

Statistický model obsazenosti dálničních parkovišť

Brabec, M.-Konár, O.-Malý, M.-Kasanický, I.-Pelikán, E.

mbrabec@cs.cas.cz

*Seminář strojového učení a modelování,
24.4.2013, MFF UK & UI AV ČR*

Projekt

- TAČR projekt 2012-2014
- TA02031411, „Zvýšení využití parkovací kapacity na dálnicích za pomoci predikčních modelů“
- Partneři:
 - FD ČVUT
 - ÚI AV ČR, v.v.i.
 - Kapsch TrafficCom Construction & Realization spol. s r.o.
 - Innoxive, s.r.o.

Cíle projektu

- Rychlý a relativně levný predikční model parkovací kapacity pro nákladní automobily podél dálnic
- Online implementace, telematické řešení přenosu dat k uživatelům/řidičům
- Využití stávajících (byť nepřímých) dat z mýtných bran vs. budování nákladné nové infrastruktury

Dostupná data

- Individuální data o průjezdech jednotlivých vozidel (OBU) z mýtného systému (Kapsch)
- Časy průjezdu jednotlivými mýtnými branami (1s -> 1min -> 5min rozlišení)
- Info o individuálních vozidlech (státní příslušnost SPZ, tonáž, Bus/Truck, apod.)

- Info o celkovém dopravním proudu (ŘSD, intenzita, hustota na vybraných branách)
- Dodatečná měření na vybraných parkovištích (počet vjíždějících/vyjíždějících vozidel, kalibrace a kontrola modelů)

Jak odhadnout intenzitu parkování z mýtných dat?

- Přímá měření počtu vozidel parkujících na daném parkovišti *nejsou* k dispozici
latentní proměnná odhadovaná z pozorovatelných dat
(o průjezdech mýtnými branami)
- Konstrukce proxy proměnné
jako aproximace počtu parkujících
časy průjezdu úseky -> průměrná rychlost
44 úseků na D5 (na 18 je parkoviště)
délka úseku 1.2-14.1 km (průměrně 6.7 km)
- Klasifikace parkujících/neparkujících
založena hlavně na prahu pro rychlost (event. i dalších proměnných)

Značení, I

- Máme individuální (krátké) řady časů průjezdů jednotlivými branami

$$T_{1j_1}, T_{1j_2}, T_{1j_3}, \dots, T_{1j_{n_1}}, \dots, T_{mj_1}, T_{mj_2}, T_{mj_3}, \dots, T_{mj_{n_m}}$$

i-té vozidlo, j-tá brána

individuální řady mají obecně různou délku a jsou různě situované

- Parkování v čase t ,
v úseku mezi $(j-1)$ -ní a j -tou branou

$$(1) \quad T_{i,j-1} + \Delta_{ij}^A < t \quad \mathbf{a} \quad (2) \quad t < T_{ij} - \Delta_{ij}^B$$

Značení, II

- (Velmi) jednoduchý přístup

$$\Delta_{ijt}^A := \Delta_j^A = K \cdot \frac{x_{P,j} - x_{I,j}}{\bar{v}_{.j}}$$

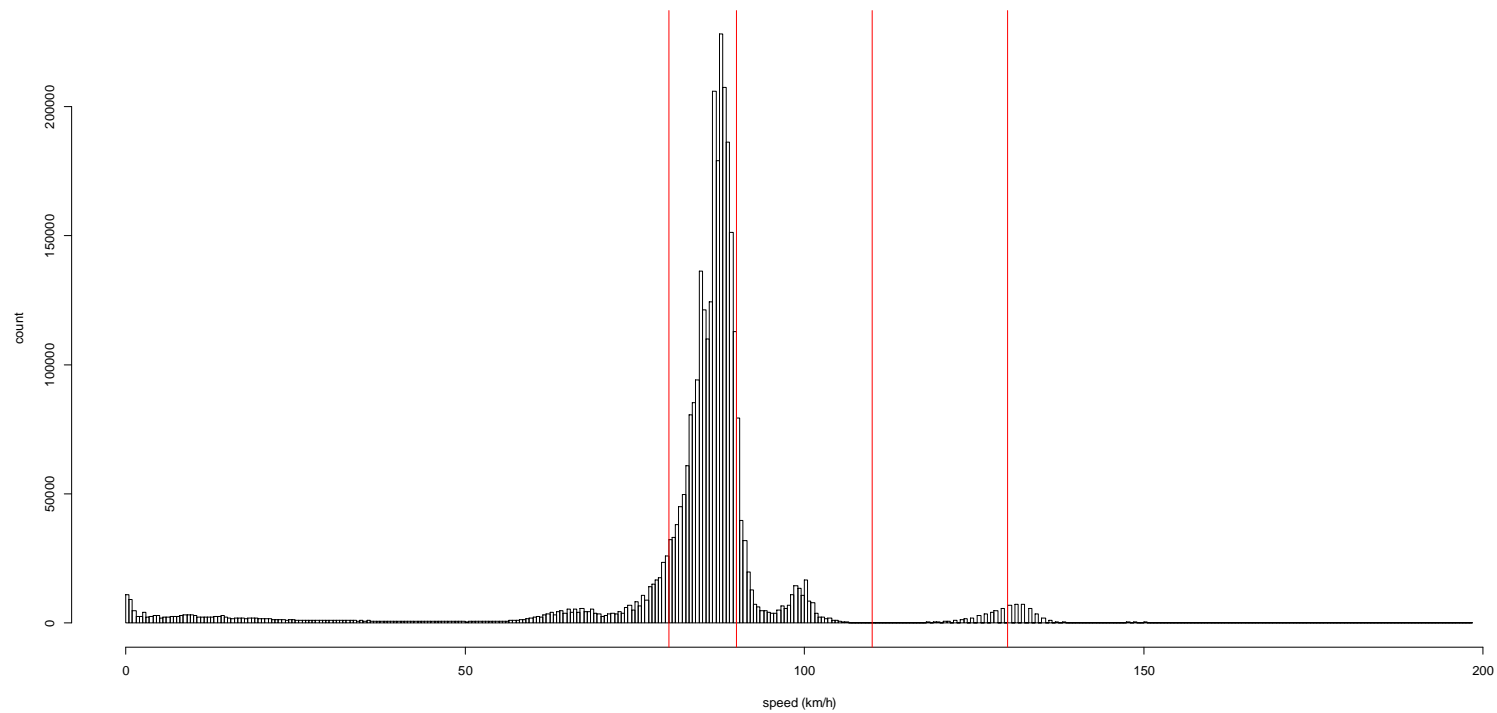
$$\Delta_{ijt}^B := \Delta_j^B = K \cdot \frac{x_{O,j} - x_{P,j}}{\bar{v}_{.j}}$$

- Δ_{ijt}^u může/měla by záviset na lokálních vlastnostech dopravního proudu, event. vozidla
- Proxy pro počet parkujících na parkovišti v úseku j (mezi branami $j, j+1$)

$$N_j(t) = \sum_i I(\text{vozidlo } i \text{ splnuje (1)}) \cdot I(\text{vozidlo } i \text{ splnuje (2)})$$

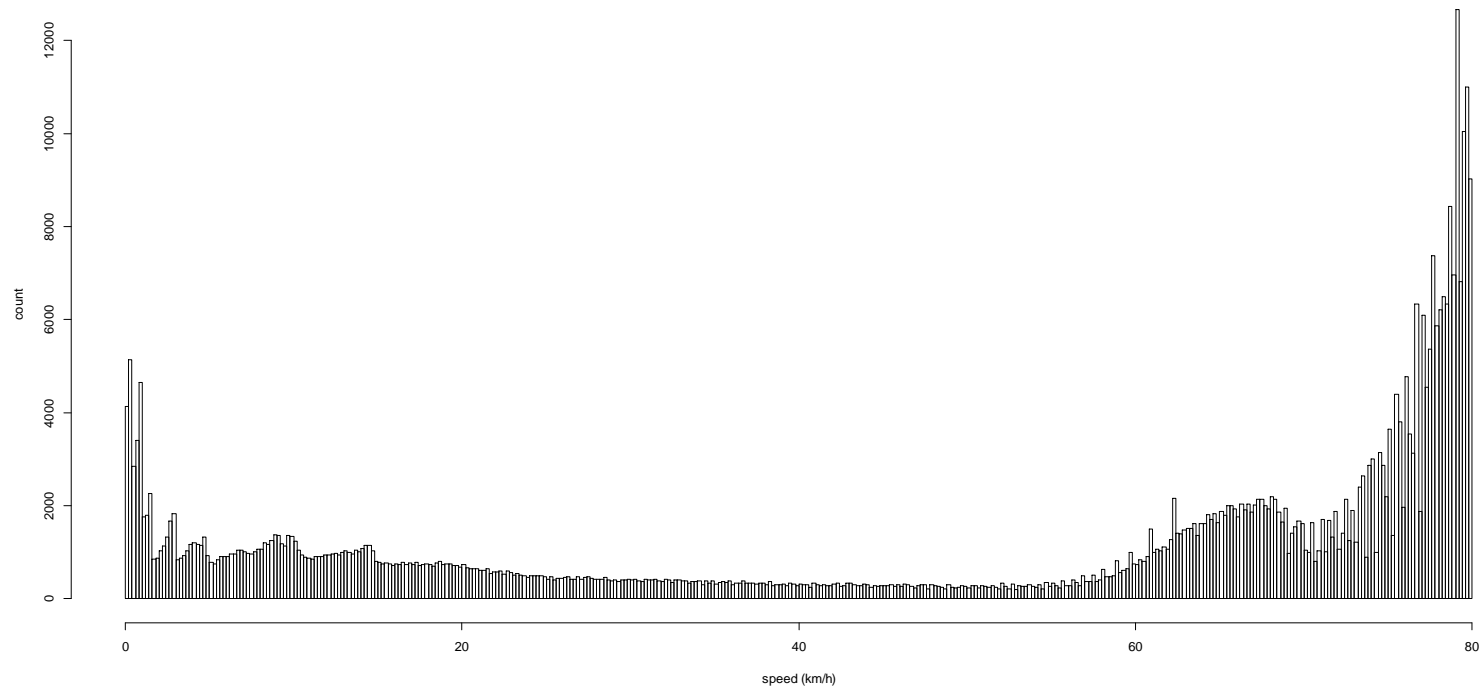
Konstrukce proxy pro příznak parkování

(pro binární A/N klasifikaci jednotlivých vozidel v daném úseku)



Relevantní jsou jen nízké rychlosti

(histogram pro rychlosti pod 80 km/hod)



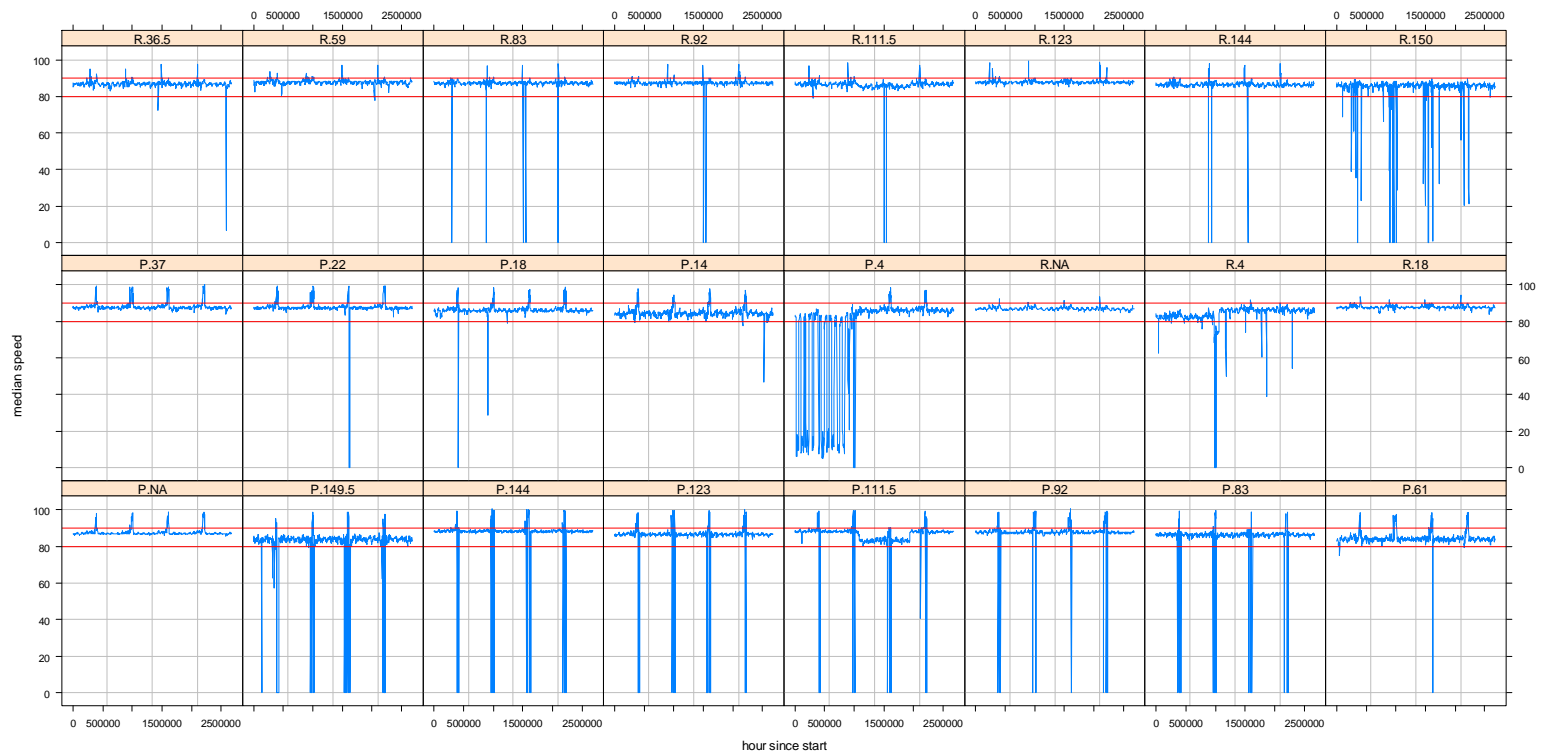
Klasifikační chyby

Parkování		Klasifikace pro proxy proměnnou z průjezdových dat	
		Ano	Ne
Skutečnost	Ano	Korektní	FN
	Ne	FP	Korektní

Konstrukce proxy

- Návrh třídy pravidel s využitím dat o skutečném stavu parkoviště
„zlatý standard“ a manuální sčítání
- Optimalizace zvoleného kritéria
FN, FP
total misclassification
weighted total misclassification
ROC (AUC)
- Klasifikační pravidla pevná/globální vs. lokální

Časová a prostorová nehomogenita (mediánová trajektorie)



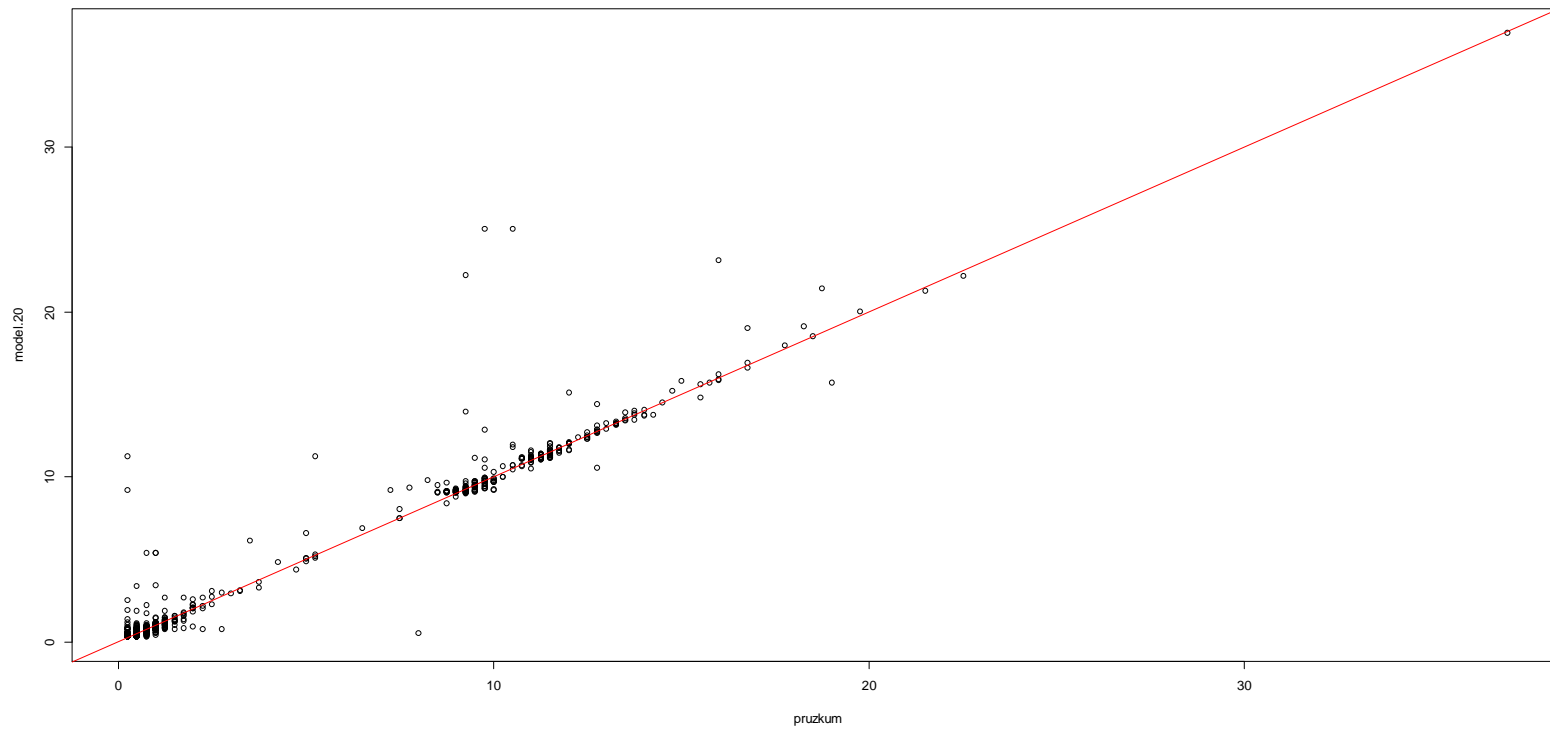
Kritéria pro klasifikaci parkování, proxy2

- čas parkování delší než 20 min (při rychlosti 85 km/hod)
- čas na parkování delší než medián+20min
- (zdánlivá) rychlost (z času průjezdu úsekem s parkovištěm) je oproti minulému úseku menší
- čas parkování kratší než 3 dny

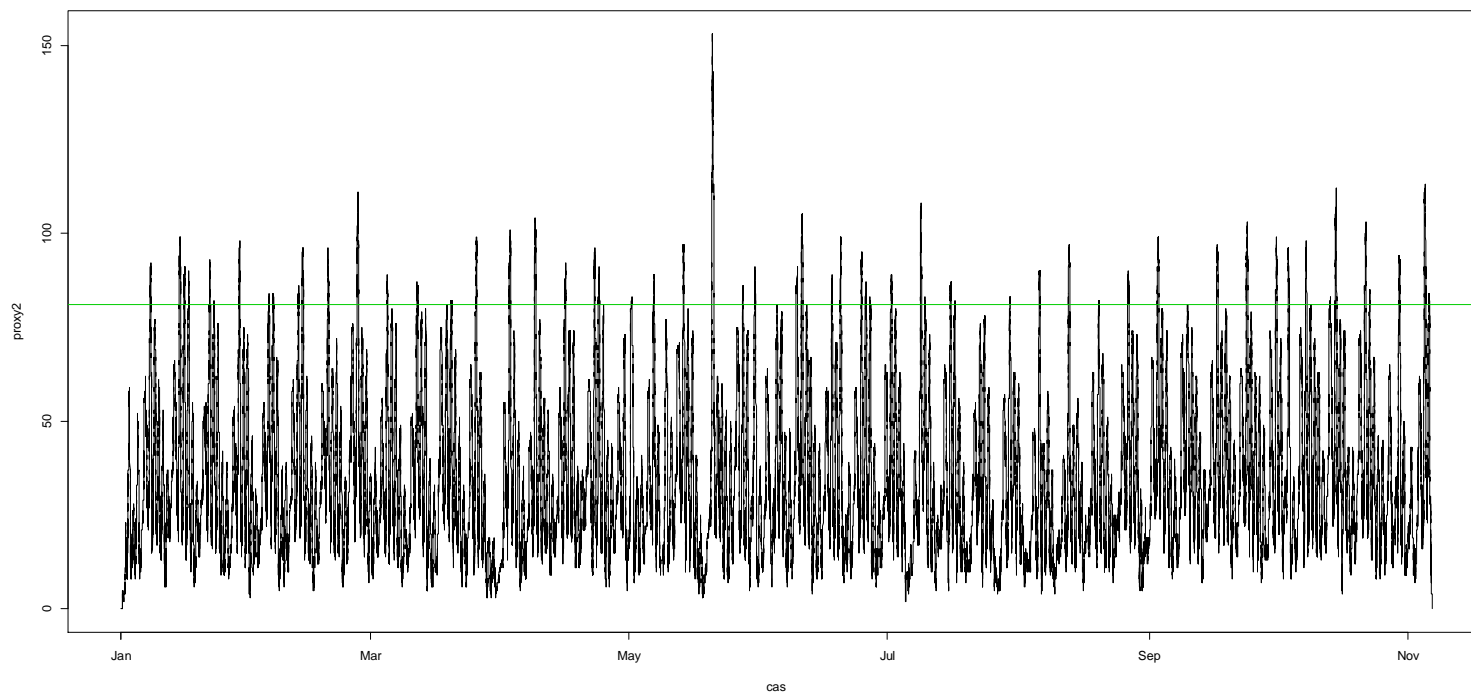
Dílčí orientační test pro FP

- Pravidlo se aplikuje i na úseky v nichž žádné parkoviště není
- Ta vozidla která pravidlo klasifikuje jako pozitivní jsou FP
- Problémy
 - úseky bez parkoviště nemusejí mít všechny vlastnosti stejné jako úseky s parkovištěm
 - FP takto můžeme „optimalizovat“, mnohem důležitější FN ale nikoli

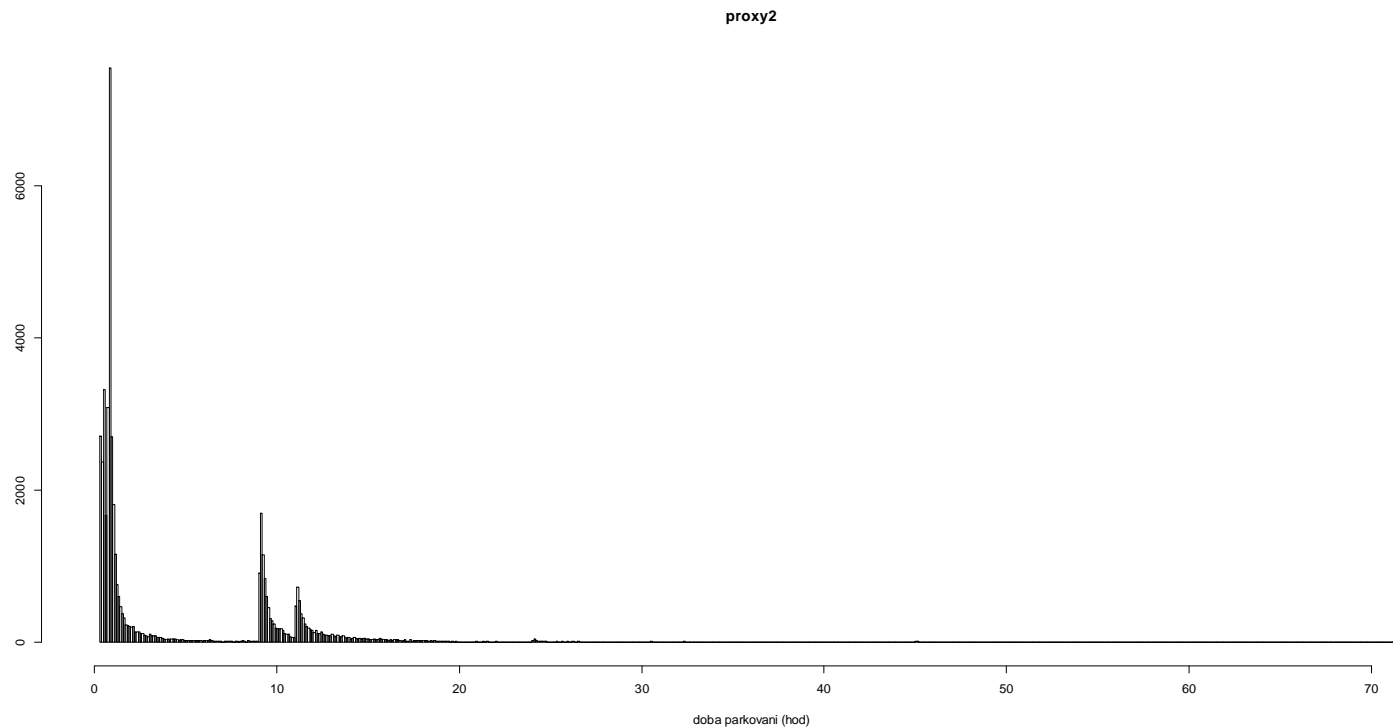
Vztah mezi manuálním sčítáním a proxy



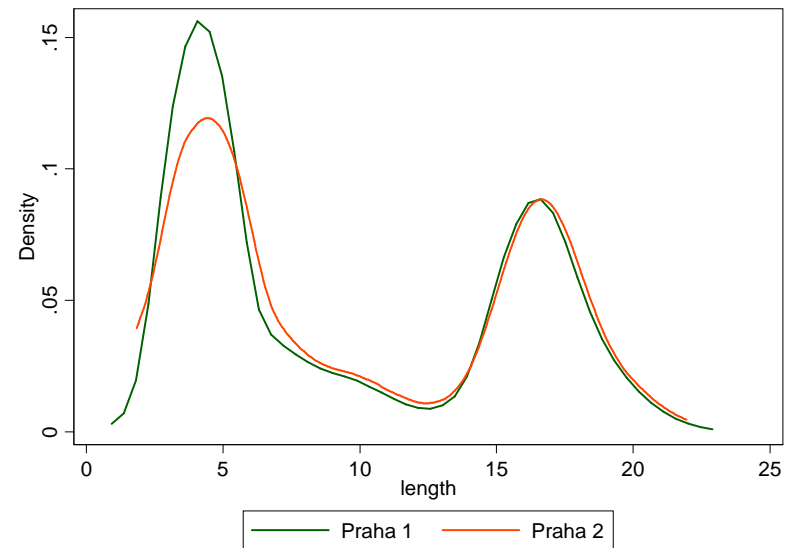
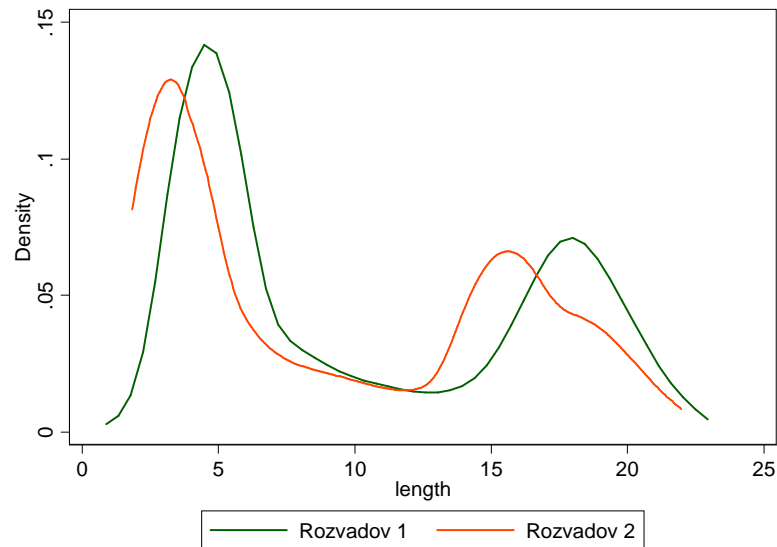
Počty parkujících odvozené z proxy proměnné a nominální kapacita parkoviště



Další odvozené proměnné (např. pro dlouhodobé plánování kapacity parkovišť): délka parkování dle proxy2



Pokus o objektivizaci, automatická čidla ...

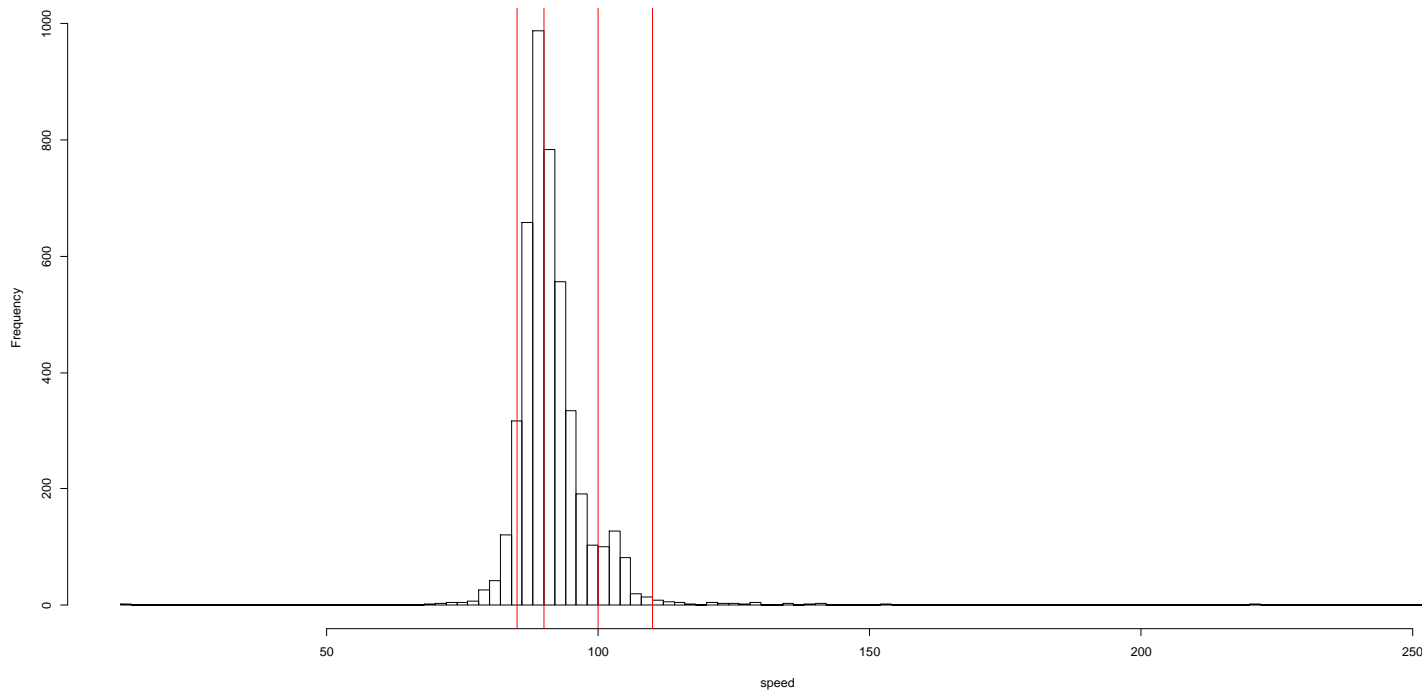


Alternativní (poněkud „strukturovanější“) přístup

- Z časů individuálních průjezdů lze snadno získat histogram (úsekových, průměrných) rychlostí
- Ten poskytuje dobrou představu o rozdělení *zdánlivých* rychlostí jízdy
- Směs rozdělení rychlostí vozidel
 - i) projíždějících a
 - ii) projíždějících + parkujících
- Snaha získat ii) - ve stylu „dekonvoluce“

Rozdělení rychlostí skutečně jedoucích vozidel

(dlouhodobá měření z jiných dálničních čidel okamžité rychlosti vozidel klasifikovaných jako nákladní, nákladní s přívěsem, bus)



Střední počet vozidel parkujících v čase t

- „vstupní“ kritéria pro parkování i -tého vozidla v j -tém úseku (někdy) v intervalu $(T_{i,j-1}, T_{ij})$ obdobná jako pro proxy2

(mj. podstatné pro zkrácení doby výpočtů)

- pro $\tau_{ijS,k} = T_{i,j-1} + \frac{x_{P,j} - x_{I,j}}{v_k}$, $\tau_{ijE,k} = T_{i,j} - \frac{x_{O,j} - x_{P,j}}{v_k}$

podmínka na rychlost v_k : $\tau_{ijE,k} - \tau_{ijS,k} > 20 \text{ min}$ $t \in (\tau_{ijE,k}, \tau_{ijS,k})$

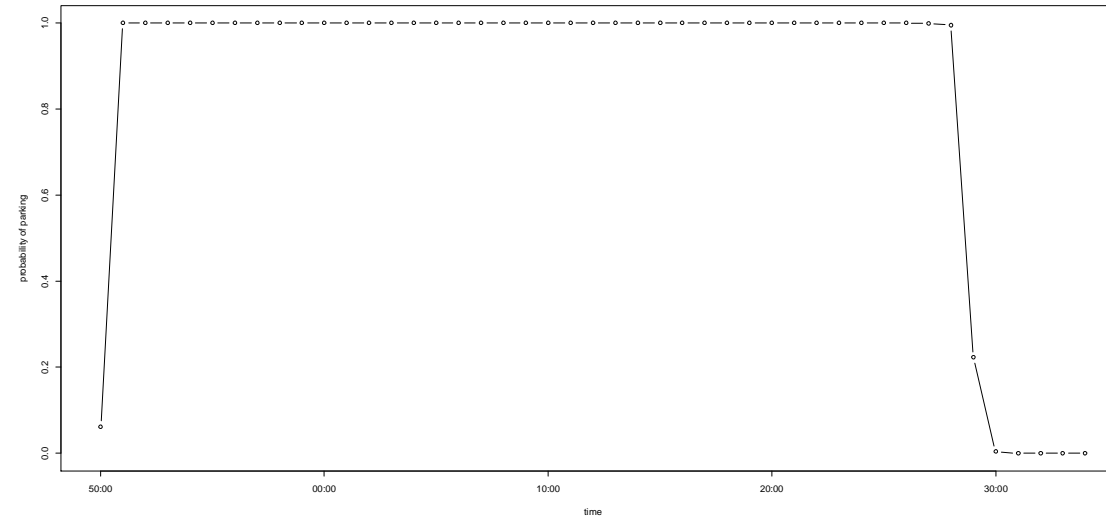
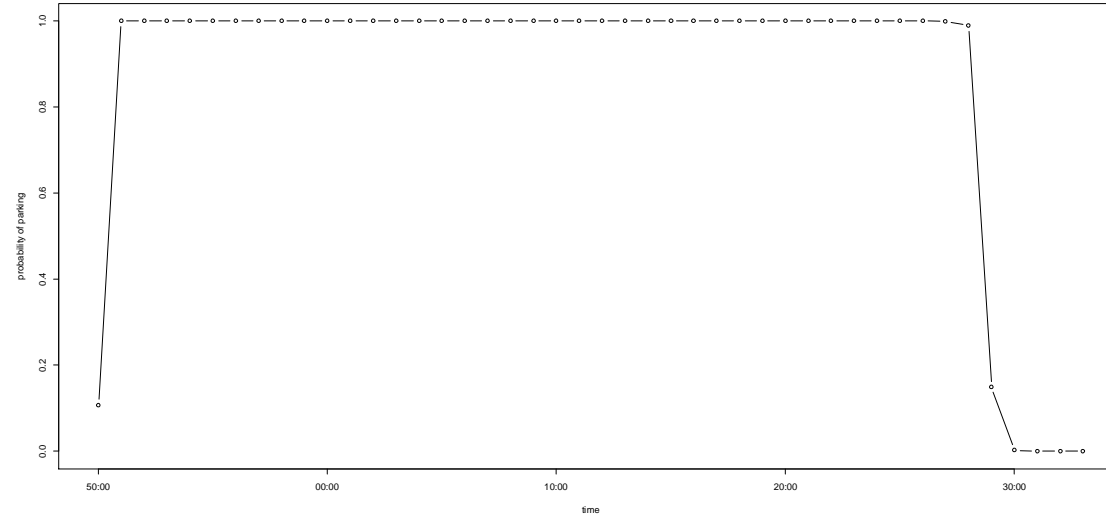
$$P[i, j, k, t] = P((\text{vstupni podm.}) \cap (\text{podm. na rychlost } v_k)) = P(v_k)$$

- střední počet vozidel parkujících v čase t

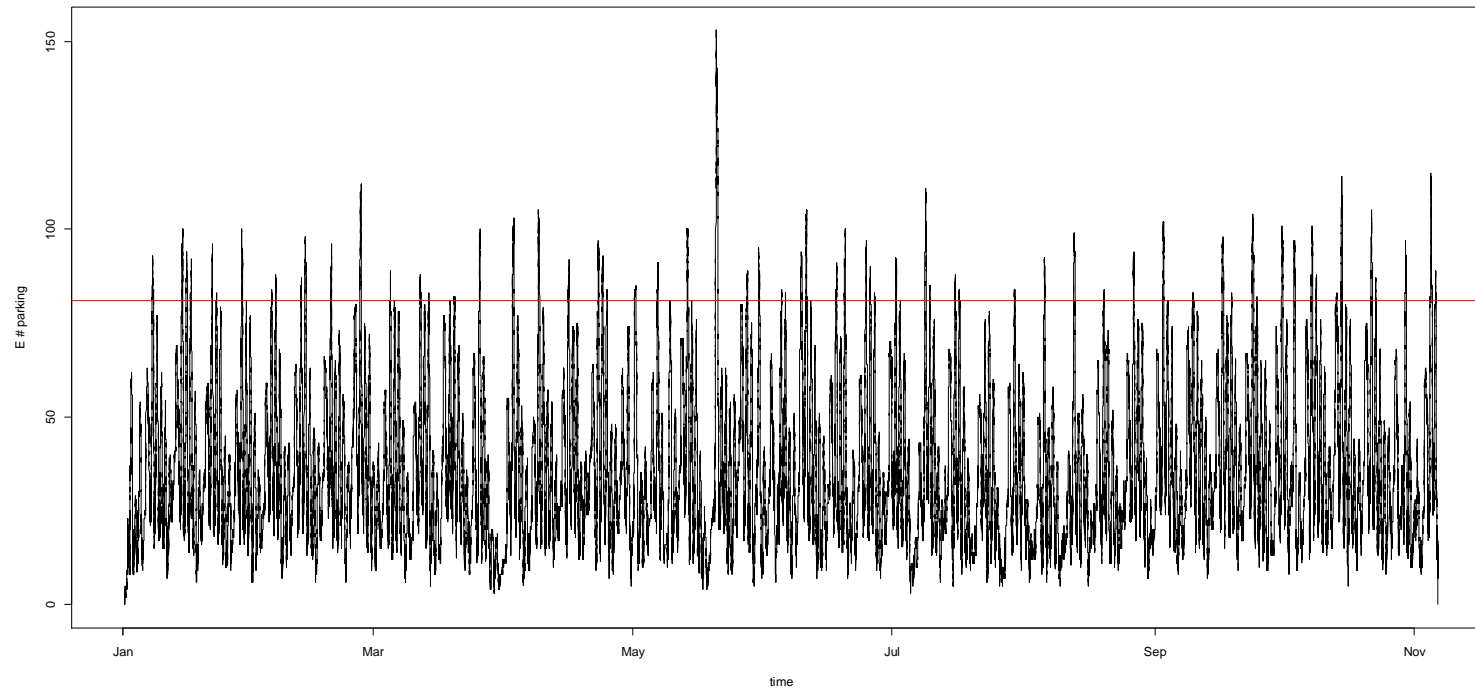
$$E[N_{j,t}] = E\left[\sum_i I(i - \text{te voz. parkuje v } j - \text{tem useku v case } t)\right] =$$

$$\sum_i P(i - \text{te voz. parkuje v } j - \text{tem useku v case } t) = \sum_i \sum_k P[i, j, k, t]$$

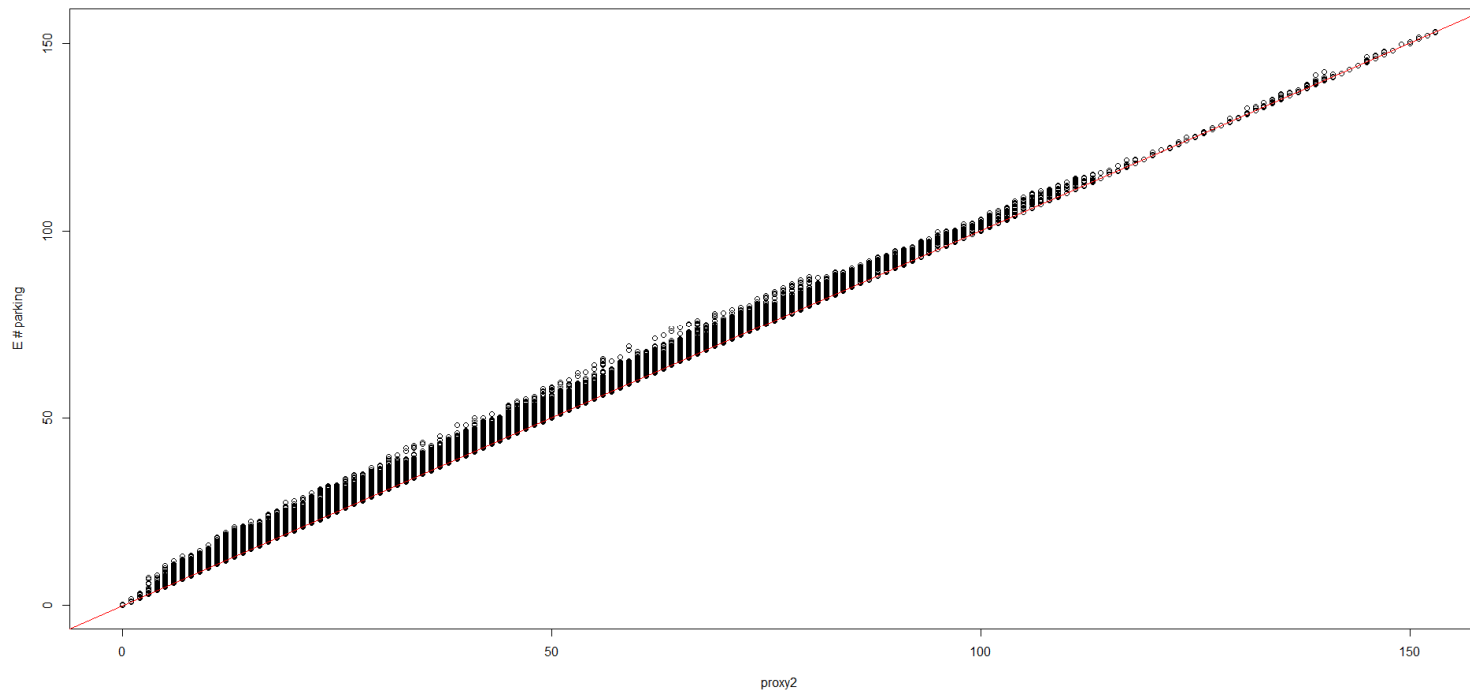
Příklad pravděpodobností parkování ($\sum_k P[i, j, k, t]$) pro minutách pro individuální průjezd

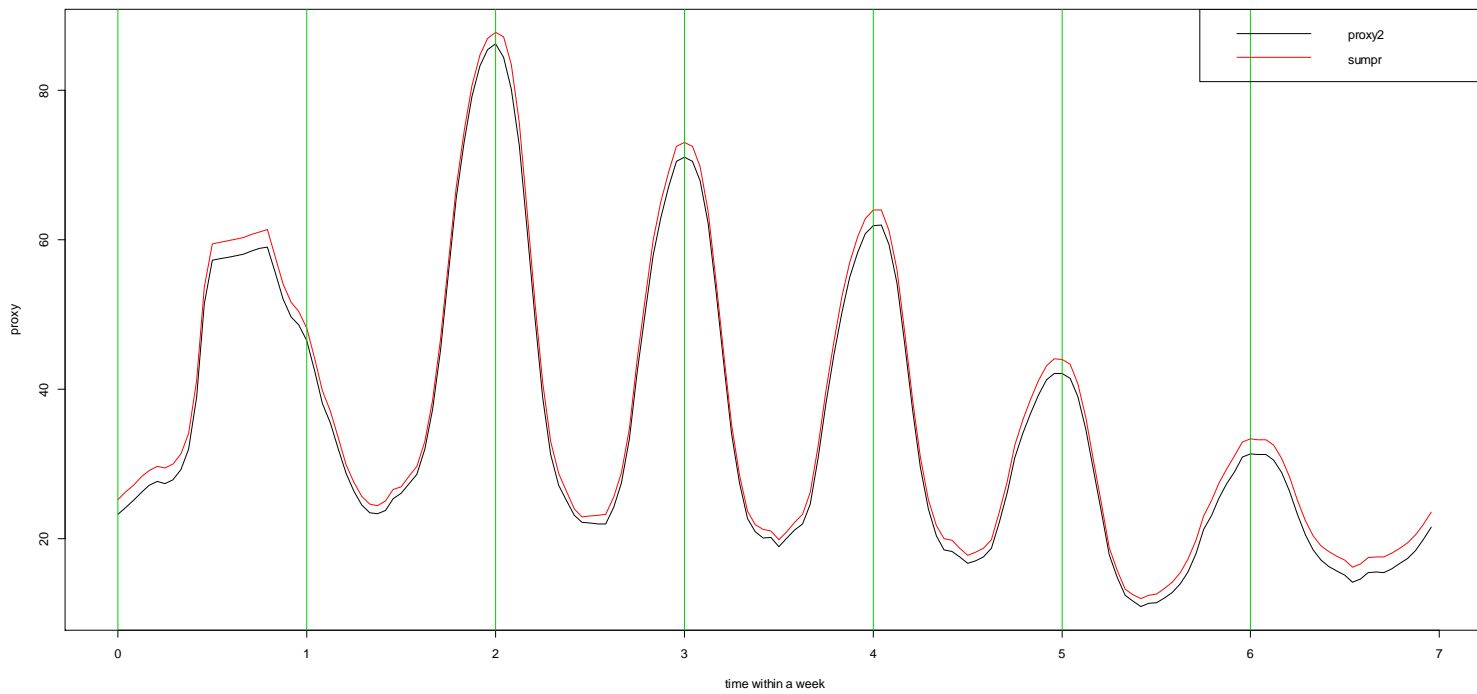


Střední hodnota počtu parkujících



Srovnání s proxy2





Dynamický model (Poissonovský model Markovského typu)

Jako kovariáty se používá nejen externích veličin (např. čas) ale i vlastní historie procesu

Markovský (nehomogenní) řetězec

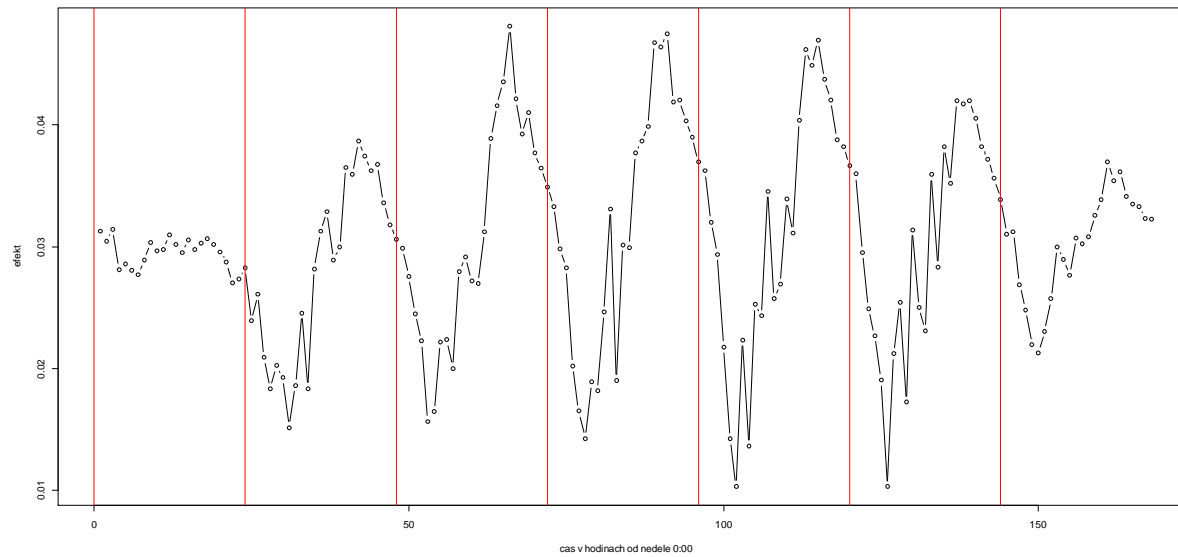
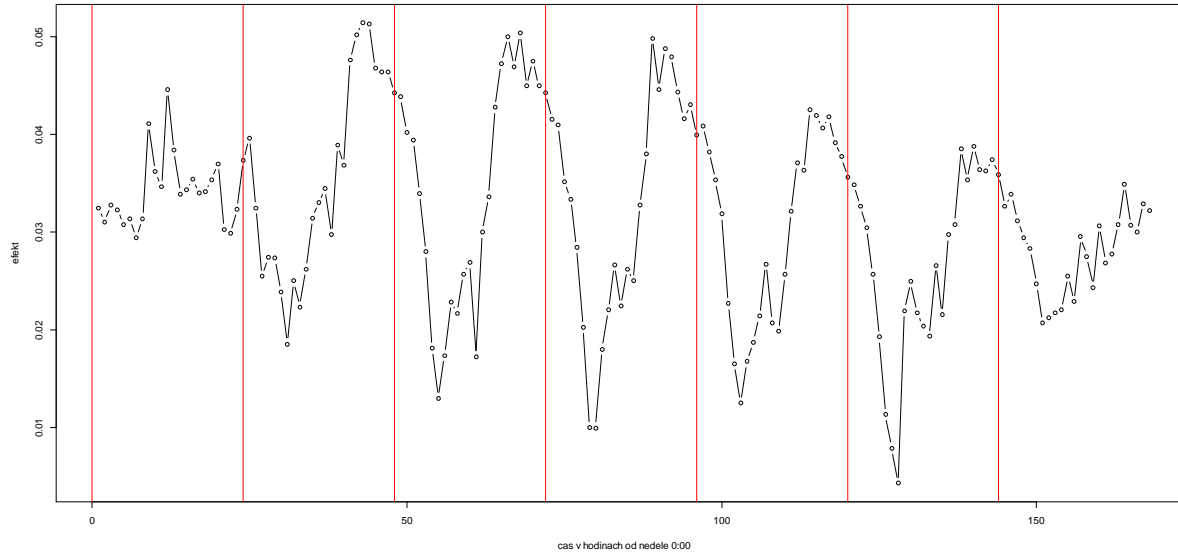
- Jednoduchá specifikace
- Jednoduchá interpretace
 - analog AR modelů na povrchu
(více komplexity uvnitř)
 - MA analog také možný, ale výrazně komplikovanější – Coxův process
- Relativně jednoduchý odhad (požívá výhod GLM třídy)
- Jednoduchá implementace v online prediktivním nástroji

Model přechodových pravděpodobností

Poissonovský, Markovovský model 1. řádu
v detailu máme ale řadu možností ...

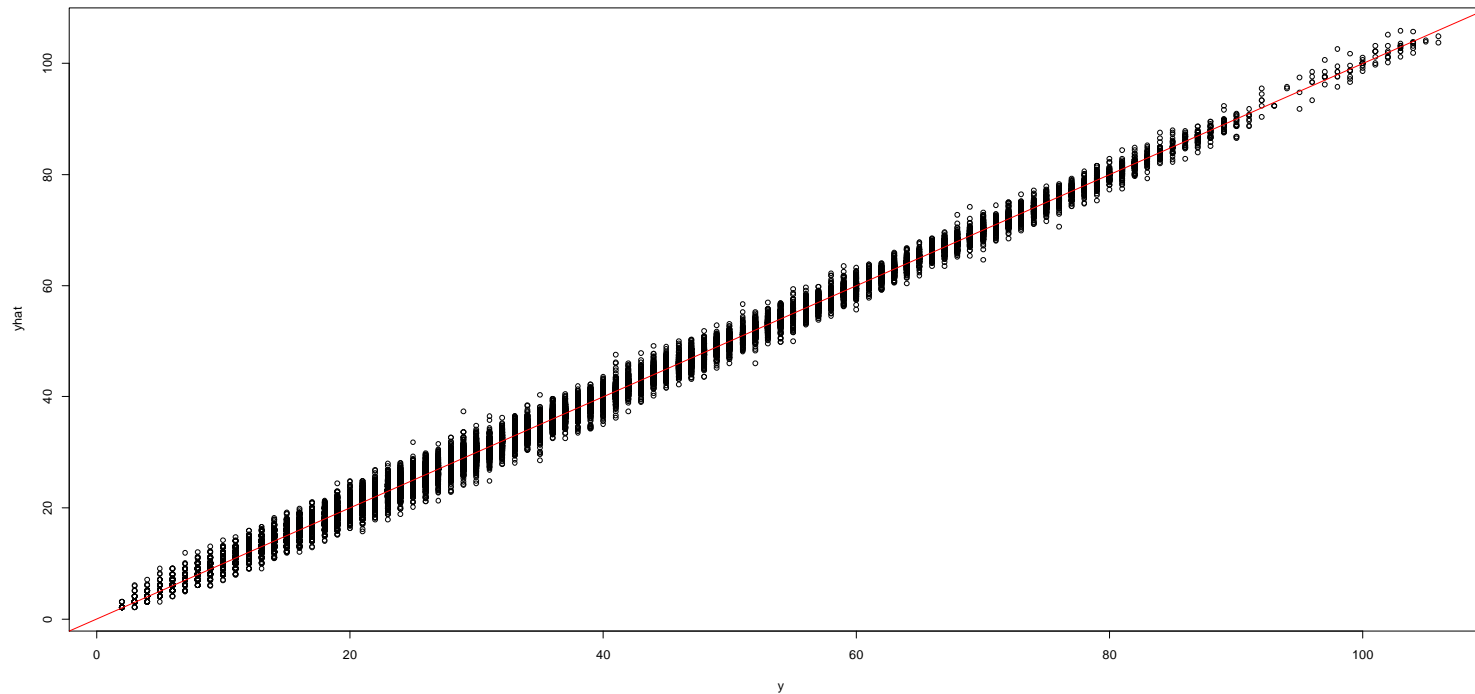
- $N_{j,t} | N_{j,t-1} \sim Poi(\mu_{jt})$
- **přímočaře, GLM:** $\log(\mu_{jt}) = f_{jt} + \beta \cdot \log(\tilde{N}_{j,t-1})$
pro $\tilde{N}_{j,t-1} = \max(N_{j,t-1}, c)$, $0 < c < 1$
a týdenní periodicitu $\exp(f_{jt})$ danou např. jako interakce dne v týdnu a hodiny (či spojitě po minutách uvnitř týdne)
- **Zeger, Quaqish (1988):** $\log(\mu_{jt}) = f_{jt} + \beta \cdot [\log(\tilde{N}_{j,t-1}) - f_{j,t-1}]$
(nelinearita – vnější iterace)
- **aditivní složky intenzity (identity link):**
$$\mu_{jt} = f_{jt} + \beta \cdot N_{j,t-1}$$
$$\mu_{jt} = f_{jt} + \beta \cdot [N_{j,t-1} - f_{j,t-1}]$$

Odhadnutý model, sezonalita (f_{jt})



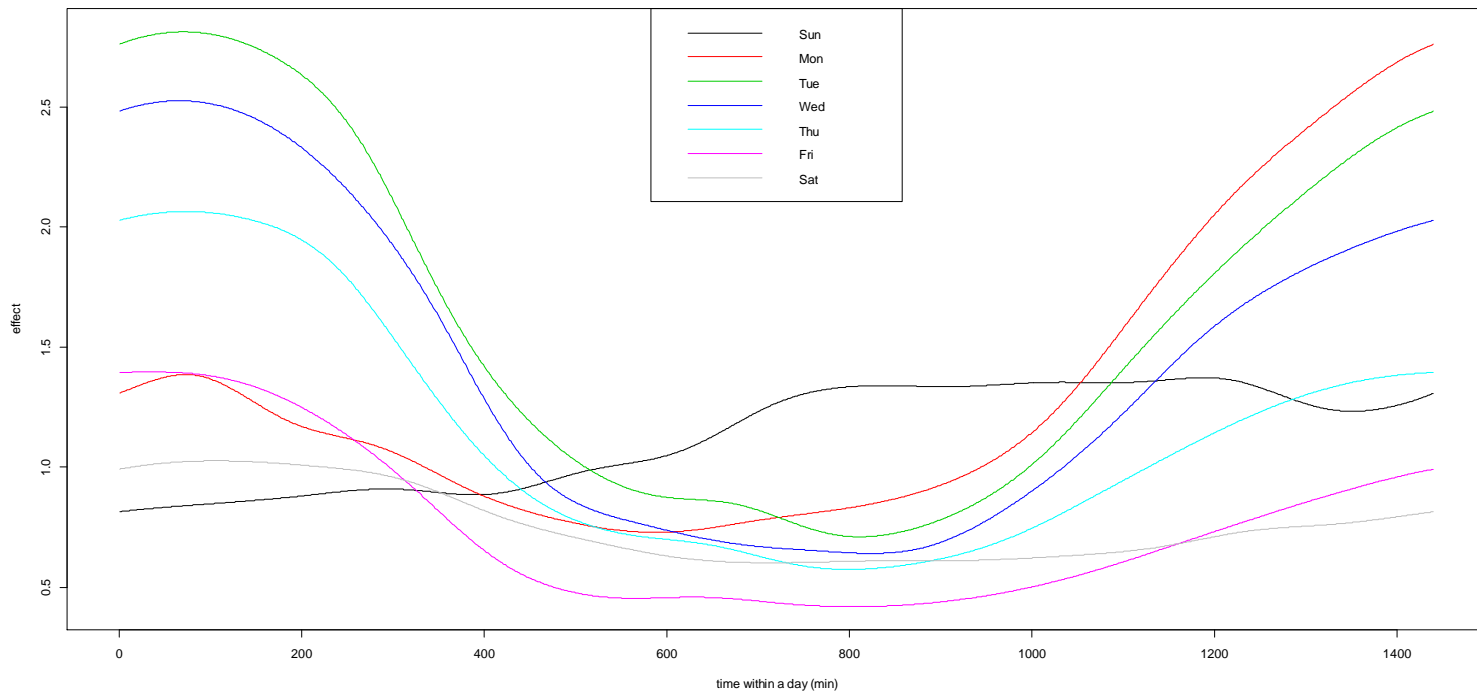
ZQ model fit (predikce o 1 krok, $\Delta = 5 \text{ min}$)

(cyclický kubický regresní spline s penalizací)



ZQ model, $\exp(f_{jt})$

(cyclický kubický regresní spline)



Některé z možností odhadu sezónní komponenty

- jako dvojná interakce dne v týdnu a hodiny resp. reparametrizace 2FI do jediného faktoru

$$f_{jt} = \sum_l \beta_l I(t \text{ odpovídá } l - \text{ tému } \Delta \text{ intervalu v týdnu})$$

- dtto s „regularizací“, tj. penalizací například 1. nebo (lépe) 2. diferencí mezi sousedními koeficienty (a periodicitou)

$$L + \lambda \cdot (H \cdot \underline{\beta})^T \cdot (H \cdot \underline{\beta})$$
$$H = \begin{pmatrix} 1 & 0 & \dots & -1 \\ -1 & 1 & & 0 \\ & & \vdots & \\ 0 & -1 & & 1 \end{pmatrix} \quad \begin{pmatrix} -2 & 1 & 0 & 1 \\ 1 & -2 & 1 & 0 \\ & & \vdots & \\ 1 & 0 & & 1-2 \end{pmatrix}$$

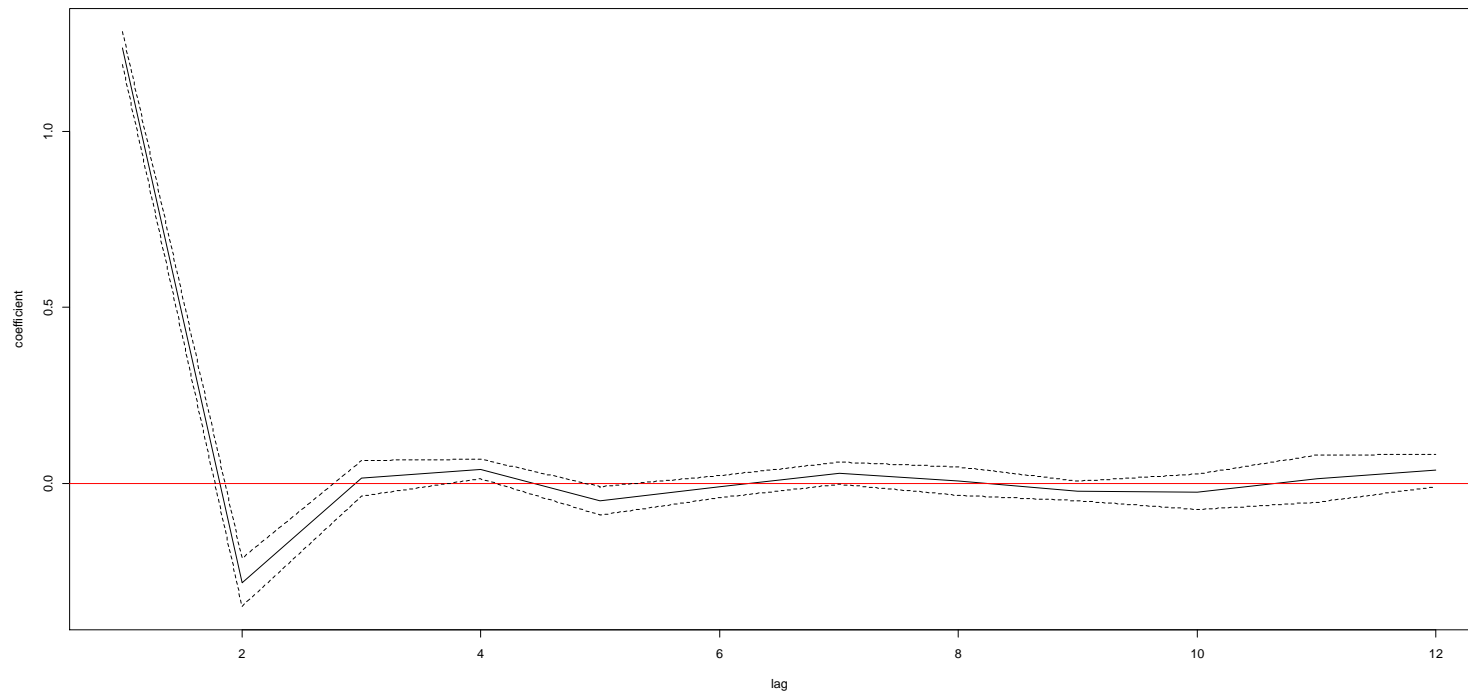
- f_{jt} pro dané j leží křivce dané splinem s periodickou podmínkou

Použití delší historie?

- Získat na efektivitě predikcí pokud použijeme více lagů simultánně? Např. lag1, lag2.
- Technicky náročné pro větší počet lagů najednou (blízké lagy jsou velmi korelované).
- Jednou možností je použít ideu of Almonova modelu (populárního v Ekonometrii) a mírně jej zobecnit za použití B-splínové (namísto polynomiální) báze pro transformaci pořadových čísel lagů

Almon model (12, 9)

(nesezónní část modelu dává nelineární, časově-invariantní filtr)



Testování (prakticky motivovaných) hypotéz

- Může být užitečné využít další info z předchozích parkovišť (proti směru jízdy)?

- Lineární prediktor expandován na:

$$\log(\mu_{jt}) = f_{jt} + \beta \cdot \log(\tilde{N}_{j,t-1}) + \gamma \cdot \log(\tilde{N}_{j',t-1-lag})$$

nebo na:

$$\log(\mu_{jt}) = f_{jt} + \beta \cdot \log(\tilde{N}_{j,t-1}) + \gamma \cdot [\log(\tilde{N}_{j',t-1-lag}) - \log(\tilde{N}_{j',t-2-lag})]$$

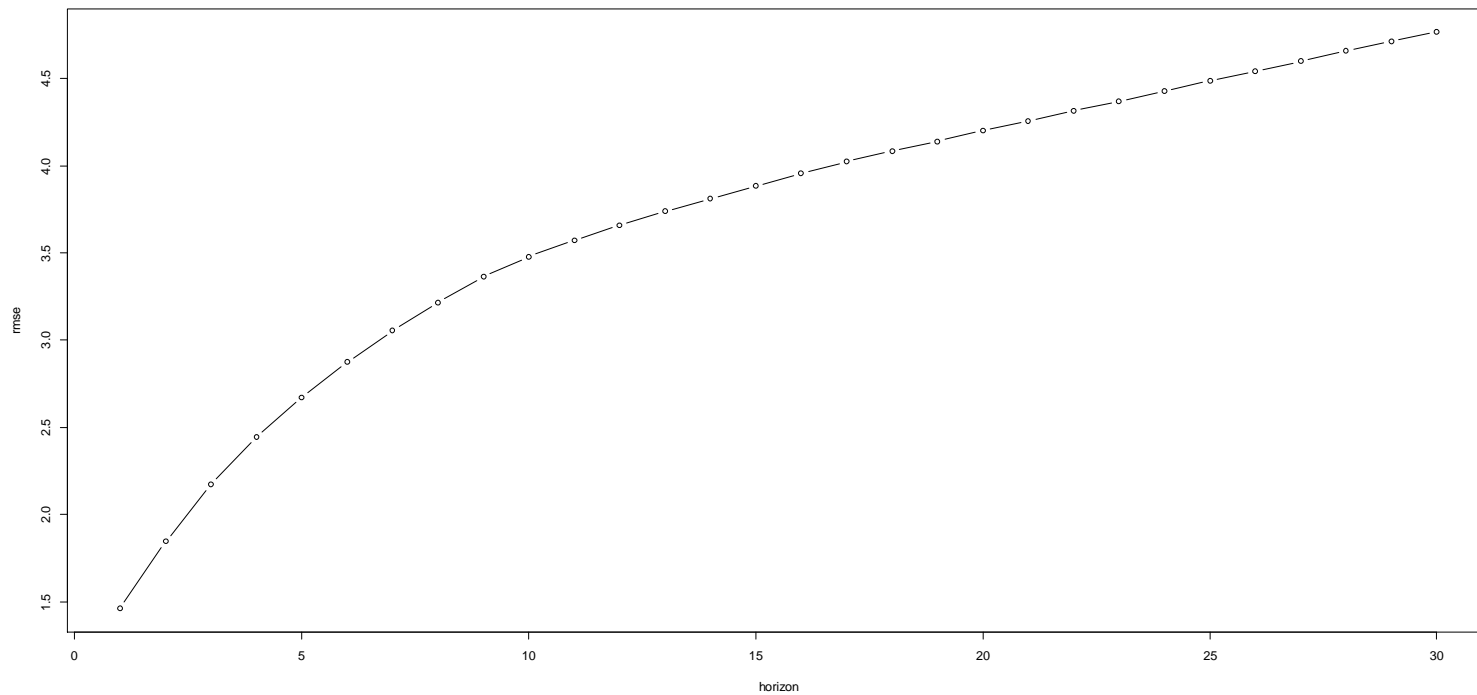
- Přídavek členu odpovídajícímu stavu předchozího parkoviště (zpožděnému o lag daný typickou rychlostí 85km/h a vzdáleností obou parkovišť) lze snadno testovat (LRT)
p-hodnota=0.1644 resp. 0.9737

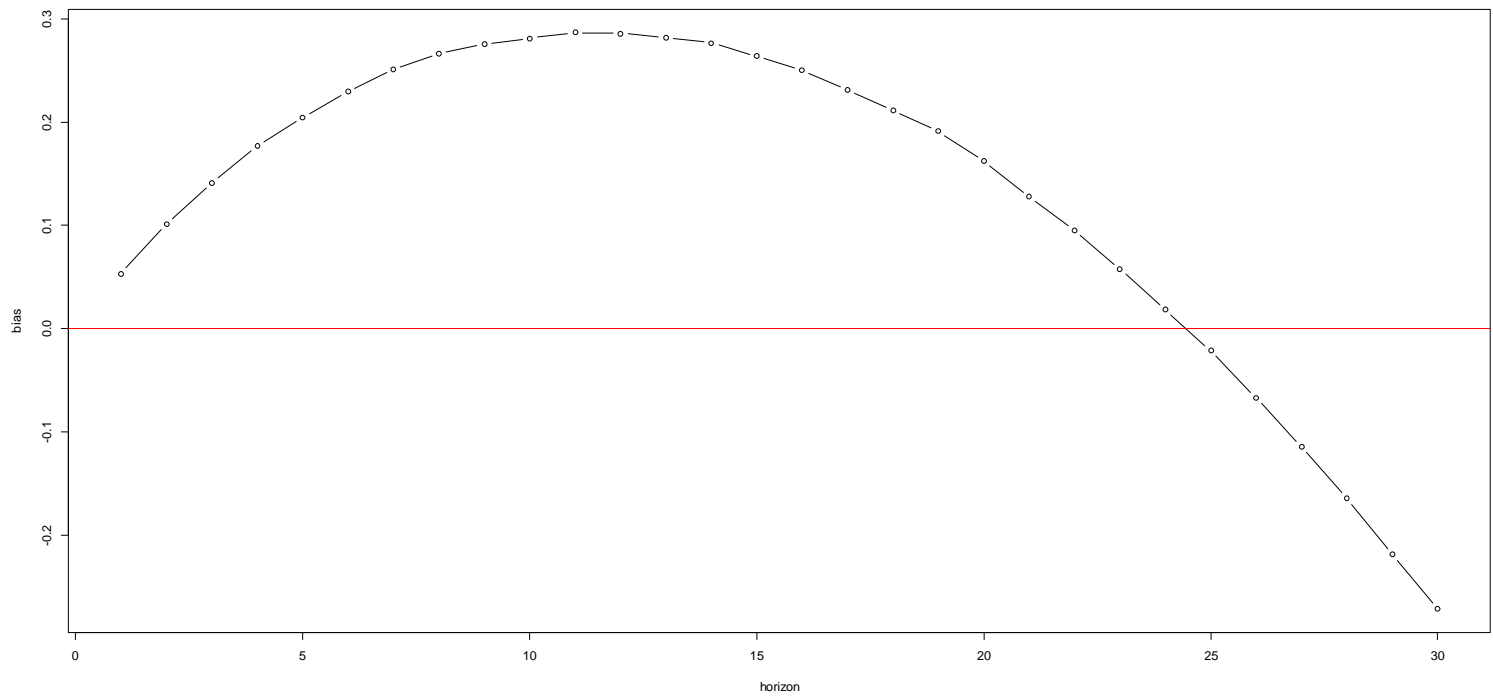
Predikce

- Násobení přechodových matic
- Stav a (velmi jednoduchá) simulace z přechodových Poi rozdělení (s intenzitou měnící se dle historie a času predikce jež určuje hodnotu sezónního příspěvku)
- “forecasting” (AI) vs. statistický přístup
odhad modelů pro každý horizont zvlášť vs. propagace dynamického modelu specifikovaného přech. pravděp. o 1 krok

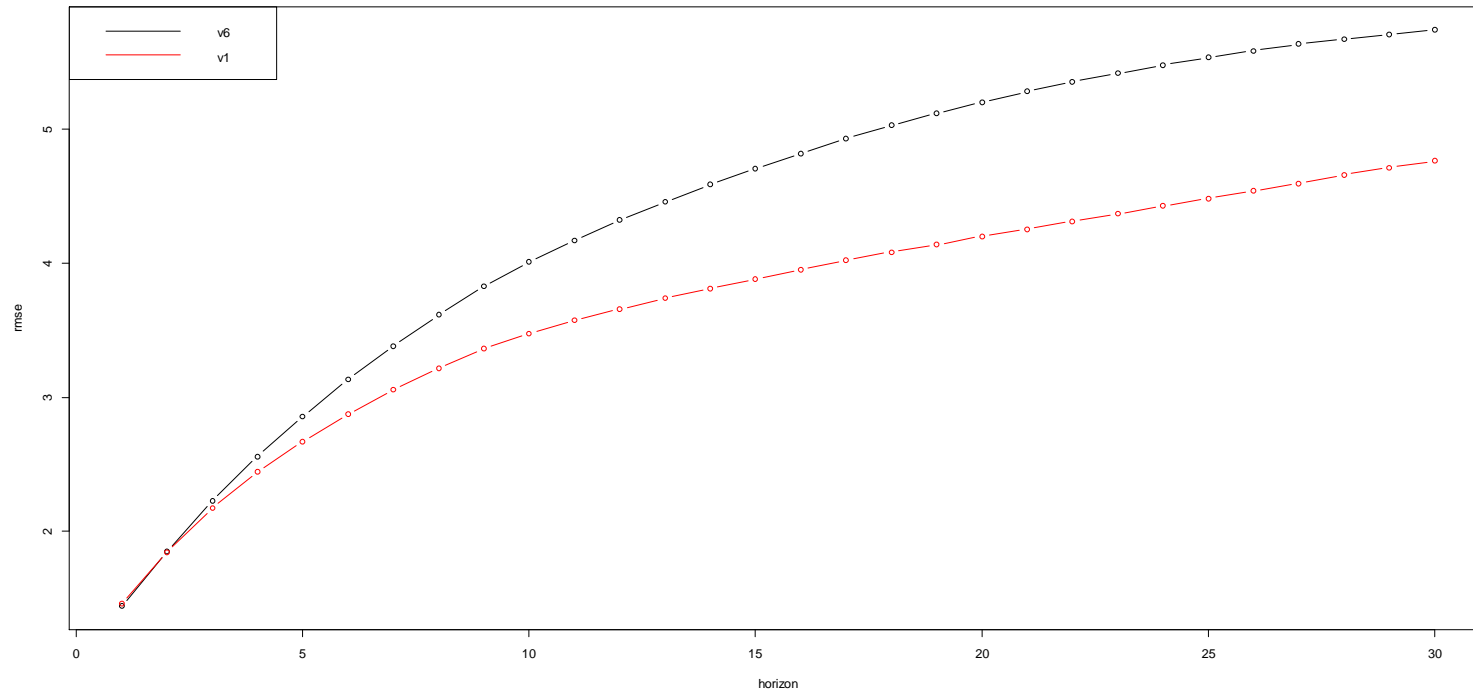
Kvalita predikce počtu parkujících pro 5- minutové intervaly a různé horizonty

(Poi model založený na proxy2, max horizont 150 minut)

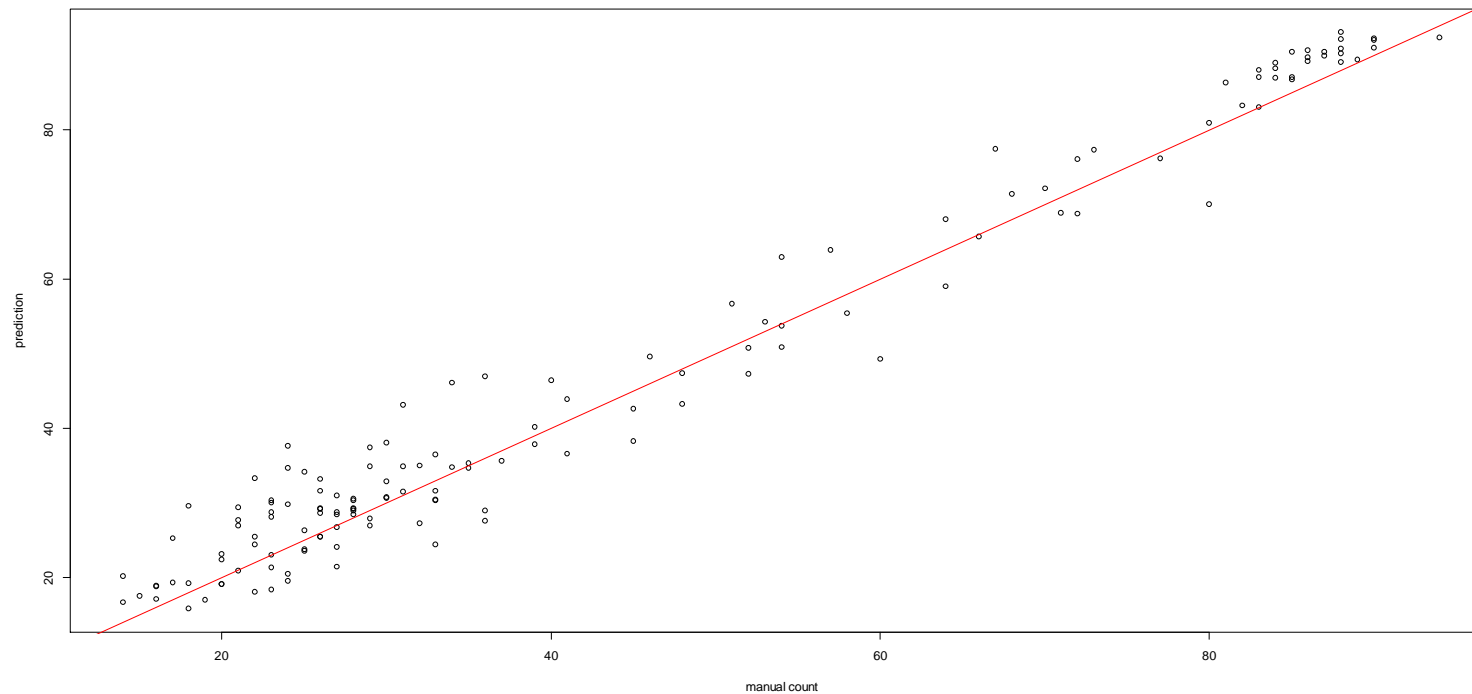




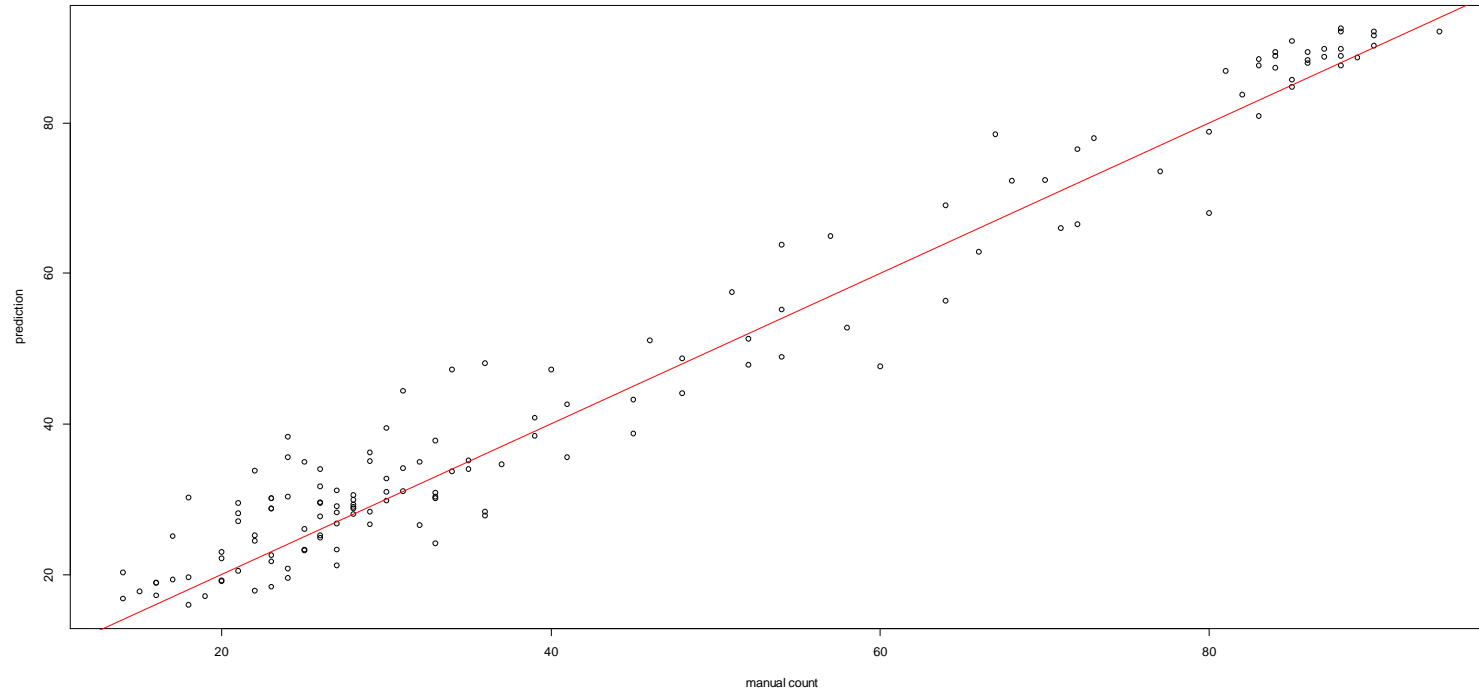
Srovnání modelů s lag1 a lag1,lag2



Predikce z modelu (na 15 min dopředu) vs. manuální sčítání



25 min



25 min

