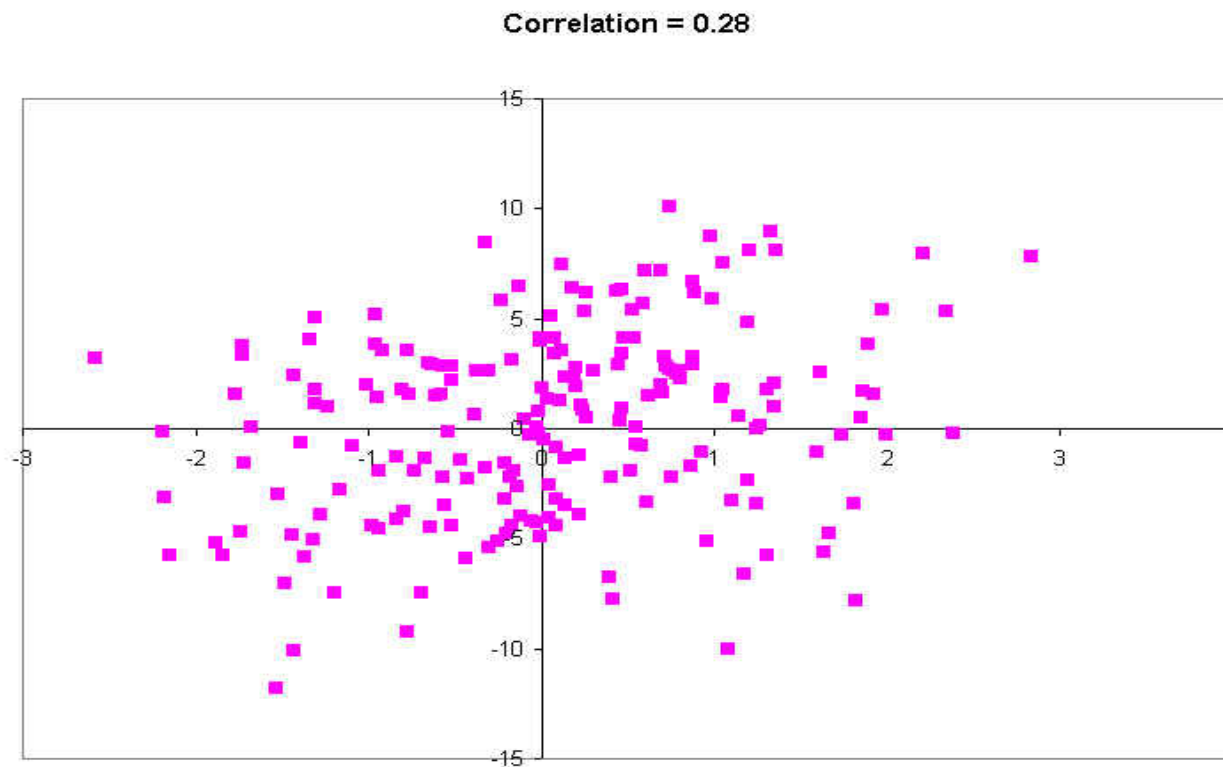


Filtrace a předzpracování dat

Korelace



Konrad Tvrdy
9.12.2010

Obsah

- Teorie – co je to korelace
- Využití korelačního koeficientu v dopravě

Korelace - teorie

- Popisuje vzájemný vztah (závislost) mezi dvěma veličinami.
- Pearsonův korelační koeficient:

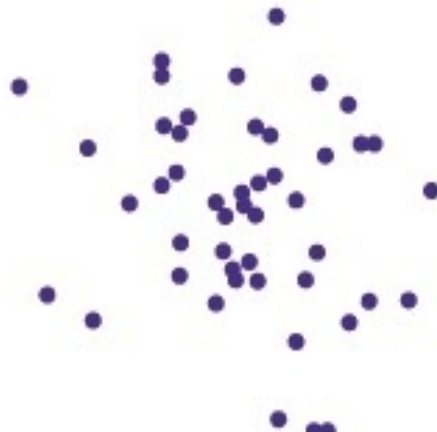
$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

- Nabývá hodnot z intervalu $\langle -1, 1 \rangle$
- 0 – veličiny jsou nezávislé
- 1, -1 – veličiny jsou lineárně závislé

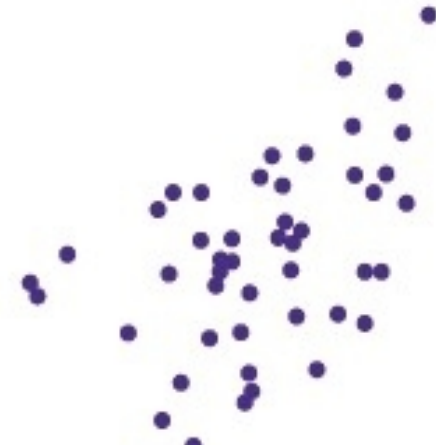
Korelance - teorije



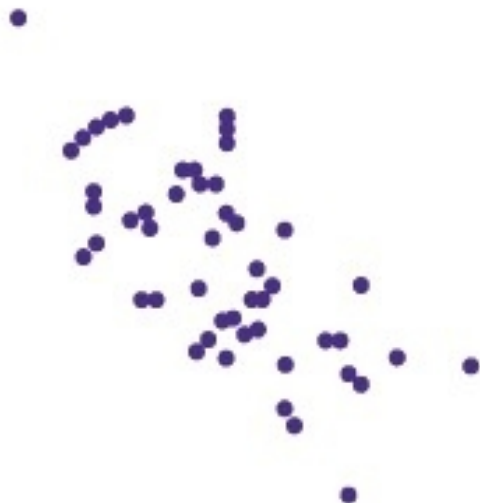
Correlation $r = 0$



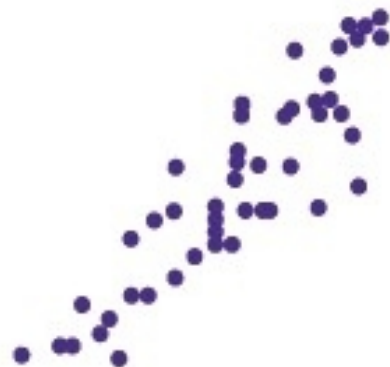
Correlation $r = -0.3$



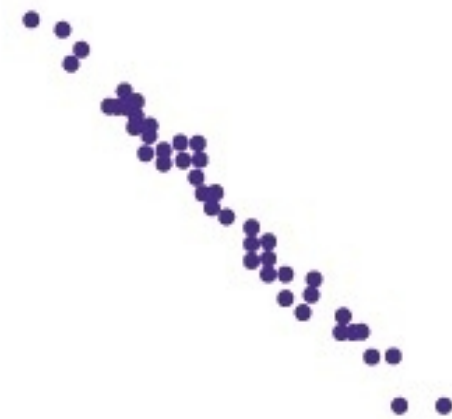
Correlation $r = 0.5$



Correlation $r = -0.7$



Correlation $r = 0.9$



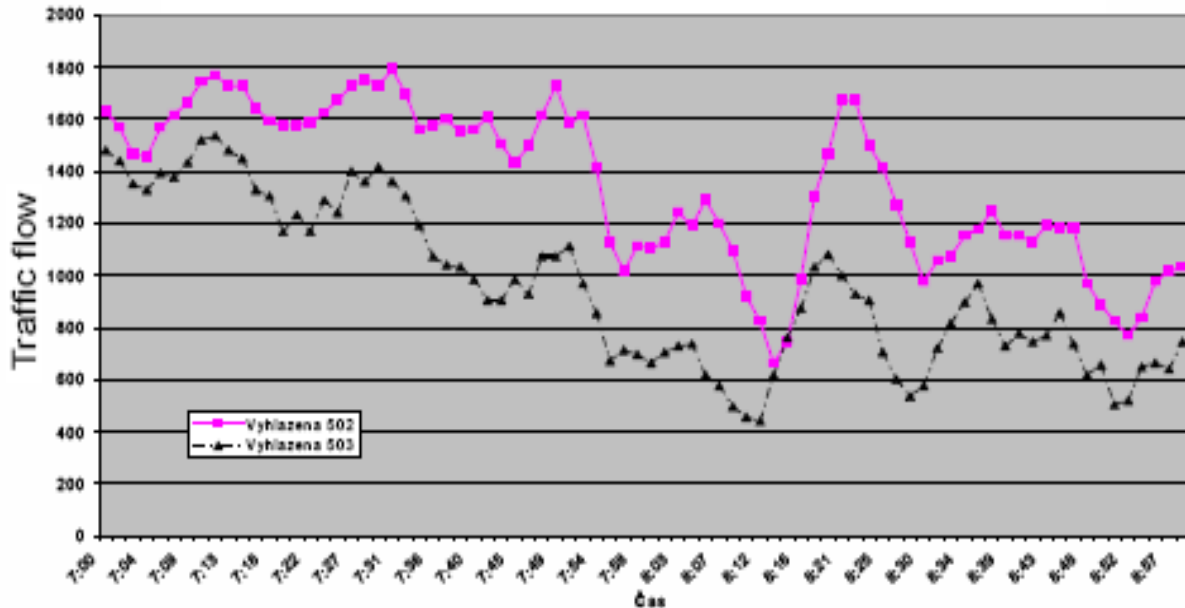
Correlation $r = -0.99$

Oblasti využití korelačního koeficientu v dopravě

- Řízení dopravy
 - **získávání dat z indukčních smyček**
 - **odhad doby jízdy**
- Souvislosti v dopravě
 - **intenzita x koncentrace škodlivých látek v ovzduší**
 - intenzita x HDP
- Plánování dopravy

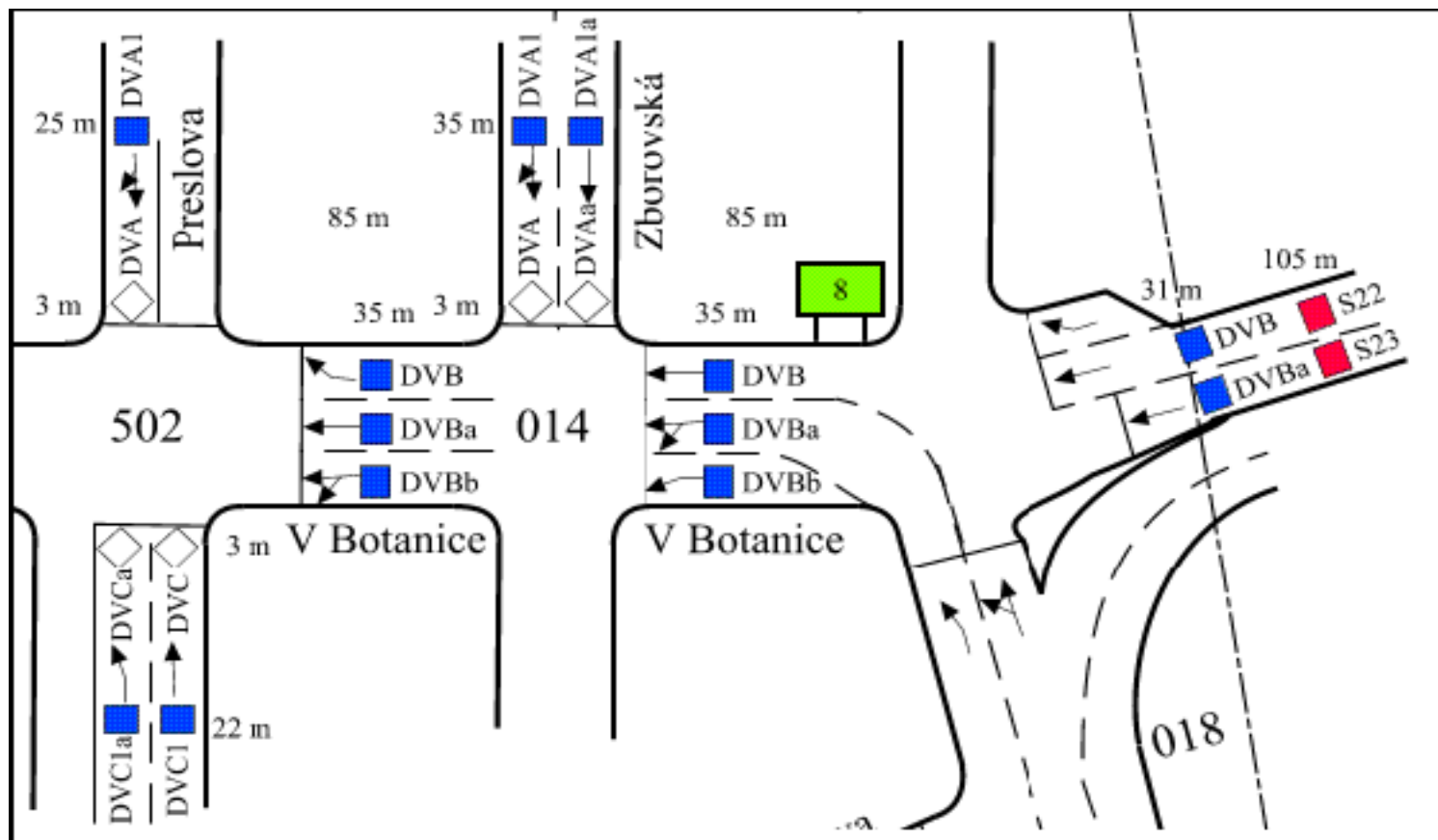
Získávání dat z indukčních smyček

- Velké množství indukčních smyček, velký objem dat se špatně zpracovává.
- Smyčky umístěné za sebou v jednom dopravním proudu mají velmi podobné výstupy.



Získávání dat z indukčních smyček

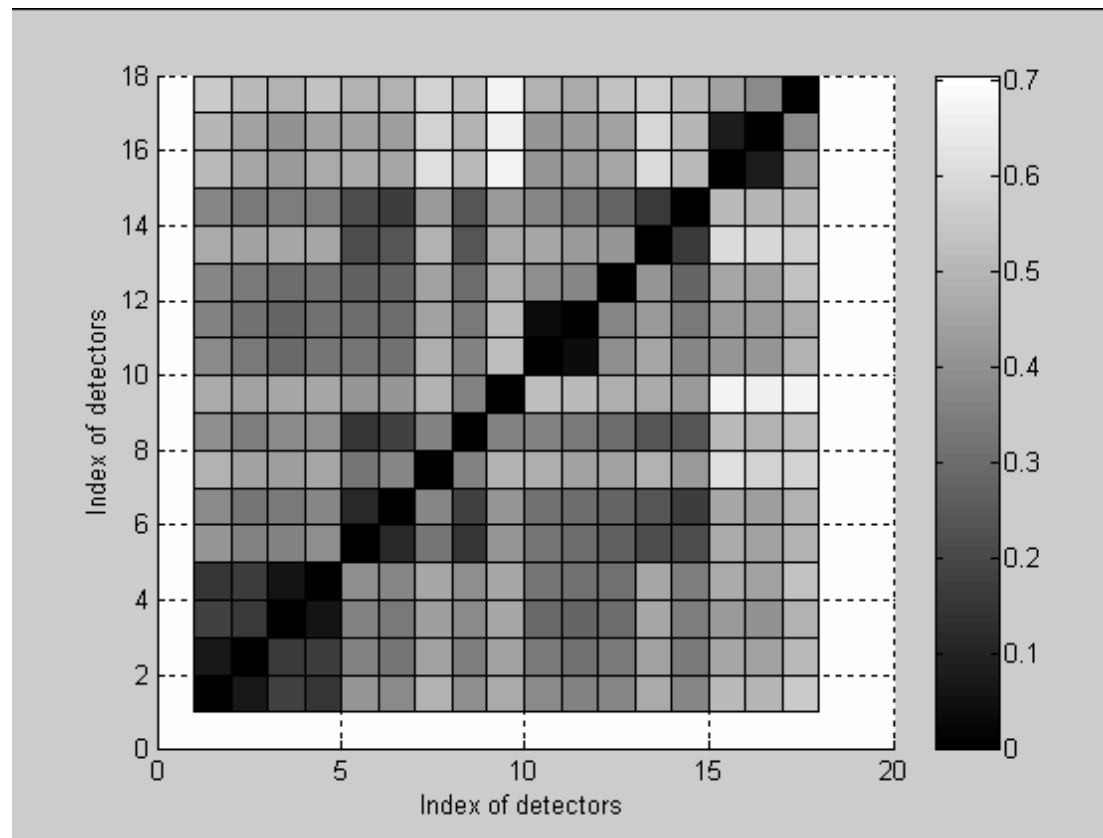
- Řešená oblast Praha – Smíchov.
- K výběru relevantních detektorů se používá shluková analýza.



Získávání dat z indukčních smyček

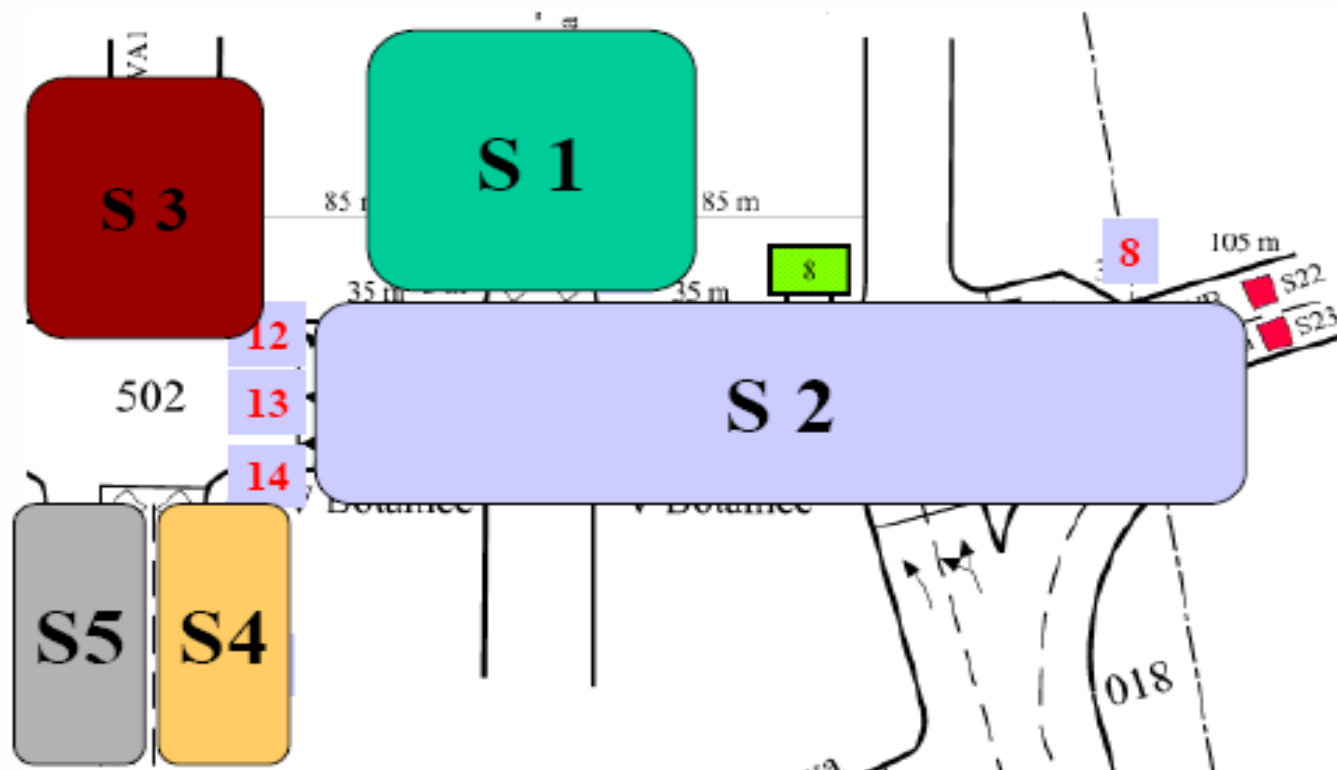
- Vstupem shlukovacího algoritmu je matice odlišností vycházející z korelací mezi výstupy z jednotlivých smyček.
- Odlišnost $d(x,y)$ je definována:

$$d(x, y) = 1 - |r_{xy}|.$$



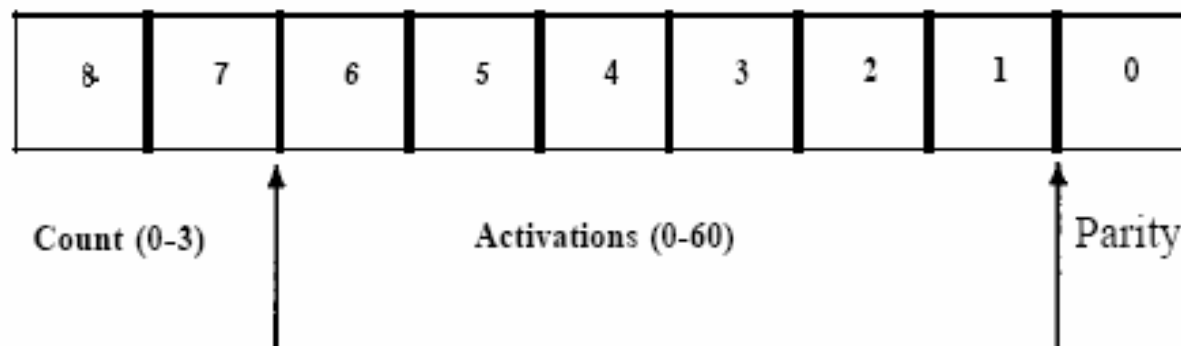
Získávání dat z indukčních smyček

- Shlukovací algoritmus vyhodnotí optimální počet shluků (clusterů) a jejich složení.



Odhad doby jízdy

- Na dálnici I5 v Seatlu, Washington jsou indukční smyčky v rozestupech 0,5 mil (800m)
- Indukční smyčka pořizuje 60 „snímků“ za sekundu. Poskytuje informace o 2 typech událostí:
 1. Počet aktivací za 1s (přechod z OFF na ON). Nabývá hodnot 0–3. Intenzita.
 2. Počet ON snímků za 1s. Nabývá hodnot 0-60. Obsazenost.
- Zaslání dat do centrály – 9-bitové komunikační schéma

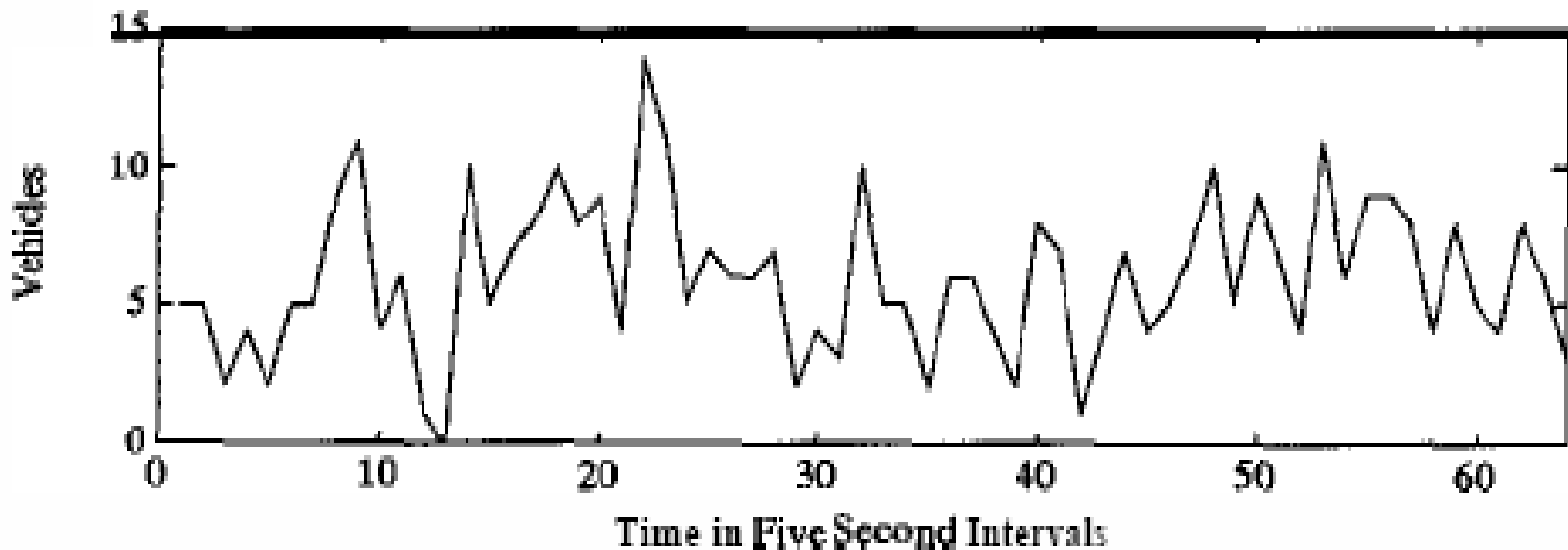


Odhad doby jízdy

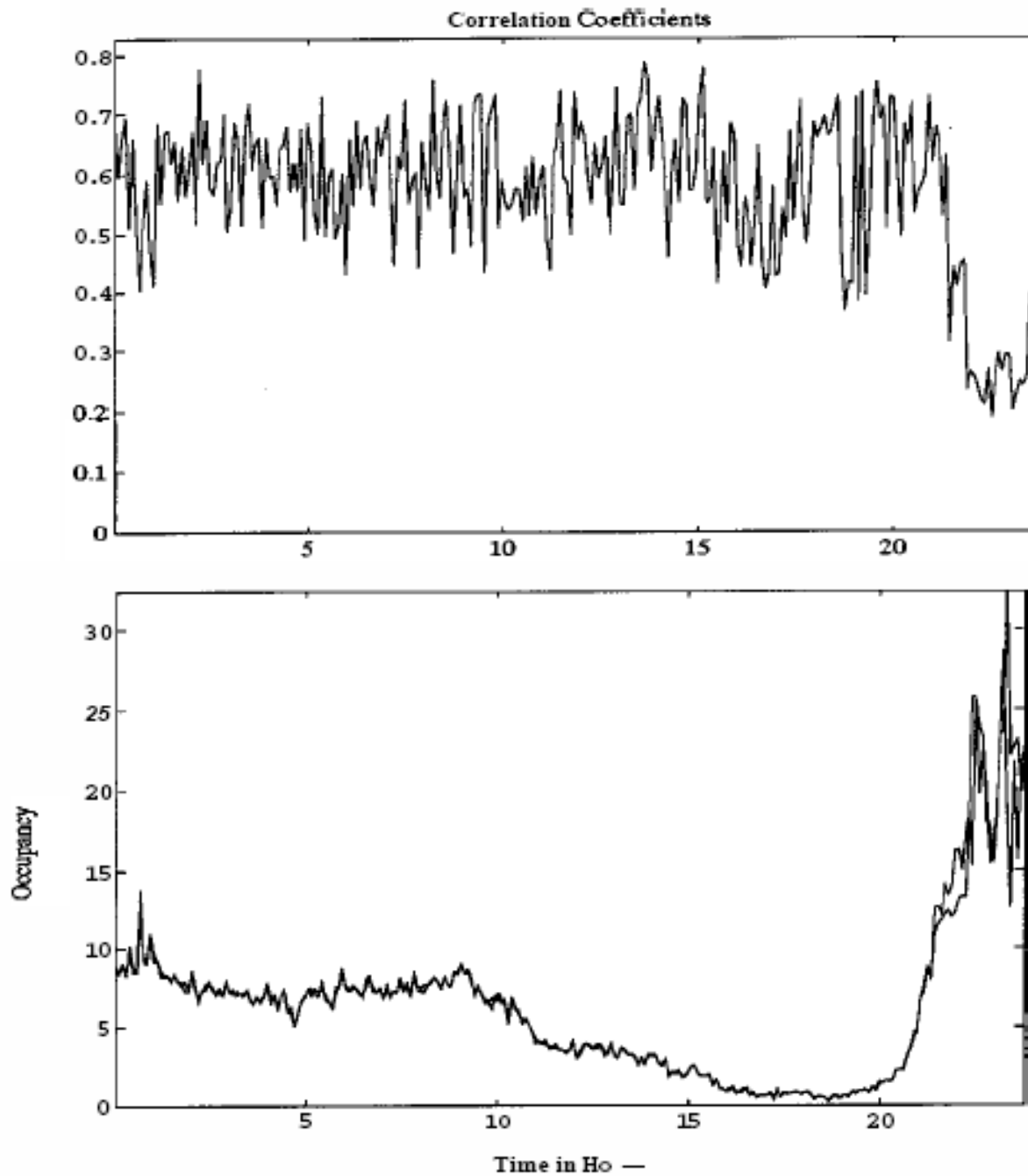
- Data z jednotlivých smyček jsou synchronizována.
- Prostorový průměr – sjednocení dat (naměřených intenzit) ze všech smyček v jedné řadě.
- Časový průměr – jednosekundové intenzity jsou sjednoceny do pětisekundových úseků.
- Jeden soubor obsahuje 64 pětisekundových úseků.
- Teď máme pětiminutový záznam fluktací intenzit na každé sadě indukčních smyček.

Odhad doby jízdy

- Pomocí metody vzájemné korelace dvou signálů se zjišťuje kolik % informací z 1. signálu se objeví ve 2. signálu. Výstupem je korelační koeficient (např. pro 77,5% shody $r = 0,6$).
- Doba jízdy se určí jako střední hodnota zpoždění mezi shodnými částmi signálů ze dvou sad smyček vzdálených 0,8 km.



Odhad doby jízdy



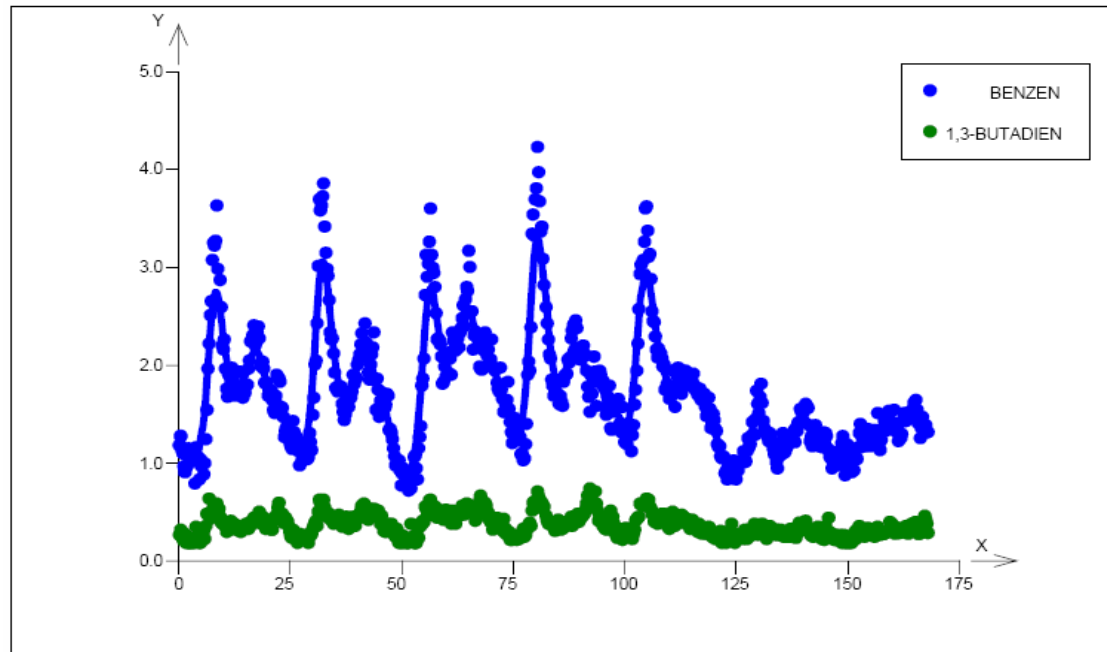
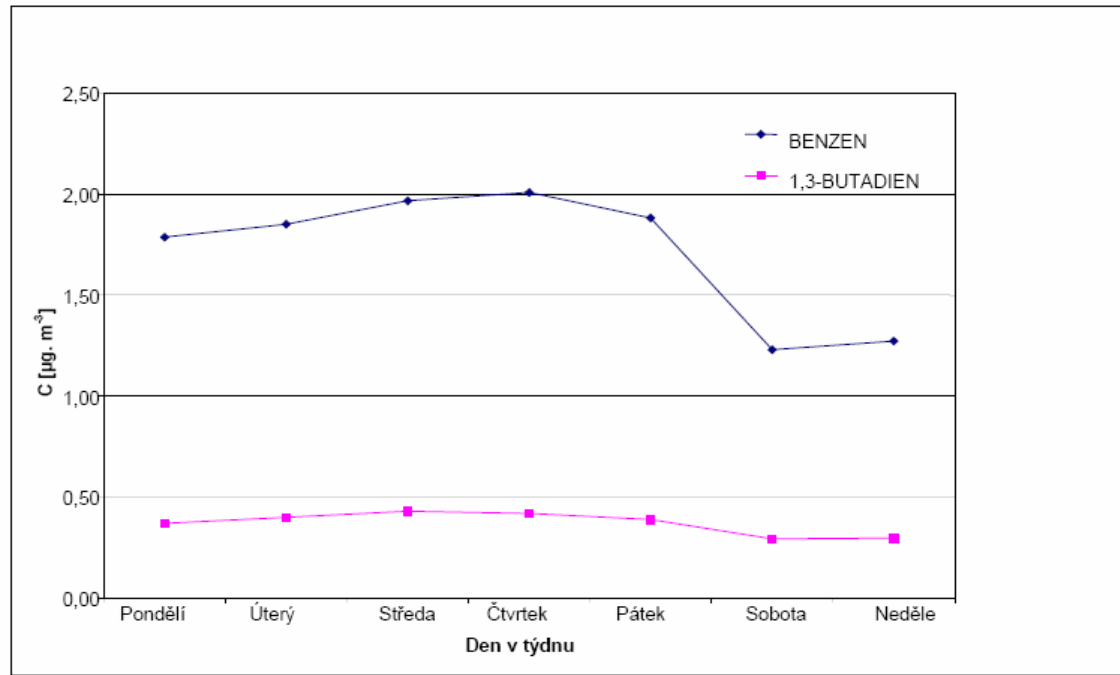
Odhad doby jízdy

- Z grafů je patrné, že hodnota korelačního koeficientu klesá s rostoucí obsazeností.
- Aby odhady doby jízdy byly relevantní, musí být $r \geq 0,4$. To nastává při obsazenosti $< 15\%$.
- Obsazenost přesáhne 15% jen během krátké doby v ranní špičce. Po většinu dne je hodnota korelačního koeficientu $> 0,5$ a odhad doby jízdy je tedy relevantní.

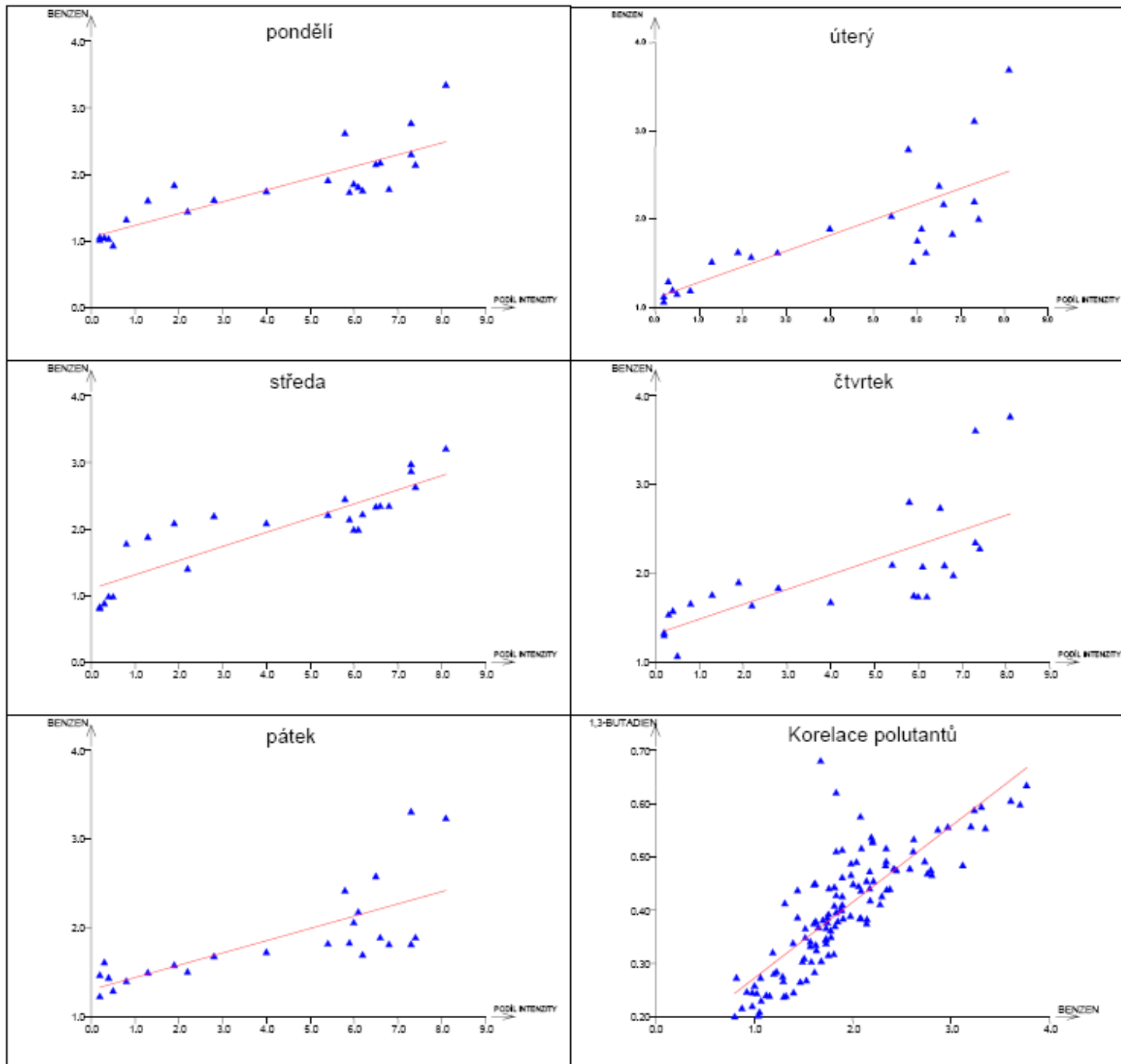
Intenzita x znečištění ovzduší

- Od 26.2.2004 do 30.6.2004 probíhalo v patnáctiminutových intervalech měření koncentrace benzenu a 1,3-butadienu v ovzduší.
- Mezi průměrnými denními obsahy v pracovních dnech nejsou statisticky významné rozdíly. Rozdíly jsou mezi hodnotami v pracovních dnech a o víkendech.
- Čtvrthodinové průměry – dvě denní maxima. Ráno 7 – 9h a odpoledne 15 -17h.
- Korelace mezi intenzitou a koncentracemi benzenu za jednotlivé dny.
- Vzájemné korelace hodinových průměrů benzenu a 1,3-butadienu za celý pracovní týden.

Intenzita x znečištění ovzduší



Intenzita x znečištění ovzduší



Intenzita x znečištění ovzduší

- Všechny korelace jsou statisticky významné. Ukazují, že denní výkyvy koncentrací polutantů v ovzduší byly způsobeny dopravními zdroji.

Korelované veličiny	Korelační koeficient		Dny
	Jednotlivé dny	Průměr za pracovní týden	
benzen - podíl intenzity	0,845	0,829	Pondělí
benzen - podíl intenzity	0,777		Úterý
benzen - podíl intenzity	0,885		Středa
benzen - podíl intenzity	0,726		Čtvrtek
benzen - podíl intenzity	0,731		Pátek
1,3-butadien - podíl intenzity		0,816	Pondělí - Pátek
benzen - 1,3-butadien	0,824	0,910	

Zdroje

- [1] Příbyl, O., Příbyl, P. Mining data from induct loops. Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE.
- [2] Dailey, D. J., Haselkorn, M. P., Nihan, N. L. Travel time estimation using cross correlation techniques. University of Washington, Seattle, Washington 1991.
- [3] Huzlík, J., Adamec, V. Koncentrace benzenu a 1,3-butadienu v městském ovzduší silně zatíženém dopravou. Centrum dopravního výzkumu 2004.
- [4] wikipedia.org



Děkuji za pozornost