculations. *J. Comput. Chem.*, Vol. 34, p. 460-465, 2013. DOI:10.1002/jcc.23154
- [24] Hillel R, Joseph A: *Environmental fluid mechanics*. Marcel Dekker Inc., New York Basel, 2001. DOI:10.1201/9780203908495
- [25] Dang Q, Matthias E: Adequate Numerical Solution of Air Pollution Problems by positive Difference Schemes on unbounded Domains. *Math. Computer Modelling*, Vol. 44, pp. 834-856, 2006. DOI:10.1016/j.mcm.2006.02.016
- [26] Dahmen W, Micchelli CA: Using the refinement equation for evaluating integrals of wavelets. *SIAM J. Numer. Anal.*, Vol. 30, p. 507, 1993. DOI:10.1137/0730024

# Queuing Models and Subspace Identification in Logistics

T. Vadvári, P. Várlaki

Széchenyi István University, System Theory Laboratory  
Egyetem tér 1., 9026 Győr, Hungary  
E-mail: varlaki@sze.hu

**Abstract:** In modern logistics it might be helpful to describe the behavior of a complex logistical process as well as to determine the strength of relations between certain parameters of the system. In this paper a subspace identification approach has been applied to estimate the relation between the features of the system based on measured input-output pairs. In order to validate the suitability of the approach for logistical processes a queuing based model has been proposed and used to generate simulation data. Our analysis as well as the obtained results clearly reflect that subspace identification approaches can advantageously be applied to model the relation between certain parameters of the system, nevertheless to characterize the strength of this relation, as well.

**Keywords:** *subspace identification, queuing models, supply chains, modeling*

## 1. Introduction

As in many fields also in logistics system modeling and identification approaches play significant role especially when accurate models of complex logistical processes (LP) are needed. Such models may be helpful to predict various features related to the modeled system, such as the response time or in case of supply chains the delivery cycle time, customer order path (related to time spent in different channels), etc. A framework to promote the better understanding of supply chain performance measurement and metrics can be followed for example in [8]. Trough monitoring of performance metrics analytic or statistical models of the observed LP can be designed.

Depending on the knowledge about the modeled system a broad range of solutions can be utilized. Since complex logistical systems are non-linear MIMO systems and are influenced by many parameters their modeling is not a trivial task. Many methods have been proposed to deal with multi-input, multi-output systems in the literature. Perhaps the most popular tool in this topic is the linear parameter varying (LPV) structure by which

non-linear systems can be modeled and controlled on the basis of linear control theories. If there is no knowledge about the inner structure of the logistic system such as for instance the concrete service strategy and other internal mechanisms only black box like solutions (mainly heuristic approaches) are utilized. In this case the system might be identified based on measured input-output data. In the literature many models (as for instance scheduling, transportation planning, flow-shop sequencing problem) related to logistic systems are based on the fuzzy set and fuzzy control theory [2][3], statistics or their combination [9][10]. Furthermore, the most recent results of the numerical algebra, such as the higher order singular value decomposition and the related tensor product transformation (making connection between LPV models and higher order tensors) offer promising tools to bridge heuristic and analytic approaches. In such a joint framework besides analytic description of the system the expert knowledge can be considered, as well. This may further improve the effectiveness and extend the applicability of the related methods [7][6].

Many times it is difficult to find a proper mathematical model in form of differential equations which would suitable approximate the behavior of the observed logistical process even if the identification of the system is considered locally. However subspace identification techniques combined with tensor product transformation seem to be promising to model complex logistical processes based on input-output data. In this case there is no need for an explicit model parametrization, which is a rather complicated matter for multi-output linear systems [1]. During this research our motivation was first of all to investigate and to show how efficiently subspace identification techniques can be used in case of systems where long delays may occur. We have been focusing first of all on logistical processes which are good examples of such systems. In order to accurately identify the system on subspace basis the input must be persistently exciting, i.e. it must contain sufficiently many distinct frequencies. The Gaussian white noise, pseudo-random binary noise, etc. are the most suitable choices for input. However in case of a logistical process the arrival of demands is considered to be a Poisson process. However arriving demands usually enter the queue of waiting demands which acts like a 'damper' thus to take directly the arrival of demands as input (depending on the size of the mentioned queue and delays in the system) is many times not suitable for identification. Therefore instead of taking the direct input, during the experiments a transformed input has been considered. Such a transformed input is efficient even in case when the arrival of demands is modeled by a Poisson or other type of processes. The main contribution of the paper is to show how efficiently a transformed input can be used to identify its relation to certain parameters of the logistical process. In addition it will be shown how the strength of these relations can be characterized with the help of the identified model.

The paper is organized as follows: Section 2 gives a brief overview of subspace identification for deterministic case, Section 3 deals with supply chain models in relation to subspace identification. In Section 4 and 5 examples are reported together with model validation, Section 6 points out the possibilities of embedding logistical processes into linear parameter varying (LPV) framework, finally conclusions and future works are reported.

## 2. Overview on Subspace Identification of LTI Systems

Before turning the focus onto logistical processes, let us give a brief description on how subspace identification techniques can be used to identify linear time invariant (LTI) vertex models in the parameter space. Let us assume that the local behavior of the logistical system is deterministic, thus it can be described in the well known state space form as follows:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k \quad (1)$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k, \quad (2)$$

where  $\mathbf{x}_k \in \mathbb{R}^n$  stands for the state vector,  $\mathbf{u}_k$  and  $\mathbf{y}_k$  represent the input and output vector respectively at time  $k$ . The goal is to find the model matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  and  $\mathbf{D}$  based on input-output pairs. As described in [1] let us first arrange the input-output pairs into so called Hankel matrices (reflecting the history of our input-output data):

$$\mathbf{U}_{1|i} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_j \\ \mathbf{u}_2 & \mathbf{u}_3 & \dots & \mathbf{u}_{j-1} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{u}_i & \mathbf{u}_{i+1} & \dots & \mathbf{u}_{j+i-1} \end{bmatrix}, \quad (3)$$

$$\mathbf{Y}_{1|i} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \dots & \mathbf{y}_j \\ \mathbf{y}_2 & \mathbf{y}_3 & \dots & \mathbf{y}_{j-1} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{y}_i & \mathbf{y}_{i+1} & \dots & \mathbf{y}_{j+i-1} \end{bmatrix}, \quad (4)$$

and let the history of states (unknown) to be estimated encode as follows:

$$\mathbf{X}_i = [\mathbf{x}_i \quad \mathbf{x}_{i+1} \quad \dots \quad \mathbf{x}_{i+j-1}]. \quad (5)$$

It can be recognized from (2) that all row vectors in  $\mathbf{Y}_{1|i}$  are in the vector space determined by the union of row space of  $\mathbf{X}_i$  and  $\mathbf{U}_{1|i}$ . Let us assume that the intersection of row space of  $\mathbf{X}_i$  and  $\mathbf{U}_{1|i}$  is empty. The most simple alternative for estimating  $\mathbf{X}_i$  (up to a constant multiple  $\mathbf{C}$ ) is to project the row space of  $\mathbf{Y}_i$  onto orthogonal complement of the row space of  $\mathbf{U}_{1|i}$ . The elements of  $\mathbf{Y}_i$  can be expressed with the help of the extended observability matrix  $\mathbf{\Gamma}_i$  and lower block triangular Toeplitz matrix  $\mathbf{H}_i$  form as follows [1]:

$$\mathbf{Y}_{1|i} = \mathbf{\Gamma}_i \mathbf{X}_1 + \mathbf{H}_i \mathbf{U}_{1|i}, \quad (6)$$

where

$$\mathbf{\Gamma}_i = [\mathbf{C} \quad \mathbf{C}\mathbf{A} \quad \dots \quad \mathbf{C}\mathbf{A}^{i-1}]^\top \quad (7)$$

and

$$\mathbf{H}_i = \begin{bmatrix} \mathbf{D} & 0 & 0 & \dots & 0 \\ \mathbf{CB} & \mathbf{D} & 0 & \dots & 0 \\ \mathbf{CAB} & \mathbf{CB} & \mathbf{D} & \dots & 0 \\ \mathbf{CA}^{i-2}\mathbf{B} & \mathbf{CA}^{i-3}\mathbf{B} & \dots & \mathbf{CB} & \mathbf{D} \end{bmatrix}. \quad (8)$$

By substituting recursively into (1) we can express the state sequence  $\mathbf{X}_{i+1}$  as follows:

$$\mathbf{X}_{i+1} = \mathbf{A}^i \mathbf{X}_1 + \Delta_i \mathbf{U}_{1|i}, \quad (9)$$

where

$$\Delta_i = [\mathbf{A}^{i-1}\mathbf{B} \quad \mathbf{A}^{i-2}\mathbf{B} \quad \dots \quad \mathbf{AB} \quad \mathbf{B}] \quad (10)$$

stands for the reversed extended controllability matrix [1]. From (6) the state sequence  $\mathbf{X}_1$  can be expressed as:

$$\mathbf{X}_1 = \Gamma^*_i \mathbf{Y}_{1|i} - \Gamma^*_i \mathbf{H}_i \mathbf{U}_{1|i}, \quad (11)$$

By substituting (11) into (9) we obtain:

$$\mathbf{X}_{i+1} = \mathbf{A}^i \Gamma^*_i \mathbf{Y}_{1|i} - \mathbf{A}^i \Gamma^*_i \mathbf{H}_i \mathbf{U}_{1|i} + \Delta_i \mathbf{U}_{1|i}. \quad (12)$$

Let us express  $\mathbf{X}_{i+1}$  as the sum of two matrices, where one of the matrices contains only the input-output values, i.e.

$$\mathbf{X}_{i+1} = \mathbf{L}_i \mathbf{W}_{1|i}, \quad (13)$$

where

$$\mathbf{L}_i = [\Delta_i - \mathbf{A}^i \Gamma^*_i \mathbf{H}_i \quad \mathbf{A}^i \Gamma^*_i] \quad (14)$$

and

$$\mathbf{W}_{1|i} = [\mathbf{U}_{1|i} \quad \mathbf{Y}_{1|i}]^\top. \quad (15)$$

Since based on (6)

$$\mathbf{Y}_{i+1|2i} = \Gamma_i \mathbf{X}_{i+1} + \mathbf{H}_i \mathbf{U}_{i+1|2i} = \Gamma_i \mathbf{L}_i \mathbf{W}_{1|i} + \mathbf{H}_i \mathbf{U}_{i+1|2i}. \quad (16)$$

Let us now project  $\mathbf{Y}_{i+1|2i}$  onto orthogonal complement of  $\mathbf{U}_{i+1|2i}$ . Since the projection of  $\mathbf{H}_i \mathbf{U}_{i+1|2i}$  onto its orthogonal complement is empty subspace we obtain [1]:

$$\mathbf{Y}_{i+1,2i}/\mathbf{U}_{i+1,2i}^\perp = \Gamma_i \mathbf{L}_i \mathbf{W}_{1|i}/\mathbf{U}_{i+1,2i}^\perp \quad (17)$$

$$(\mathbf{Y}_{i+1,2i}/\mathbf{U}_{i+1,2i}^\perp)(\mathbf{W}_{1|i}/\mathbf{U}_{i+1,2i}^\perp)^{-1} = \Gamma_i \mathbf{L}_i, \quad (18)$$

$$\underbrace{(\mathbf{Y}_{i+1,2i}/\mathbf{U}_{i+1,2i}^\perp)(\mathbf{W}_{1|i}/\mathbf{U}_{i+1,2i}^\perp)^{-1} \mathbf{W}_{1|i}}_{\mathbf{O}_{i+1}} = \Gamma_i \underbrace{\mathbf{L}_i \mathbf{W}_{1|i}}_{\mathbf{X}_{i+1}}, \quad (19)$$

$$\mathbf{O}_{i+1} = \Gamma_i \mathbf{X}_{i+1} \quad (20)$$

Let us investigate the structure of  $\mathbf{O}_{i+1}$ . Based on (7) and (5) it can be expressed as:

$$\mathbf{O}_{i+1} = [\mathbf{C} \quad \mathbf{CA} \quad \dots \quad \mathbf{CA}^{i-1}]^\top [\mathbf{x}_{i+1} \quad \mathbf{x}_{i+2} \quad \dots \quad \mathbf{x}_{i+j}]. \quad (21)$$

Based on (21) the rank of  $\mathbf{O}_{i+1}$  equals to the rank of the state sequence matrix  $\mathbf{X}_{i+1}$ . Equivalently, the dimensionality of the state vector  $\mathbf{x}$  equals to the dimensionality of  $\mathbf{O}_{i+1}$ . The rank of  $\mathbf{O}_{i+1}$  can be determined by singular value decomposition (SVD) as follows[1]:

$$\mathbf{O}_{i+1} = \mathbf{U}_1 \mathbf{S}_1 \mathbf{V}_1 \quad (22)$$

$$\Gamma_i \mathbf{X}_{i+1} = \mathbf{U}_1 \mathbf{S}_1^{1/2} \mathbf{T} \mathbf{T}^{-1} \mathbf{S}_1^{1/2} \mathbf{V}_1, \quad (23)$$

where  $\mathbf{T}$  is an arbitrary invertible square matrix representing a similarity transformation.

$$\mathbf{X}_{i+1} = \mathbf{T}^{-1} \mathbf{S}_1^{1/2} \mathbf{V}_1 \quad (24)$$

$$\tilde{\mathbf{X}}_{i+1} = \mathbf{S}_1^{1/2} \mathbf{V}_1 \quad (25)$$

The system matrix can be estimated in the least squares sense from the following set of equations:

$$\begin{bmatrix} \tilde{\mathbf{X}}_{i+2} \\ \tilde{\mathbf{Y}}_{i+1} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{A}} & \tilde{\mathbf{B}} \\ \tilde{\mathbf{C}} & \tilde{\mathbf{D}} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{X}}_{i+1} \\ \tilde{\mathbf{U}}_{i+1} \end{bmatrix}, \quad (26)$$

where  $\mathbf{U}_{i+1}$  and  $\mathbf{Y}_{i+1}$  are input and output block Hankel matrices, respectively having one block row.

### 3. Modeling Supply Chains on Subspace Basis

In this section let us introduce our proposed queuing model suitable to describe and analyze supply chains or loading systems. As depicted in Fig. 1 the system is composed of resource pools, queues, servicing processes. Incoming customers or demands stand for the input of the system. The resource pools are categorized according to the type of the resource. To each resource pool a FIFO queue is connected. Resources waiting in the resource queue are assigned to demands (depending on the service the customer is requesting for). To each service a servicing time is assigned. Let us denote it  $T_S$ . After servicing a given demand the corresponding customer can leave the system while the used resources are released and directed back to the pool of resources of the given type. Servicing a request might be considered as an oriented graph where the nodes represent sub-services and the edges correspond to the ordering and delay of execution between these nodes.

By using such a concept various types of supply chains can be simulated and analyzed. For simplicity in the followings let us assume that services are composed of one node. However the same approach might be applied for services divided to numerous sub-services,

as well. During this investigation our main goal was to show how queuing models can be identified based on input-output data by using subspace identification techniques.

As already mentioned in the introduction the arrival of demand is modeled by a pseudo-random binary noise however due to the queue of waiting demands (acting like a 'damper') its impact on the system behavior is less significant (depending on the internal structure and parameters of the system). Thus several kind of processes (Poisson process, burst arrival, etc.) can be used to model the arrival of demands, due to long delays it does not significantly affect the efficiency of the identification. Therefore during these experiments a transformed input has been considered (see later in this section). In addition, identification based on input-output data might also be suitable to evaluate the strength of relation between certain parameters of the system.

The parameters of the system are for instance the queue lengths  $L_{qi}$ , number of resources  $N_{Ri}$ , servicing times  $T_{Sj}$ , where  $i$  and  $j$  stand for the number of resource pools and number of sub-services, respectively. In the following sections let us investigate a structure where  $i = 1$  and  $j = 1$ , i.e. there is only one type of resource and one service.

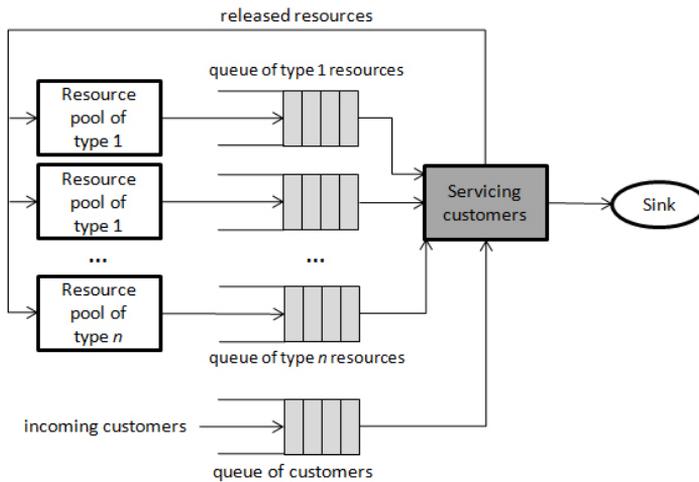


Figure 1. Illustration of the system architecture

#### 4. Example 1

In this example our goal was to estimate a deterministic state space model describing the relation between the average waiting time of resources  $m_r(t)$  and the average waiting

time of demands  $m_d(t)$ . Before going further let us show the configuration of the analyzed system:

- $N_R = 3$
- $L_q = 3$
- $L_c = 1000$
- $T_S$  is an exponential random variable with mean  $E[T_S] = 3$
- Arrival of demands  $d(t)$ : pseudo-random binary sequence (customer generation event is triggered at rising and falling edges)
- Limit of simultaneously serviced customers: 3

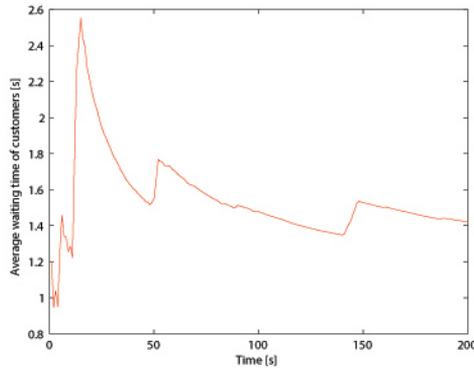


Figure 2. The transformed input  $m_d(t)$

By considering the above configuration the relation between the two mentioned features of the system has been modeled by subspace identification technique (considering the deterministic case). Fig. 2 shows the input signal reflecting the average waiting time of demands in the corresponding queue. Furthermore, in Fig. 3 the measured and modeled average waiting time of resources in the resource queue can be followed. It is clear that the input in this case stands for a transformed input namely  $m_d(t)$ . In the first parts of the experiment 200 input-output pairs have been used (generated by the above system) for model estimation. The system matrices have been determined based on the described subspace identification approach. As it can be seen the obtained model nicely follows the characteristics of the measured output. In the second part of this experiment only the first 100 input-output pairs have been considered during the model estimation and the rest 100 pairs have been used for validation (see Fig. 4). As reflected by Fig. 4 also in this case the model output nicely follows the characteristics of the measured data. On the other hand if the incoming rate of customers  $d(t)$  is considered as input (a pseudo-random number

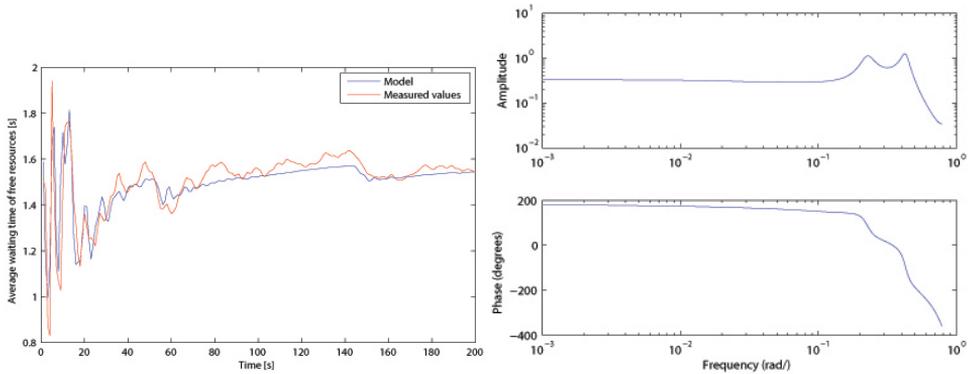


Figure 3. The measured average waiting time of resources  $m_r(t)$  (red), the output of the identified model (blue), Input:  $m_d(t)$  (left), Bode diagram of the identified model (right)

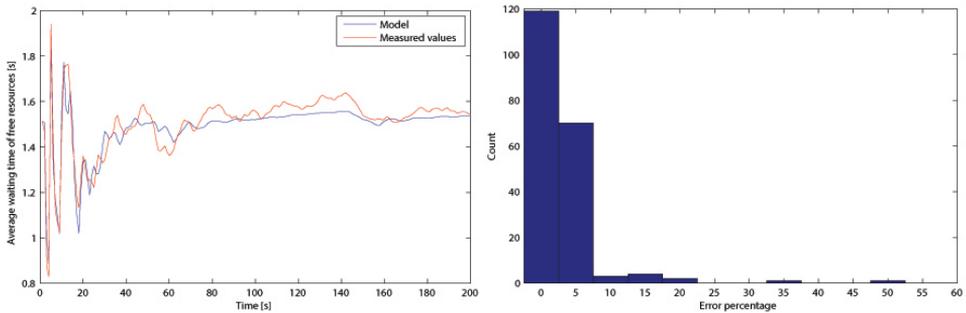


Figure 4. Validation of the identified model. The measured average waiting time of resources  $m_r(t)$  (red), the output of the identified model (blue), Input:  $m_d(t)$  (the first half of samples has been used for model estimation while the second half for validation (left), Error percentage histogram corresponding to the estimated model (right)

in our case) the obtained model is inaccurate (see Fig. 5). The primary reason for this is the length of queues and the related long delays in the system. If the maximal length of the customer queue is set to a smaller value, more accurate model is obtained. It can be assumed that the accuracy of the obtained model is strongly influenced by the strength of dependence between the two selected features.

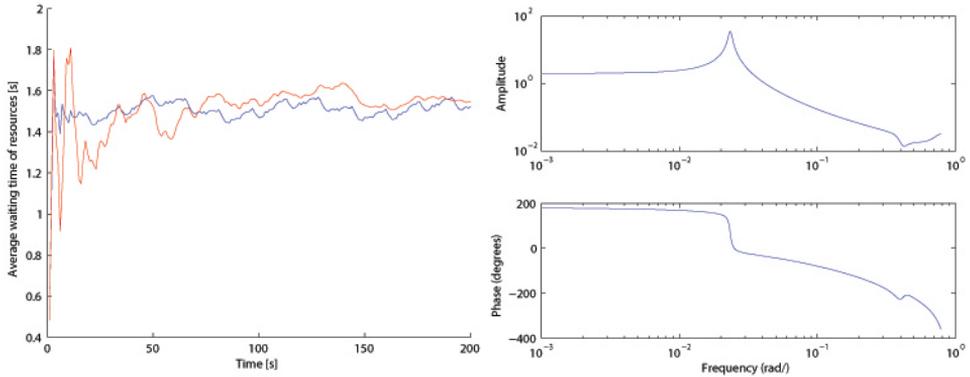


Figure 5. The measured average waiting time of resources  $m_r(t)$  (red), the output of the identified model (blue), Input:  $d(t)$ . As it can be seen, the model is of low accuracy which indicates weak relation between the used input-output variables (left), Bode diagram corresponding to the model (right)

## 5. Example 2

The main goal was to model the number of entities in the queue of resources. The input in this example is  $T_S$  while the output the number of available resources waiting in the queue. Let us consider the following system configuration:

- $N_R = 40$
- $L_q = 40$
- $L_c = 1000$
- $T_S$  is an exponential random variable with mean  $E[T_S] = 3$
- Arrival of customers: pseudo-random binary sequence (customer generation event is triggered at rising and falling edges)
- Limit of simultaneously serviced customers: 40

The input signal can be followed in Fig. 6. It represents the duration of the servicing assigned to a given demand over time. In the first part of this experiment – similarly to the previous example – 200 input-output pairs have been used for model estimation. The measured number of free resources over time together with the output of the identified model can be followed in Fig. 7. The matrices of the obtained deterministic state space

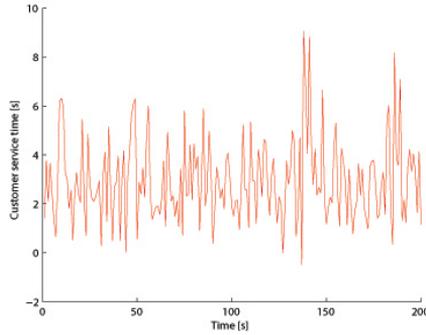


Figure 6. The input signal  $T_S$

model are as follows:

$$\mathbf{A} = \begin{bmatrix} 0.41814 & -0.080494 & -0.28302 \\ -0.27051 & -0.86992 & -0.066373 \\ 0.16624 & -0.36498 & -0.27266 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0.033479 \\ 0.014832 \\ 0.0090137 \end{bmatrix},$$

$$\mathbf{C} = [-8.94 \quad 0.51886 \quad -1.5867], \quad \mathbf{D} = [0], \quad \mathbf{X} = \begin{bmatrix} 0.77666 \\ -0.31588 \\ 1.1808 \end{bmatrix}.$$

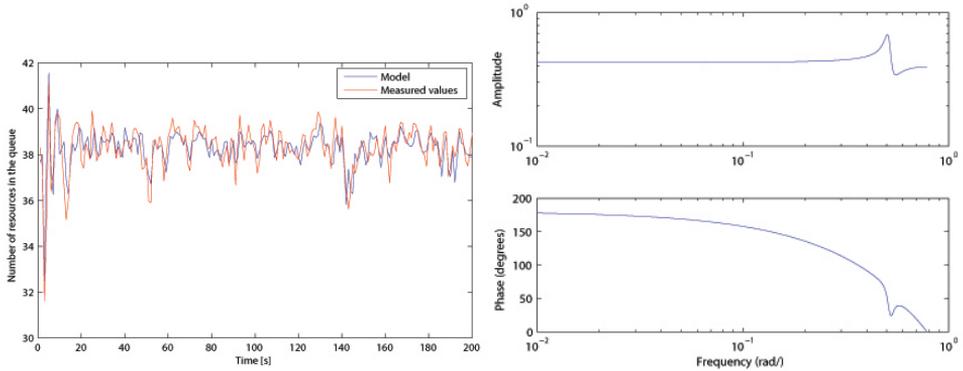


Figure 7. The measured number of free resources (red), output of the identified model (blue), Bode diagram of the identified model (right)

In the second part of the experiment – also similarly to the previous example – only the first 100 input-output pairs have been considered during the model estimation and

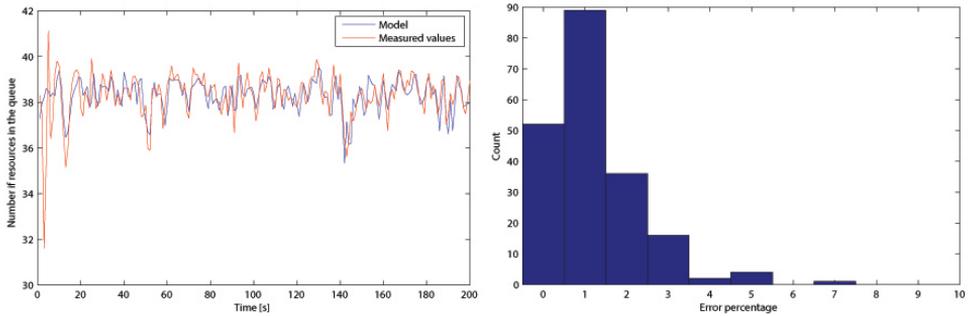


Figure 8. Validation of the identified model. The measured number of free resources (red), output of the identified model (blue) (left), Error percentage histogram corresponding to the estimated model (right)

the rest 100 pairs have been used for validation. The obtained model nicely follows the characteristics of the measured output in both cases (see Fig. 8). The error percentage histogram in Fig. 8 well reflects the accuracy of the estimated model.

## 6. Queuing Systems in LPV Framework

In this section let us give a brief description how queuing systems can be described by linear parameter varying models. It is clear that in complex queuing systems there might be many parameters which may strongly affect the behavior of the whole system. If we discretize the parameter space over a hyper-rectangular grid, for each grid point a linear time invariant (LTI) model can be assigned. The global behavior of the system can be obtained by "blending" the local models properly. Depending on the number of discretization points the number of identified local models might be significant, thus to reduce their number by keeping the accuracy at acceptable level plays another important task in this field [5]. Here the higher order singular value decomposition (HOSVD) plays significant role [4].

Let us consider the following linear parameter varying (LPV) state-space model [5]:

$$\begin{pmatrix} \dot{\mathbf{x}}(t) \\ \mathbf{y}(t) \end{pmatrix} = \mathbf{S}(\mathbf{p}(t)) \begin{pmatrix} \mathbf{x}(t) \\ \mathbf{u}(t) \end{pmatrix}, \quad (27)$$

where  $\mathbf{u}(t) \in \mathbb{R}^u$  stands for the system input,  $\mathbf{y}(t) \in \mathbb{R}^v$  represents the output of the system and  $\mathbf{x}(t) \in \mathbb{R}^k$  denotes the state vector. Furthermore  $\mathbf{p}(t) = (p_1(t), \dots, p_N(t)) \in \Omega$ ,

$$\Omega = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_N, b_N] \subset \mathbb{R}^N,$$

and

$$\mathbf{S}(\mathbf{p}(t)) = \begin{pmatrix} \mathbf{A}(\mathbf{p}(t)) & \mathbf{B}(\mathbf{p}(t)) \\ \mathbf{C}(\mathbf{p}(t)) & \mathbf{D}(\mathbf{p}(t)) \end{pmatrix} \in \mathbb{R}^{(k+v) \times (k+u)}. \quad (28)$$

$\mathbf{S}(\mathbf{p}(t))$  for arbitrary  $\mathbf{p}(t)$  parameter vector can be expressed in tensor product form as follows:

$$\begin{pmatrix} \dot{\mathbf{x}}(t) \\ \mathbf{y}(t) \end{pmatrix} = (\mathbf{S} \boxtimes_{n=1}^N \mathbf{w}_n^T(p_n)) \begin{pmatrix} \mathbf{x}(t) \\ \mathbf{u}(t) \end{pmatrix}, \quad (29)$$

where vector  $\mathbf{w}_n^T(p_n) \in \mathbb{R}^{I_n}$ ,  $n = 1..N$ , contains continuous and bounded functions  $w_{n,i_n}(p_n)$  on interval  $[a_n, b_n]$ ,  $i_n = 1..I_n$ . The  $(N + 2)$ -dimensional tensor  $\mathbf{S} \in \mathbb{R}^{I_1 \times \dots \times I_{N+2}}$  contains the system matrices of linear time invariant vertex systems [5]:

$$\mathbf{S}_{i_1 \dots i_N} = \{S_{i_1 \dots i_N, \alpha, \beta}, 1 \leq \alpha \leq I_{N+1}, 1 \leq \beta \leq I_{N+2}\}$$

$$\mathbf{S}_{i_1 \dots i_N} \in \mathbb{R}^{I_{N+1} \times I_{N+2}}$$

Applying HOSVD on the first  $N$  dimensions of  $\mathbf{S}$  we obtain the following:

$$\begin{aligned} \begin{pmatrix} \dot{\mathbf{x}}(t) \\ \mathbf{y}(t) \end{pmatrix} &= [(\mathbf{D} \boxtimes_{n=1}^N \mathbf{U}_n) \boxtimes_{n=1}^N \mathbf{w}_n^T(p_n)] \begin{pmatrix} \mathbf{x}(t) \\ \mathbf{u}(t) \end{pmatrix} \\ \begin{pmatrix} \dot{\mathbf{x}}(t) \\ \mathbf{y}(t) \end{pmatrix} &= \left[ \mathbf{D} \boxtimes_{n=1}^N \underbrace{\mathbf{U}_n \mathbf{w}_n^T(p_n)}_{\tilde{\mathbf{w}}_n^T(p_n)} \right] \begin{pmatrix} \mathbf{x}(t) \\ \mathbf{u}(t) \end{pmatrix}, \end{aligned} \quad (30)$$

where  $\mathbf{D}$  stands for the core tensor, and functions  $\tilde{w}_{n,j_n}(p_n)$  are the weighting functions [5] In order to reduce the number of LTI vertex systems the rightmost columns of matrices  $\mathbf{U}_n$  may be removed. Some recent applications related to LPV systems and TP model transformation can be found in [11],[12].

## 7. Future work and Conclusions

In the present paper a queuing approach for modeling logistical processes has been proposed. It was shown how the relation between parameters of such queuing models can be identified on subspace basis. The results clearly reflect that in case of strong relation between two system features the identified model nicely approximates the modeled system. On the other hand if this relation is weak the identified model (based on the corresponding input-output pairs) reflects significantly lower accuracy. In addition it was briefly shown how complex logistical systems could be modeled on LPV basis by blending locally identified linear models together. In our case the vertex systems stand for state space models identified based on input-output data. These models are then embedded into tensor representation and transformed into tensor product form. In such form model reduction can also be directly executed.

## References

- [1] van Overschee P, de Moor BL: Subspace Identification for Linear Systems: Theory - Implementation - Applications. Kluwer Academic Publishers, 2011.  
DOI: 10.1007/978-1-4613-0465-4
- [2] Harmati I, Orbán G, Várlaki P: Takagi-Sugeno Fuzzy Control Models for Large Scale Logistics Systems. International Symposium on Computational Intelligence and Intelligent Informatics, Agadir, pp. 199–203, 2007.  
DOI: 10.1109/ISCIII.2007.367389
- [3] Orbán G, Várlaki P: Fuzzy Modelling for Service Strategy and Operational Control of Loading Systems. Acta Technica Jaurinensis, Series Logistics, Vol. 2, No. 3, pp. 375–391, 2009.
- [4] De Lathauwer L, De Moor B, Vandewalle J: A multilinear singular value decomposition. SIAM Journal on Matrix Analysis and Applications, Vol. 21, No. 4, pp. 1253–1278, 2000.  
DOI: 10.1137/S0895479896305696
- [5] Szeidl L, Várlaki P: HOSVD Based Canonical Form for Polytopic Models of Dynamic Systems. Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 13, No. 1, pp. 52–60. 2009.
- [6] Nagy S, Petres Z, Baranyi P: TP Tool - a MATLAB Toolbox for TP Model Transformation. Proceedings of 8th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics, budapest, pp. 483–495, 2007.
- [7] Szeidl L, Baranyi P, Petres Z, Várlaki P: Numerical Reconstruction of the HOSVD Based Canonical Form of Polytopic Dynamic Models. International Symposium on Computational Intelligence and Intelligent Informatics, Agadir, pp. 111–116, 2007.  
DOI: 10.1109/ISCIII.2007.367372
- [8] Gunasekaran A, Patel C, McGaughey RE: A framework for Supply Chain Performance Measurement. International Journal of Production Economics, Vol. 87, No 3, pp. 333–347, 2004.  
DOI: 10.1016/j.ijpe.2003.08.003
- [9] Yao JS, Lin FT: Constructing a Fuzzy Flow-Shop Sequencing Model Based on Statistical Data. International Journal of Approximate Reasoning, Vol. 29, No 3, pp. 215–234, 2002.  
DOI: 10.1016/S0888-613X(01)00064-0
- [10] Sevastjanov PV, Róg P: Fuzzy Modeling of Manufacturing and Logistic Systems. Mathematics and Computers in Simulation, Vol. 63, No. 6, pp. 569–585, 2003.  
DOI: 10.1016/S0378-4754(03)00064-8

- [11] Baranyi P, Petres Z, Korondi P, Yam Y, Hashimoto H: Complexity Relaxation of the Tensor Product Model Transformation for Higher Dimensional Problems. *Asian Journal of Control*, Vol. 9, No. 2, pp. 195–200, 2007.  
DOI: 10.1111/j.1934-6093.2007.tb00323.x
  
- [12] Baranyi P, Tikk D, Yam Y, Patton RJ: From Differential Equations to PDC Controller Design via Numerical Transformation. *Computers in Industry*, Vol. 51, No. 3, pp. 281–297, 2003.  
DOI: 10.1016/S0166-3615(03)00058-7

# Design Considerations by the Optimization of PMS Motors

**B. Kollár**

**Széchenyi István University, Department of Automation  
Research Center of Vehicle Industry  
Egyetem tér 1, H-9026 Győr, Hungary  
e-mail: kollar.bettina@sze.hu**

**Abstract:** The determination of the targets, the input, implicit and design parameter is very important part in the design optimization of the permanent magnetic synchronous (PMS) motors for electric vehicles. The connections were defined between them, investigations were made which assisted to the appropriate optimization. In this paper the researches for the design optimization are presented which define the targets correctly and describes the results of the calculations and simulations of the slot fill factor.

**Keywords:** *PMS motor, optimization, demagnetisation, slot fill factor, penalty function, electric vehicle*

## 1. Introduction

It is important to aim to find the optimum of the specified parameters by the design of electric motors. Usually the target of the optimization is to minimise the losses.

The objectives like minimum building size, total cost, induced voltage, demagnetisation and size of the permanent magnets will be important too if the electric motor will drive a vehicle [1-4].

In this paper those design methods will be introduced which were necessary to execute the optimization of permanent magnetic synchronous motors.

## 2. The model of the PMS motor

Permanent magnetic synchronous motor was used with outer rotor construction during the design and optimization process. The motor will drive vehicle so the design parameters were planned along these lines. The first input of the process is the required driving cycle. The operating points of the vehicle are determined from the driving cycle in view of the mass of the vehicle, the transmission and the diameter of the wheels [6]. In addition the maximum outer diameter of the rotor is known before the design process which implies the building size of the motor.

In this work the finite element software ANSYS Maxwell is used to design, calculate and simulate the models of the motor. The model of the motor is shown in Fig. 2. The

Design and Optimization Toolbox of Matlab is used to makes the optimization process with multi-objective genetic algorithm. The design parameters of the optimization are illustrated in Fig. 1:

- tooth gap width ( $B_{s0}$ )
- tooth tang depth ( $H_{s0}$ )
- slot depth ( $H_{s2}$ )
- air gap thickness
- length of the stator
- magnet gap
- magnet thickness
- number of turns
- wire diameter
- tooth width ( $f$ )
- rotor outer diameter
- stator inner diameter

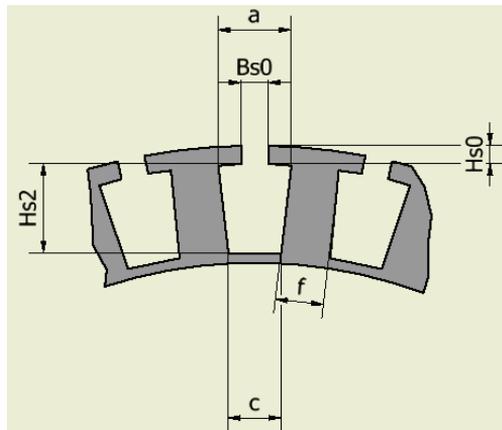


Figure 1. Sizes of the slot

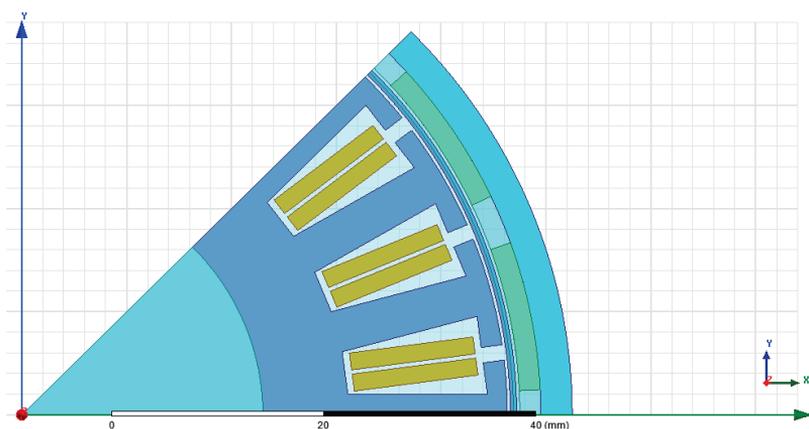


Figure 2. The 2D model of the motor-segment in ANSYS Maxwell

Lower and upper limit values were given to these parameters. The rotor outer diameter was handled like an input of the optimization as the upper limit value. The teeth of the stator are parallel with each other because of the smaller cogging torque [12].

The next list includes the implicit parameters which were determined from the design parameters:

- Wires of conductor
- Slot width in the upper part (a)
- Slot width in the lower part (c)
- Rotor inner diameter
- Stator outer diameter
- Slot fill factor
- Tooth gap width ( $H_{s0}$ )

The connections between the parameters are determined with equations, for example the winding size could not be bigger than the slot.

The targets of the optimization were to minimize the total losses, the weight of the motor and minimize the amount of the penalty functions. The calculation of the total loss was given by the next equations. The core loss ( $p_c$ ) is calculated from the amount of the hysteresis loss and the eddy current loss [5, 6].

$$p_c = K_h f |B|^2 + K_e (f|B|)^2 \quad (1)$$

where  $K_h$  is the hysteresis loss constant,  $K_e$  is the eddy current loss constant,  $B$  is the magnetic flux density and  $f$  is the sinusoidal varying frequency of  $B$ . The value of  $p_c$  is calculated by ANSYS Maxwell.  $K_h$  and  $K_e$  are the parameters of the chosen material in the software.

The winding loss was calculated according to this equation:

$$P_w = 3 \left( \frac{I^2}{\sqrt{2}} \right) R_w l N p \quad (2)$$

where  $I$  is the amplitude of the phase current,  $R_w$  is the resistance of the wire,  $l$  is the length of the stator,  $N$  is the number of the turns,  $p$  is the pole number of the motor.

The summarized loss is [6]:

$$P_{\text{sum}} = P_w + \int_V p_c \quad (3)$$

### 3. Determination of the penalty functions

The results of a multi-objective optimization can be displayed graphically more beneficial in 2D if the number of the objectives are not higher than 3 [1]. For this the third objective of the optimization is to minimize the sum of the torque ripple and the demagnetisation.

These parameters were defined with penalty functions which can handle the limits of the parameters and can give penalty when the value is higher than the specified limit. These functions are the logistic functions (P(t)) which is shown in Fig. 3.

$$P(t) = \frac{1}{1+e^{-t}} \quad (4)$$

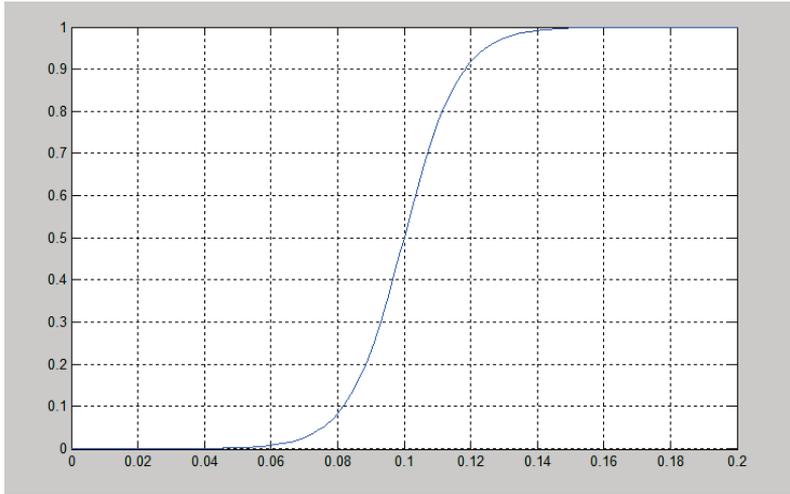


Figure 3. The logistic function

The limits were appointed by the definition of the penalty functions, the first is where the value of the given parameter is good (the penalty is close to 0) the other is where the value of the parameter is absolute incorrect (the penalty is close to 1).

### 3.1. The penalty function of the cogging torque

The torque of the motor is created by the interaction of the current which flows in the winding and the flux density distribution which was generated by the permanent magnetic rotor. As the flux density distribution and the stator currents change in time the torque which is constant in time is ensured for the motor, these were fitted to each other that the product of them is permanent. If it does not come into existence perfect there will be cogging torque.

The cogging torque has vibration induced effect to the mechanical elements of the vehicles so it is necessary to keep it under a determined limit [12]. The two limits of the logistic function are defined by the experience of the motor design and the literature research to 2,5% and 5%.

### 3.2. The penalty function of the demagnetization

The demagnetization curve is the part of the hysteresis loop curve where the magnetic flux density (B) is positive and the magnetic field intensity (H) is negative. This curve is important because the operation point of the built-in magnetized permanent magnet moves in the demagnetisation part of the hysteresis loop.

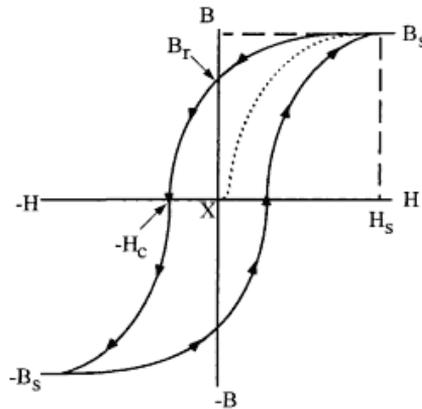


Figure 4. Hysteresis loop [7]

One of the important points is the remanent magnetic flux density ( $B_r$ ). It is the magnetic flux density which corresponds to zero field intensity. If it increases the flux and the inducted voltage of the motor get larger. The other point is the coercive field strength ( $H_c$ ) which releases the magnetic flux density to zero in the open-loop magnetic circle permanent magnet [8, 9, 13].

The remanent magnetic flux density and the coercive field strength are temperature-dependent values. If the temperature of the magnet transcended the specified limit the operation point of the magnet can slide down from the linear part of the demagnetization curve. After that the original magnetic flux density is not insured to get back to room temperature only if it will be magnetise again. The target is avoid the work in the linear part.

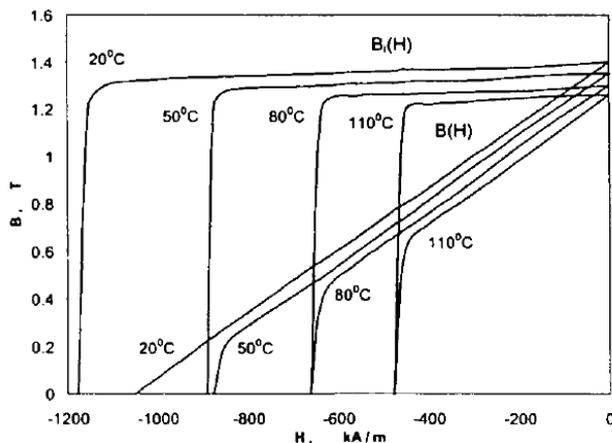


Figure 5. Demagnetisation curves and their variations with the temperature for sintered NdFeB [9]

The maximum operation temperature was specified like an input parameter of the optimization. Thereto the insulation class of the winding materials was considered and proportion the permanent magnet to this temperature.

The temperature-dependence can be chosen of the magnet material in ANSYS Maxwell. These data are in the catalogue of the magnet factories. These parameters are the temperature coefficient of the remanent magnetic flux density ( $RTC_{Br}$ ) and the temperature coefficient of the coercive field strength ( $RTC_{Hc}$ ) [8, 9, 13].

$$B_r(T) = B_{r20} \left( 1 - \frac{RTC_{Br}(T-20)}{100} \right) \quad (5)$$

where  $B_r(T)$  is the remanent magnetic flux density in the specified maximum operating temperature,  $B_{r20}$  the remanent flux density in room temperature (catalogue data),  $T$  is the specified maximum working temperature [9].

$$H_c(T) = H_{c20} \left( 1 - \frac{RTC_{Hc}(T-20)}{100} \right) \quad (6)$$

where  $H_c(T)$  is the coercive field strength in the specified maximum working temperature,  $H_{c20}$  is the coercive field strength in room temperature (catalogue data) [9].

The values of the remanent magnetic flux density and the coercive field strength in the maximum operating temperature should be considered at the same time. So the optimizer has to give penalty to the demagnetisation to the member of the optimization if the  $B_r(T)$  is higher and  $H_c(T)$  is lower than the calculated values.

#### 4. Simulations to determinate the slot fill factor

Currently the slots were handle with double layer winding. It reduces the cogging torque of the motor [10]. The insulation class of the winding materials are H (180 °C). During the reeling a slot liner is placed first in the slots, which thickness is 0.31 mm. The slots are closed with a slot wedge which is usually from bakelite material by this shape of slots. The thickness of it is 1 mm and it is placed under the tooth tang. The shape of the slots is trapeze, like in the Fig. 1. The sizes of the slot ( $a$ ,  $c$ ,  $H_{s2}$ ) change because of the slot liners and slot wedge. This area is calculated for each of the optimization's member.

$$T_{slotV} = \frac{a-(2 \cdot 0.31)+c-(2 \cdot 0.31)}{2} (H_{s2} - (2 \cdot 0.31 + 1)) \quad (7)$$

Round enamelled copper wires are used which are placed in the slots like in the schematic drawing in Fig. 6. When the wires are placed next to each others there are losses which are air. So a round wire occupies a hexagonal area in the slot.

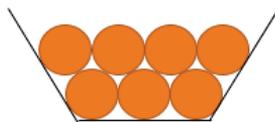


Figure 6. Round copper wires in the slot

In theory the slot fill factor can be better with wires with smaller diameter. Smaller current can flow across the smaller diameter wire because the resistance is bigger, so the winding losses will be bigger too. Therefore the software controls the winding losses according to the eq. (2) after it chose the size of the wire.

Table 1. The data of the wires [11]

Effective diameter	Enamelled diameter	DC resistance at 20 °C [ $\Omega/m$ ]
0.5	0.584	0.0871
0.53	0.738	0.07748
0.56	0.779	0.06940
0.6	0.823	0.06046
0.63	0.876	0.05484
0.67	0.928	0.4848

The slot fill factor was considered during the optimization like one member has the values  $a, c, Hs2$ . The software calculates the slot area with slot liner and slot wedge ( $T_{slotV}$ ) and without them too. Afterwards it calculates the maximum reachable slot fill factor.

$$T_{slot} = \frac{a+c}{2} Hs2 \quad (8)$$

$$k = 0,9075 \frac{T_{slotV}}{T_{slot}} \frac{T_w}{T_wE} \quad (9)$$

where  $k$  is the maximum slot fill factor,  $T_wE$  is the area of the enamelled wire,  $T_w$  is the effective area of the wire. The multiplier 0.9075 is the proportion of the areas of the circle and the hexagon.

After the software calculated the maximum slot fill factor it starts to determinate the number of wires per set with observance the number of turns ( $N$ ) and specifies the real slot fill factor:

$$nw = \frac{T_{slotV}}{2 T_{wH} N} \quad (10)$$

$$k_v = \frac{T_{slotV} \cdot T_w}{T_{wH} T_{slot}} \quad (11)$$

where  $nw$  is the number of wire per set  $T_{wH}$  is the hexagonal area of the wire which was calculated from the area of the enamelled copper wire,  $k_v$  is the real slot fill factor. The multiplier 2 comes from the double layer winding in eq. (10).

To the more precise value of the slot fill factor more simulations were made in the finite element software Infolytica to get the slot fill factor and wire diameters pairs. The fill factor was calculated for the different wire diameters and the results were compared. The simulations and the calculations were made for two different slot sizes, with double layer winding. The number of turns was 7 in both of the cases.

The software needed the value of the wire effective diameter and the number of wires per set. The table (see Table 2.) contains the wire sizes which were used during the comparison.

Table 2. Diameters of the wires [11]

<i>Wire diameter</i>	
<i>Effective</i>	<i>Enamelled</i>
0.500	0.584
0.530	0.738
0.560	0.779
0.600	0.823
0.630	0.876
0.670	0.928
0.710	0.980
0.750	1.032
0.800	1.083
0.850	1.145
0.900	1.209
0.950	1.271
1.000	1.343
1.060	1.415
1.120	1.498
1.180	1.600
1.250	1.701
1.320	1.804
1.400	1.908
1.500	2.012
1.600	2.113
1.700	2.237
1.800	2.358
1.900	2.482
2.000	2.625
2.120	2.778
2.240	2.930
2.360	3.132

The slot fill factor was calculated according to the eq. (9). The results are in Fig. 7. and Fig. 8.

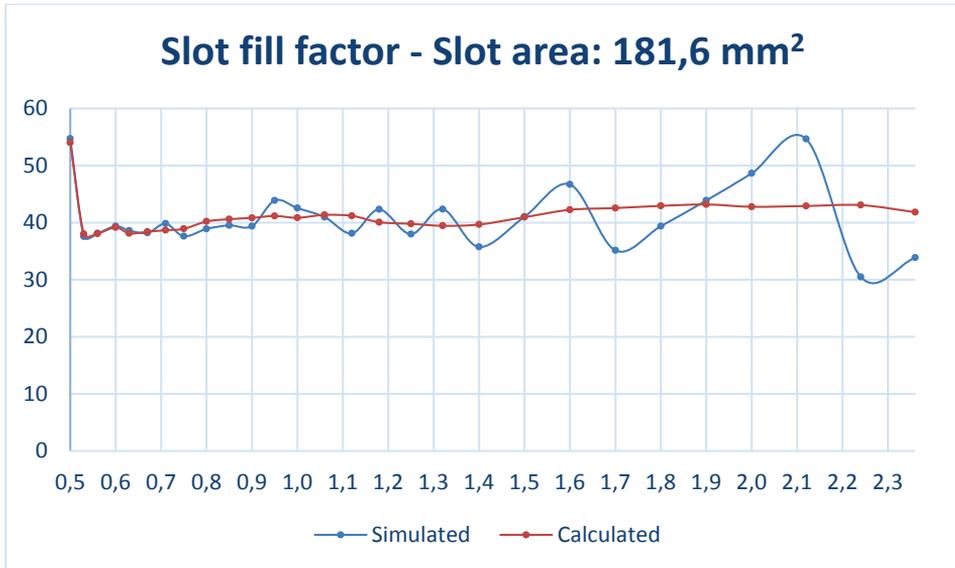


Figure 7. The slot fill factor results for the slot area 181,6 mm<sup>2</sup>

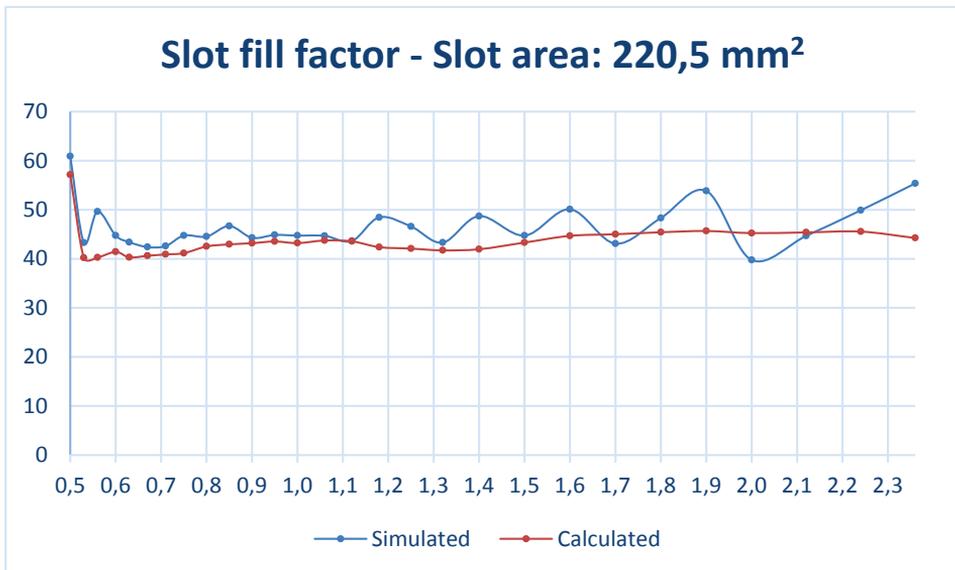


Figure 8. The slot fill factor results for the slot area 220,5 mm<sup>2</sup>

The calculated slot fill factor values are almost the same by the two slot sizes, the difference is marginal by the bigger slot size, the shapes of the curves are very similar. The simulated results are almost similar by the slot sizes but the main value of simulated fill factors by the bigger slot is 6% higher.

It would be useful to fill the slot with double layer windings with the same number of turns by different wire diameters and complete these results with the real values.

## **Conclusion**

In these researches the design was focused during the optimization to build the model correctly with the aware describing of the results. The optimization process varies the value of the parameters in wide ranges therefor it points out such kind of parameter combinations which were not expect during the design. The target was to make a general model of the outer rotor PMS motor and describe them prudently to use the model to give make good solutions for the different applications.

Later the optimization will be enlarged with better consideration the thermometric behaviour of the motor. The other plans are to complete the penalty function with the inducted voltage and to continue the investigation of the slot fill factor with check and complete the results with a real winding measurement.

## **Acknowledgement**

The research work presented in this paper was carried out as part of the TÁMOP-4.2.2.A-11/1/KONV-2012-0012 project in the framework of the New Széchenyi Plan. The realization of this project is supported by the European Union, and co-financed by the European Social Fund.

## **References**

- [1] Bittner F, Hahn I: Kriging-Assisted Multi-Objective Particle Swarm Optimization of Permanent Magnet Synchronous Machine for Hybrid an Electric Cars. IEEE International Electric Machines and Drives Conference, Chicago, IL, USA, 2013. DOI: [10.1109/IEEMDC.2013.6556123](https://doi.org/10.1109/IEEMDC.2013.6556123)
- [2] Duan Y, Harley RG: Method for Multi-Objective Design and Optimization of Three Phase Induction Machines. IEEE Transactions on Industry Applications, 2011. DOI: [10.1109/TIA.2011.2156372](https://doi.org/10.1109/TIA.2011.2156372)
- [3] Geest Mvd, Polinder H, Ferreira JA, Zeilstra D: Optimization and Comparison of Electrical Machines using Particle Swarm Optimization. XXth International Conference on Electrical Machines, Marseille, 2012. DOI: [10.1109/ICEIMach.2012.6350058](https://doi.org/10.1109/ICEIMach.2012.6350058)
- [4] Milot A, Korkosz M, Lukaniszyń M: Iron loss and eddy-current loss analysis in a low power BLDC motor with magnet segmentation. Archives of Electrical Engineering, 1. kötet, pp. 33-46, 2012. DOI: [10.2478/v10171-012-0003-5](https://doi.org/10.2478/v10171-012-0003-5)
- [5] Krotzsch J, Piepenbreie B: Hybrid Algorithm for Multi-objective Optimization of PMSM using massively distributed Finite Element Analysis. 12th International Conference on Optimization of Electrical and Electronic Equipment, 2010. DOI: [10.1109/OPTIM.2010.5510429](https://doi.org/10.1109/OPTIM.2010.5510429)
- [6] Kuslits M: Driving cycle based cost function for energetic optimization of PMS motors applied in electric vehicles. Workshop on Design, Simulation,

- Optimization and Control of Green Vehicles and Transportation, Győr, Hungary 2014.
- [7] Jordan MI: Why the logistic function? A tutorial discussion on probabilities and neural networks. Computational Cognitive Science, Massachusetts Institute of Technology, 1995.
  - [8] Mclyman CWT: Transformer and inductor design handbook, NY, U.S.A.: Marcel Dekker, 2004.  
DOI: [10.1201/9780203913598.ch8](https://doi.org/10.1201/9780203913598.ch8)
  - [9] Gieras JF: Permanent magnet motor technology: Design and applications, New York: MARCEL DEKKER Incorporated, 2002. ISBN: 0-8247-0739-7
  - [10] Szénásy I: Some actual questions at the development of up-to date PMSM motors. [Performance]. Workshop on Design, Simulation, Optimization and Control of Green Vehicles and Transportation, 2014.
  - [11] Specifications for particular types of winding wires - Part 0-1: General requirements - Enamelled round copper wire, IEEE Standard IEC 60317-0-1, 2013.
  - [12] Zhu ZQ, Howe D: Influence of design parameters on cogging torque in permanent magnet machines. IEEE Transactions on Energy Conversation, Vol. 15, No. 4, pp. 407-412, 2000.  
DOI: [10.1109/60.900501](https://doi.org/10.1109/60.900501)
  - [13] Hanselman D: Brushless permanent motor design. Orono: MAGNA PHYSICS PUBLISHING, 1994. ISBN: 1-881855-15-5

# A Review of Current Challenges in Production Planning

P. Szármes

Széchenyi István University, Multidisciplinary Doctoral School  
of Engineering Sciences  
Egyetem tér 1, H-9026 Győr, Hungary  
e-mail: peter.szarmes@sze.hu

**Abstract:** For manufacturing companies growing customer expectations, shortening product life cycles and increasing product variant numbers mean great challenges. Overcoming these is a major task for production planning, which is responsible for coordinating all business processes directly related to manufacturing. This paper presents important activities of production planning giving an integrative overview, and presents the digital factory concept with special attention to problems and their possible solutions. It concludes with challenges still waiting to be resolved and future development trends.

**Keywords:** *production planning, digital factory, manufacturing process optimization*

## 1. Introduction

Today's rapidly changing business environment with ever increasing customer expectations creates great challenges for manufacturing companies: they must quickly adapt their products, manufacturing processes and even their technologies to current demands and new technical solutions. Failing to do so may result in a loss of market share and eventually in financial troubles since manufacturing happens to be their most expensive and time-consuming activity.

Developing, testing and introducing new production processes and technologies is a major task of production planning. Production planning is a planning and coordinating activity embracing the entire manufacturing process: it involves not only planning but also keeping close contact with machine and technology suppliers, quality managers and manufacturing engineers in order to control and support the whole manufacturing process.

On this department hinges to a great extent the competitiveness of manufacturing companies. Not only technical constraints have to be considered here, but time, quality and economic factors as well. Production planning must ensure that the company carries out its manufacturing operations with modern and appropriate technology and well-trained workforce through effective and efficient processes.

The digital factory concept is an integrated approach in production planning, using various tools to enhance the design and manufacturing processes of products. It can be also seen as a business initiative that aims to manage information and collaboration in order to solve manufacturing problems and facilitate volume and variant production, which are determining factors in decreasing costs and increasing sales revenue [1].

## **2. Production planning in general**

Production planning is a complex activity bridging the gap between product design and production and involves all activities related to setting up a production system with all the resources and processes [2]. The benefits from a good product design can be lost easily if delays or cost overruns occur in production due to inefficient, expensive and unpredictable manufacturing processes.

Production costs are determined mostly by product characteristics but production process properties also have a significant impact. Since there is a constant price pressure in the market, cost cutting measures have to start in the product design phase, but then production processes must be planned with cost-consciousness as well [3].

Manufacturing companies must therefore steadily develop flexible and efficient production processes. New methods and production techniques must be introduced constantly to keep pace with external developments and satisfy the customer needs for customized and personalized products, faster deliveries, better service offerings and so on.

In sections 3-6 major parts of production planning (product engineering, production engineering, logistics optimization, operative production control) are examined in a more detailed way with special attention to the digital factory concept. Section 7 reviews challenges and difficulties in the implementation of the digital factory concept and gives a short description of a promising new type of manufacturing execution systems: the holonic manufacturing execution system.

## **3. Product engineering**

The increasing involvement of customers in the design (or even in the concept phase) of the product they are going to purchase requires a stronger collaboration between design and marketing people, customers and suppliers through co-development or simply through web interfaces [4]. This customization or even personalization coupled with other services is more and more important in satisfying individual customer needs.

The consequence of this co-design or co-creation is a growing number of variants and thus increasing complexity on the production lines. There are however design principles to tackle this problem, like modularization and product families. Products are often built from standardized modules whose designs are reused and updated. A product family is a group of similar products from similar components, which can be built on flexible production lines with minimal setting modifications [5].

The digital factory concept strives to streamline and optimize this complex product engineering process through various IT tools, like CAD systems which also allow collaboration among all teams working on product design and manufacturing. Perhaps

the most important element of product engineering is product lifecycle management (PLM).

PLM aims to manage all aspects of product engineering from design inception through manufacturing and maintenance to its disposal. It helps to enforce company specific engineering workflows, what can increase product flexibility, facilitate design and manufacturing concurrency through reuse of components, avoid design flaws as early as possible, enable collaboration among design and manufacturing engineers, suppliers and other partners, and integrate different IT tools into a coherent process [6].

#### 4. Production engineering

Production engineering or production planning focuses on the optimization of manufacturing processes. It encompasses numerous elements from layout design through material flow optimization, production line and process optimization to the simulation of machine operations. The general objectives are to shorten time-to-market and time-to-volume, improve production efficiency through optimizing necessary investments in machines, buffer levels and scrap ratio.

Computer planning, modelling and simulation techniques enable a thorough and detailed analysis that ensure that design problems and waste in production processes are discovered before the company ramps up for production and also help to start efficient volume production sooner.

##### 4.1. Layout planning

There are special CAD tools for factory layout planning where a detailed factory model with all its production lines can be built from predefined elements (conveyors, cranes, containers and even machines). With three-dimensional objects a 3D factory model is the result, enabling virtual walks, measurements and inspections in a not yet existing factory.

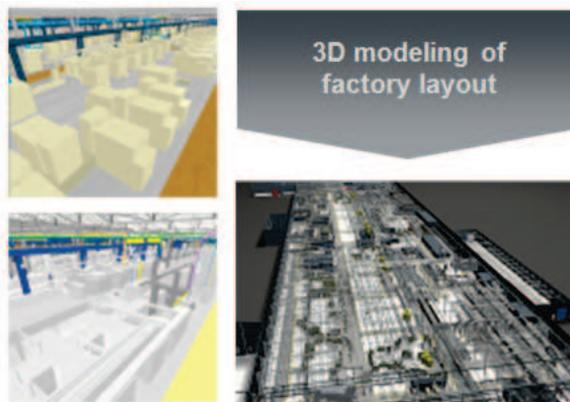


Figure 1. HLS 3D layout (preliminary, detailed, final plans) [7]

3D modeling facilitates earlier design changes for a machine or the production line (if necessary), and in this way modeling enables the timely and cost efficient implementation of projects (see Fig. 1).

#### **4.2. Simulation of production processes**

Simulation is a crucial part of digital manufacturing and is used as a planning aid for analyzing and planning complex manufacturing and logistics systems. The rather static view of the planned factory or production line in the CAD system can be extended by simulation software. Simulation can create a dynamic view of the production process based on machine parameters (e.g. loading, unloading and processing times) and material flow parameters (distances, frequencies etc.).

These simulation models allow the analysis of production line variants, what-if analyses and thus help to optimize material flow, resource utilization and logistics before even investing in the components of the new production line. Bottlenecks and problems can be detected in advance avoiding costly corrections during volume production. Simulation models can also be used to optimize an already running production line trying out planned measures in a virtual reality.

#### **4.3. Simulation of manufacturing operations**

Simulation of manufacturing operations means to simulate operation steps in processing NC/CNC machines and robotic workcells in order to calculate operating cycle times, error rates, resource utilization, work-in-process and buffer sizes. It also enables to detect collisions of manufacturing instruments or robots. Robotic work cells are complex equipment where the kinematics of robot components should be analyzed to generate optimal sequence of operations, avoid collisions and check whether robotic arms can reach the points of products to be processed. These tools can greatly help and accelerate the creation of robot programs [6].

#### **4.4. Simulation of human resources**

Today there are software tools even for simulating manual operations performed by human operators in order to optimize execution times and prevent work-related injuries. These software tools support the detailed design of operations, calculate execution time, make ergonomic analyses and generate work instructions.

### **5. Logistics optimization**

The optimization of production systems is a very complex multi-criteria decision-making task encompassing layout optimization, fine-tuning of certain system parameters and generating production program [9]. On a strategic level the whole supply chain must be considered since it is an interdependent network of business partners and its characteristics may heavily influence tactical and operational decisions.

Layout optimization means the exact configuration of production line components (machines, conveyors, containers etc.) This activity was already mentioned in the previous section about production engineering since it is subject to technical constraints

before logistical ones and should be decided early upon because major changes can prove quite costly afterwards.

Fine-tuning of parameters involves selecting optimal production and logistical characteristics of the production system, e.g. number of production and logistics employees, number and type of forklift trucks, inventory levels, frequency of deliveries from or to the warehouse etc. [9] These parameters can be modified more or less easily even in a running system, unlike configuration changes for an already existing production line.

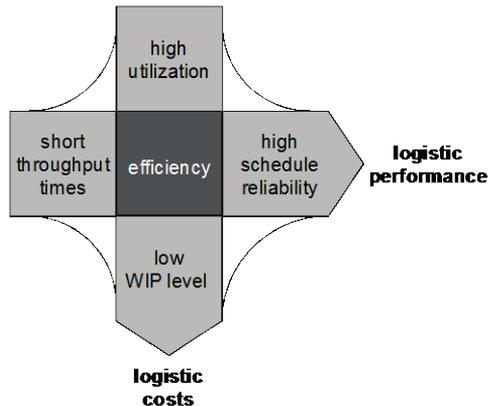


Figure 2. Objectives in production logistics [10]

When generating actual daily or weekly production programs decisions must be made which order to be processed on which production line/machine (mapping) and in which sequence and time (sequencing and scheduling). The production program has a high impact on delivery time and delivery reliability, which are important buying criteria beside price and quality. High delivery reliability and short delivery times require high schedule reliability and short throughput times, and logistics costs should also be optimized as Fig. 2 shows [10].

These logistics objectives are conflicting (it is called the scheduling dilemma) since it is not possible to maximize the utilization of a system while also minimizing throughput times. High utilization demands higher work-in-progress levels, which lead to longer throughput times, and also reduce schedule reliability. Hence a rational trade-off must be reached between these objectives to ensure a satisfactory level for all of them [10].

These decisions can be optimized using methods from operations research and it can be shown that many resource allocation problems can be formulated as special Markov decision processes [11]. Markov decision problems can be solved by dynamic programming, which solves complex problems by breaking them down into simpler sub problems.

## **6. Operative production control**

Production management system ensures that the planned production program is properly executed on the shop floor and enables capturing and sharing real-time manufacturing data for control and analysis. This system helps to improve order allocation on production lines, to make necessary program adjustments more easily, improve quality, reduce work-in-progress and scrap costs and creates better transparency about production processes [6].

Product and production tracking technologies (e.g. RFID technologies) can automatically gather data and provide a detailed real-time perspective about manufacturing and logistics operations [12]. The real-time data from the shop floor about operations, equipment status, scrap products, work-in-progress levels enables rapid problem detection and intervention in production processes, supports various analyses for improving efficiency and helps to create more realistic simulations.

The products themselves can become important control elements but this requires steady access to their relevant properties stored on their RFID chips. The Internet of Things (IoT), a relatively new paradigm means “connectivity to anything anytime” [13]. It integrates several technologies: identification and tracking, wired and wireless sensor and actuator networks, embedded intelligence for smart objects. The Internet of Things may revolutionize manufacturing processes and help to realize the full potential of the digital factory concept. The ultimate goal is to create manufacturing systems whose elements have cognitive capabilities [14].

## **7. Challenges and future developments in production planning**

A key prerequisite for the digital factory concept is the integration of IT tools supporting product design, production engineering and planning, simulation and shop floor control. It can be mostly accomplished by data integration via a central data warehouse, but it also requires interoperability between various software components and networks, and information transparency across different business partners and their IT tools. This integration and sharing of information links the enterprise resource planning system and other planning systems to the operative production control system and this system to the network of machines, sensors and actuators on the factory floor. These linkages are the essence of the digital manufacturing concept [15].

However these IT tools in the developmental, design, and planning stages are so numerous, and often differ from company to company across the supply chain, that creating all the necessary interfaces is an almost impossible and very costly task. The required standardized networking of these tools is therefore hardly feasible. In addition to this problem of compatibility the problems of version and data management should be solved as well. These difficulties nevertheless generate a strong motivation to create industry-independent open standards and frameworks. These standards with the potential benefits inherent in the paradigm of the Internet of Things may overcome the difficulties mentioned above and fully realize the digital factory concept.

There is much research on a new type of manufacturing execution system called the holonic manufacturing execution system. Its coordination and control mechanisms are inspired by natural systems, i.e. food foraging behavior in ant colonies [16]. This

control by distributed intelligence or swarm intelligence is very different from hierarchical control structures in traditional systems, and it can be realized using IoT technologies.

The holonic system is based on holons, which are single independent units (similar to agents) communicating with each other, thus also functioning as parts of a larger system. In the PROSA system architecture three types of holons (product holon, resource holon and order holon) have been defined. Resource holons describe the manufacturing resources containing data about manufacturing physics and all necessary control systems. These holons are responsible for resource scheduling and allocation. Product holons are abstractions of products with product design and product manufacturing information. Order holons represent orders for products. The three types of holons together establish a distributed intelligence system for manufacturing control [16].

These systems are still under development but promise numerous benefits. They show a fractal design, which repeats itself on the various levels of a production network, lending scalability, robustness and high adaptability to manufacturing control.

## **8. Conclusion**

Production planning plays a crucial role in the competitiveness of manufacturing companies and it might gain even greater significance in the future as digital factory concept and the Internet of Things paradigm unfolds. It is almost an art to create large and complex manufacturing systems from people, machines, materials and processes. These systems have to be not only efficient but also robust and adaptable. Building such systems requires a holistic view with steady learning and incessant developing because technology progresses rapidly while customer needs are also changing.

Embracing new concepts, methods and technologies, however, brings no automatic benefits. These tools must be tuned to a company's specific methods of operation, integrated into its business processes and aligned with its goals and objectives to realize their full potential.

## **References**

- [1] Schraft RD, Bierschenk S: Digitale Fabrik und ihre Vernetzung mit der realen Fabrik. Zeitschrift für wirtschaftlichen Fabrikbetrieb, Vol. 100, No. 1-2, pp. 14-18, 2005.
- [2] Dangelmaier W: Fertigungsplanung - Planung von Aufbau und Ablauf der Fertigung. Springer, Berlin, 2001.
- [3] Menges R: Frühzeitige Produktbeeinflussung und Prozessabsicherung. Zeitschrift für wirtschaftlichen Fabrikbetrieb, Vol. 100, No. 1-2, pp. 25-31, 2005.
- [4] Tomiyama T, Gu P, Jin Y, Lutters D, Kind C, Kimura F: Design Methodologies: Industrial and Educational Applications. CIRP Annals – Manufacturing Technology, Vol. 58, No. 2, pp. 543-565, 2009.  
DOI: [10.1016/j.cirp.2009.09.003](https://doi.org/10.1016/j.cirp.2009.09.003)

- [5] Vánca J, Monostori L, Lutters D, Kumara SR, Tseng M, Valckenaers P, Van Brussel H: Cooperative and responsive manufacturing enterprises. *CIRP Annals - Manufacturing Technology*, Vol. 60, No. 2, pp. 797-820, 2011.  
DOI: [10.1016/j.cirp.2011.05.009](https://doi.org/10.1016/j.cirp.2011.05.009)
- [6] Kühn W: Digital factory - simulation enhancing the product and production engineering process. *Proceedings of the 2006 Winter Simulation Conference IEEE*, pp. 1899-1906, 2006.  
DOI: [10.1109/WSC.2006.322972](https://doi.org/10.1109/WSC.2006.322972)
- [7] Audi Hungaria Motor Kft., Production Planning Department
- [8] Verbund der Deutschen Automobilhersteller: Ausführungsanweisung: Ablaufsimulation in der Automobil- und Automobilzuliefererindustrie, V. 2.0. VDA, 2008.
- [9] Kapp R, Löffler B, Wiendahl HP, Westkämper E: The logistics bench: Scalable logistics simulation from the supply chain to the production process. *CIRP Journal of Manufacturing Systems*, Vol. 34, pp. 5-54, 2005.
- [10] Wiendahl HP: Fertigungsregelung: Logistische Beherrschung von Fertigungsabläufen auf Basis des Trichtermodells. Carl Hanser, München, 1997.
- [11] Monostori L, Csáji BC, Kádár B, Pfeiffer A, Ilie-Zudor E, Kemény Z, Szathmári M: Towards adaptive and digital manufacturing. *Annual Reviews in Control*, Vol. 34, No. 1, pp. 118-128, 2010.  
DOI: [10.1016/j.arcontrol.2010.02.007](https://doi.org/10.1016/j.arcontrol.2010.02.007)
- [12] Monostori L, Kemény Z, Ilie-Zudor E, Szathmári M, Karnok D: Increased Transparency Within and Beyond Organizational Borders by Novel Identifier-Based for Enterprises of Different Size. *CIRP Annals - Manufacturing Technology*, Vol. 58, No. 1, pp. 417-420, 2009.
- [13] ITU Internet Reports: The Internet of Things. Geneva: Int. Telecommunication Union, 2005.
- [14] Záh MF, Reinhart G, Ostgathe M, Geiger F, Lau C: A Holistic Approach for the Cognitive Control of Production Systems. *Advanced Engineering Informatics*, Vol. 24, No. 3, pp. 300–307, 2010.  
DOI: [10.1016/j.aei.2010.05.014](https://doi.org/10.1016/j.aei.2010.05.014)
- [15] Kapp R, Aldinger L, Westkämper E: Real-time Factory Cockpit System. Implementation and Assurance of Economic Transformability, Glasgow: *Proceedings of the 20th Computer-Aided Production Engineering*, pp. 99-108, 2007.
- [16] Valckenaers P, Van Brussel H: Holonic Manufacturing Execution Systems. *CIRP Annals – Manufacturing Technology*, Vol. 54, No. 1, pp. 427–432, 2005.  
DOI: [10.1016/S0007-8506\(07\)60137-1](https://doi.org/10.1016/S0007-8506(07)60137-1)