esf evropský sociální fond v ČR

EVROPSKÁ UNIE

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání pro konkurenceschopnost

AMathNet síť pro transfer znalostí v aplikované matematice

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Statistický model pro predikci parkovací kapacity z nepřímých dat

Brabec,M.-Konár,O.-Kasanický,I.-Pelikán,E.
mbrabec@cs.cas.cz

Workshop Matematické modely a aplikace
Podlesí 5.-6.9.2013

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

  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 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 vjíždělících/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 parkující/neparkující

  založena hlavně na prahu pro rychlost (event. i dalších proměnných)

Značení, I

- Máme individuální (krátké) řady časů průjezdů jednotlivými branami

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

  i-té vozidlo, j-tá brána

  individuální řady mají obecně různou délku

- Parkování v úseku mezi j-tou a (j+1)-ní branou

  (1) $T_{ij} + \Delta_{ijt} < t$ **a** (2) $t < T_{i,j+1} - \Delta_{i,j+1,t}$

Značení, II

- (Velmi) jednoduchý přístup

  (1´) $T_{ij} < t$ a $\Delta_{i,j+1,t} := \Delta_{j+1} = K.\frac{x_{B,j+1} - x_{P,j}}{\bar{v}_{.,j}}$

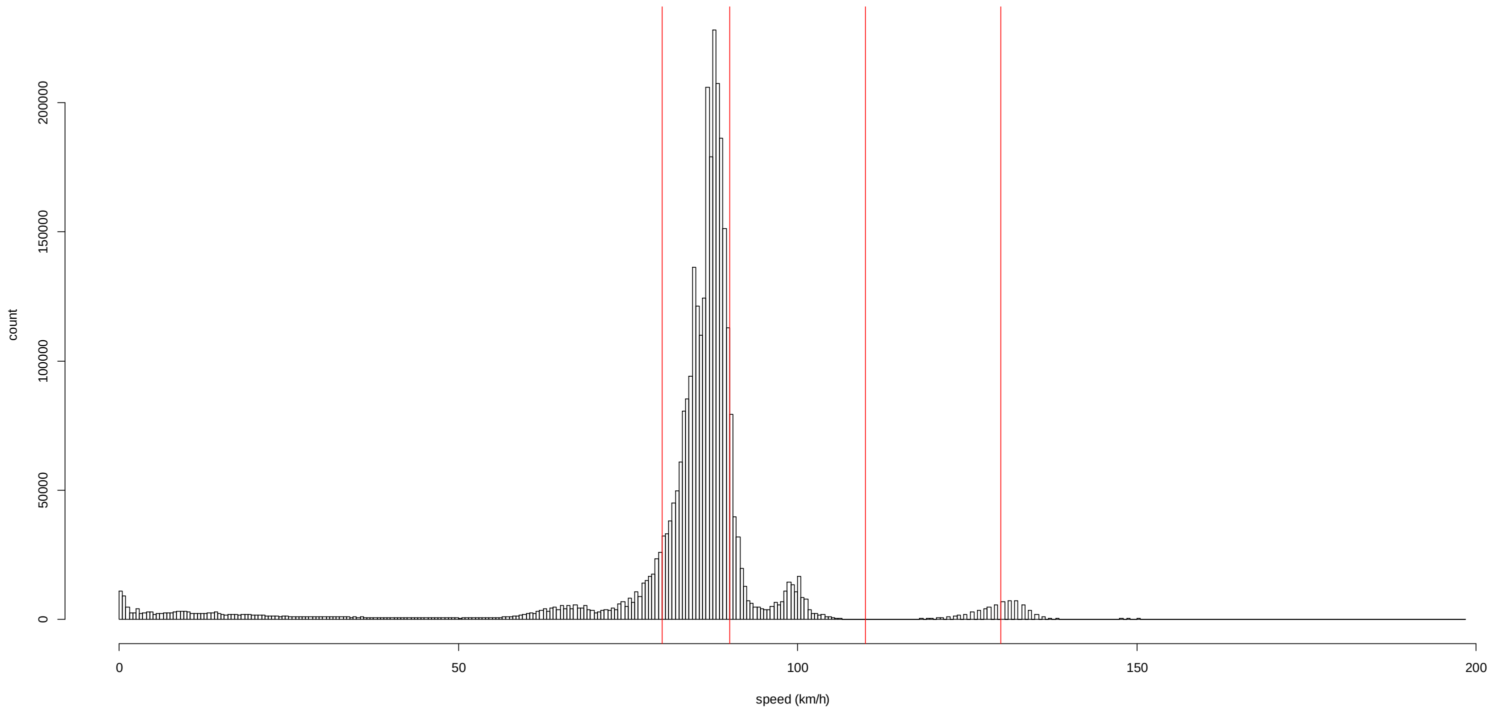
- $\Delta_{j+1,t}$ může/měla by záviset na lokálních vlastnostech dopravního proudu

- Proxy pro počet parkujících na parkovišti v úseku j (mezi branami j, j+1)

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

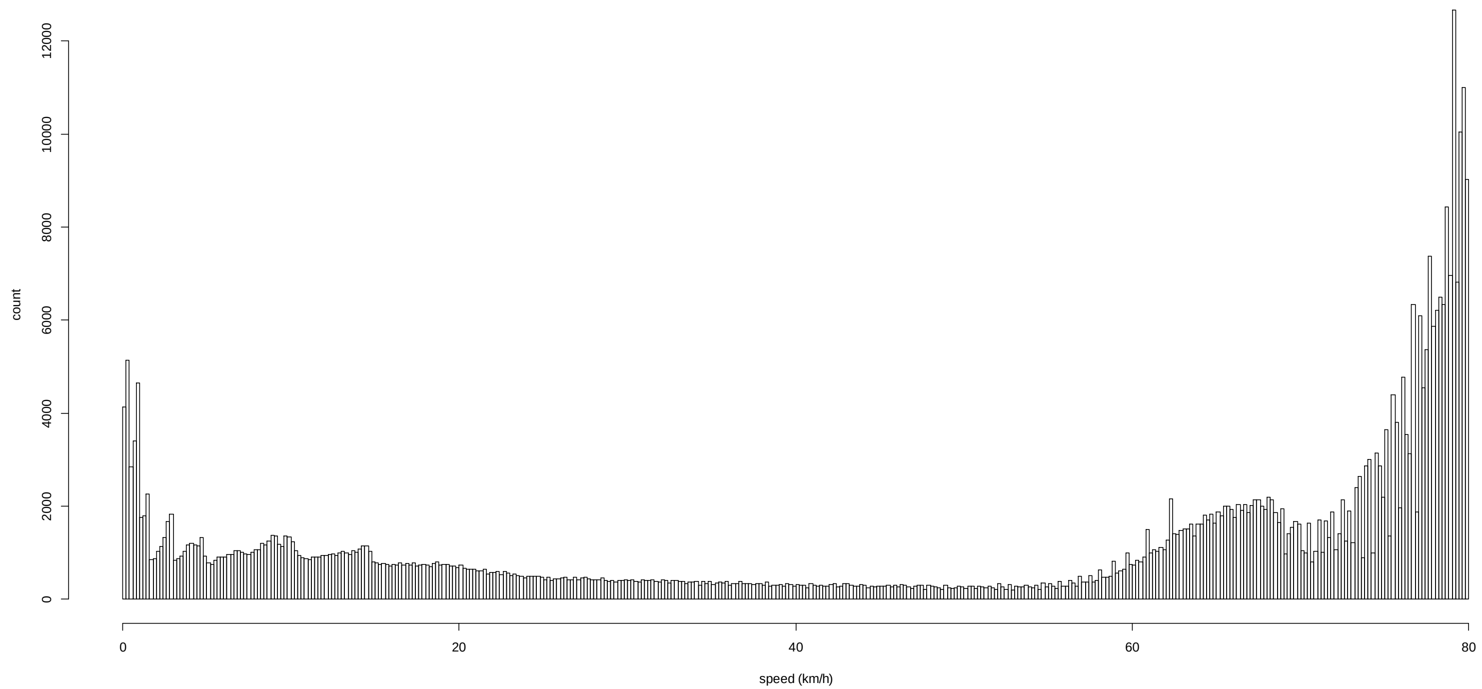
Konstrukce proxy pro příznak parkování

(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

- Optimalizace zvoleného kritéria

  FN, FP

  total misclassification

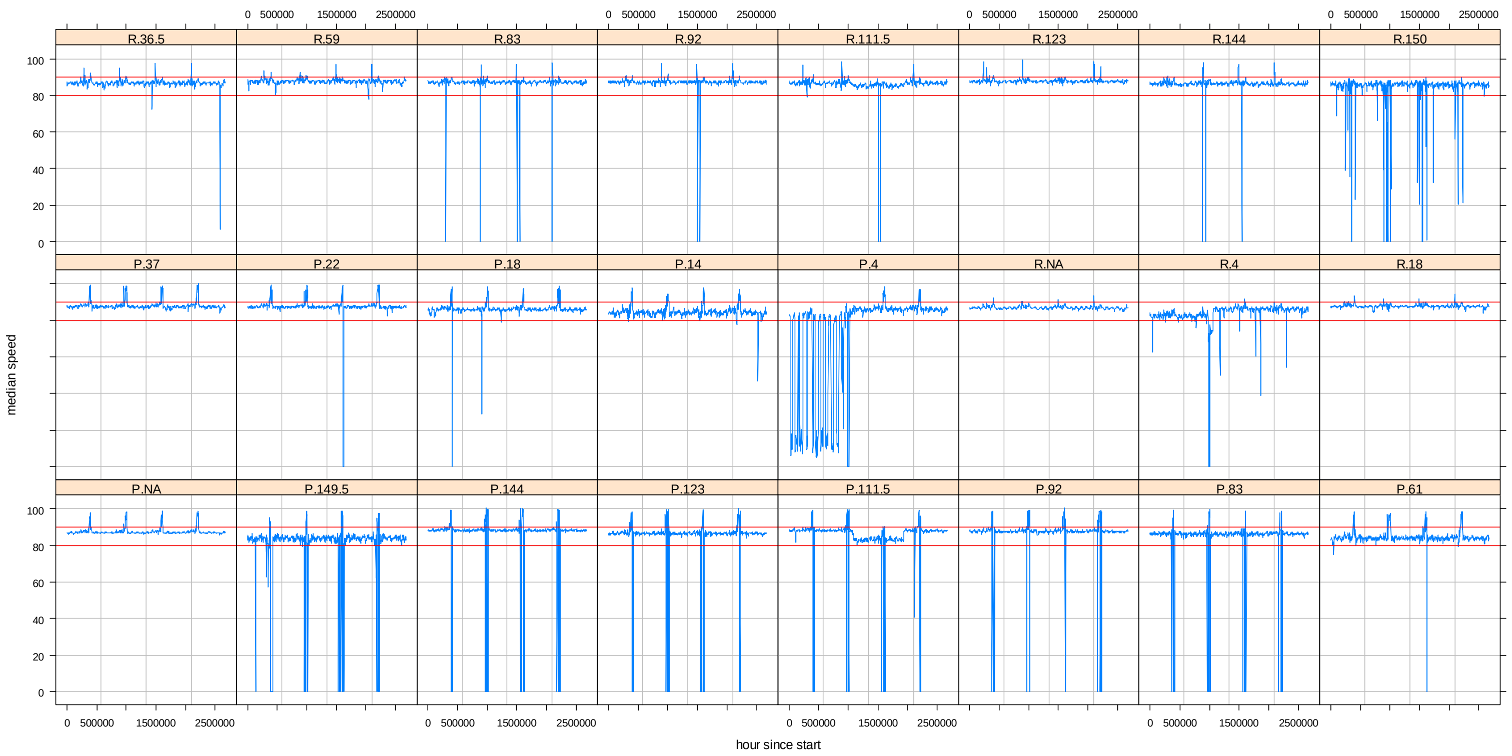
  weighted total miscalssification

  ROC (AUC)

- Klasifikační pravidla pevná/globální vs. lokální

Časová a prostorová nehomogenita

(mediánová trajektorie)



Kritéria pro klasifikaci parkování

- Rychlost

  lokální pravidlo

  1/3 mediánu rychlosti pro daný úsek v době průjezdu

A

$$Z = \text{Cas\_dif.min} - 60 * \text{Vzdalenost/(dana rychlost)}$$

- Doba zbývající pro parkování

  při mediánové rychlosti průjezdu úsekem alespoň B minut

Důsledky pro konstrukci proxy proměnné

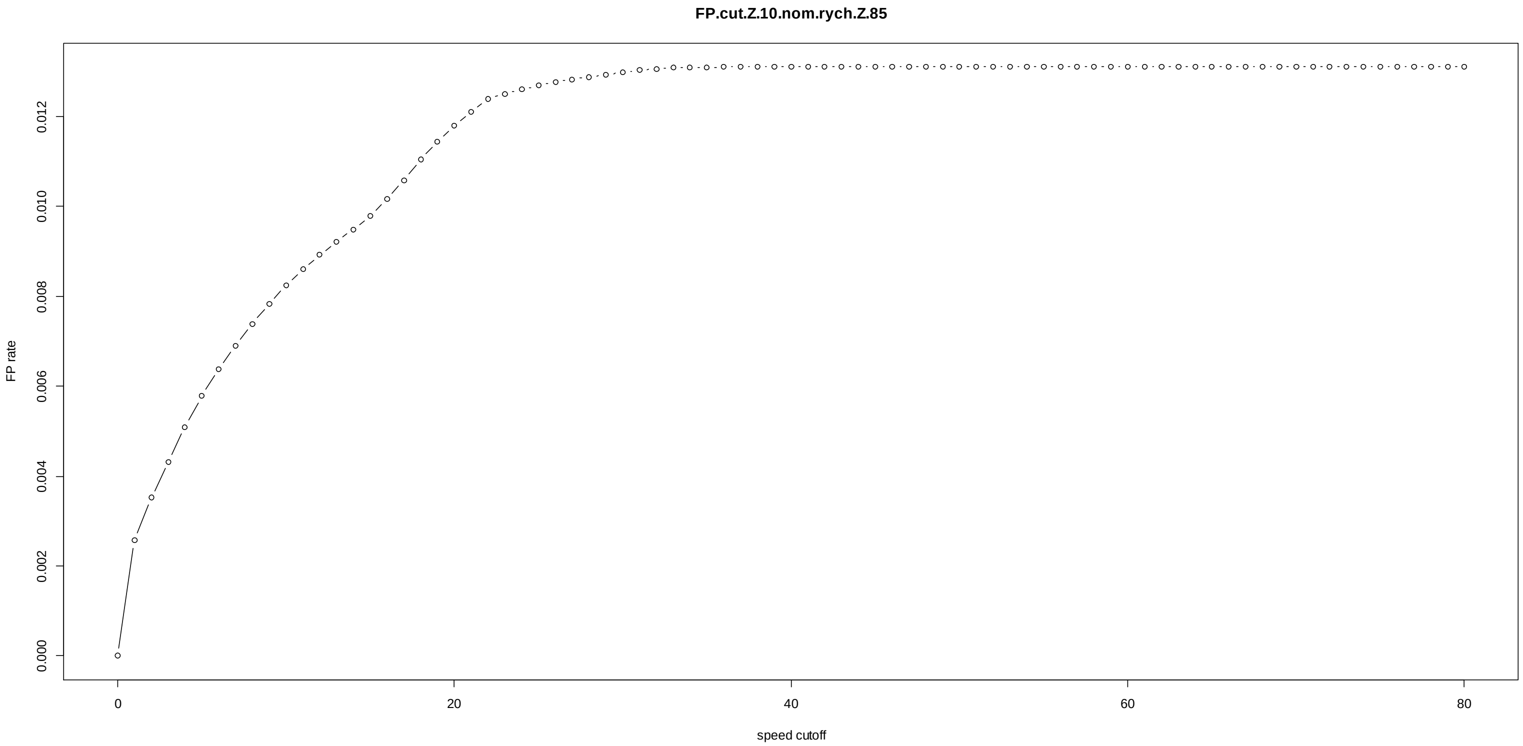
- Kritéria pro hodnocení parkovacího statusu musí být lokální (v prostoru i čase)

- Musí zohledňovat nepravidelností dané sníženou rychlostí dopravního proudu v důsledku

  - odlišného výškového profilu úseku
  - congestion
  - dopravních omezení

    bylo by dobré aby predikční systém o plánovaných omezeních „věděl“

Vliv cutoffu rychlosti na FN



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

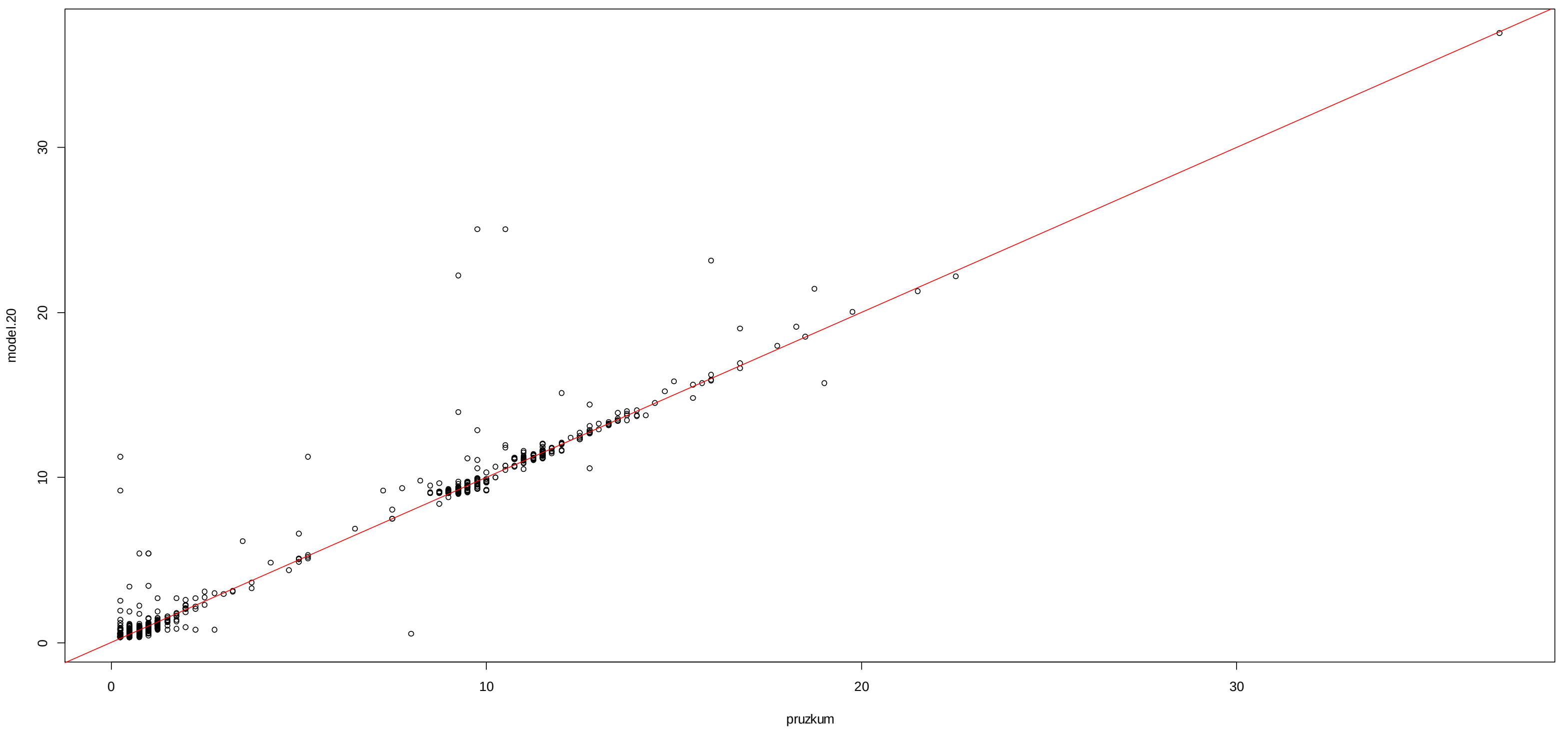
  - úseky bez parkoviště nemusejí mít všechny vlastnosti stejné jako úseky s parkovištěm

  - FP takto můžeme „optimalizovat“, mnohem důležitější FN ale nikoli

<table>
<tr><td colspan="2" rowspan="2">Misclassifikace z
manuální sčítací kampaně

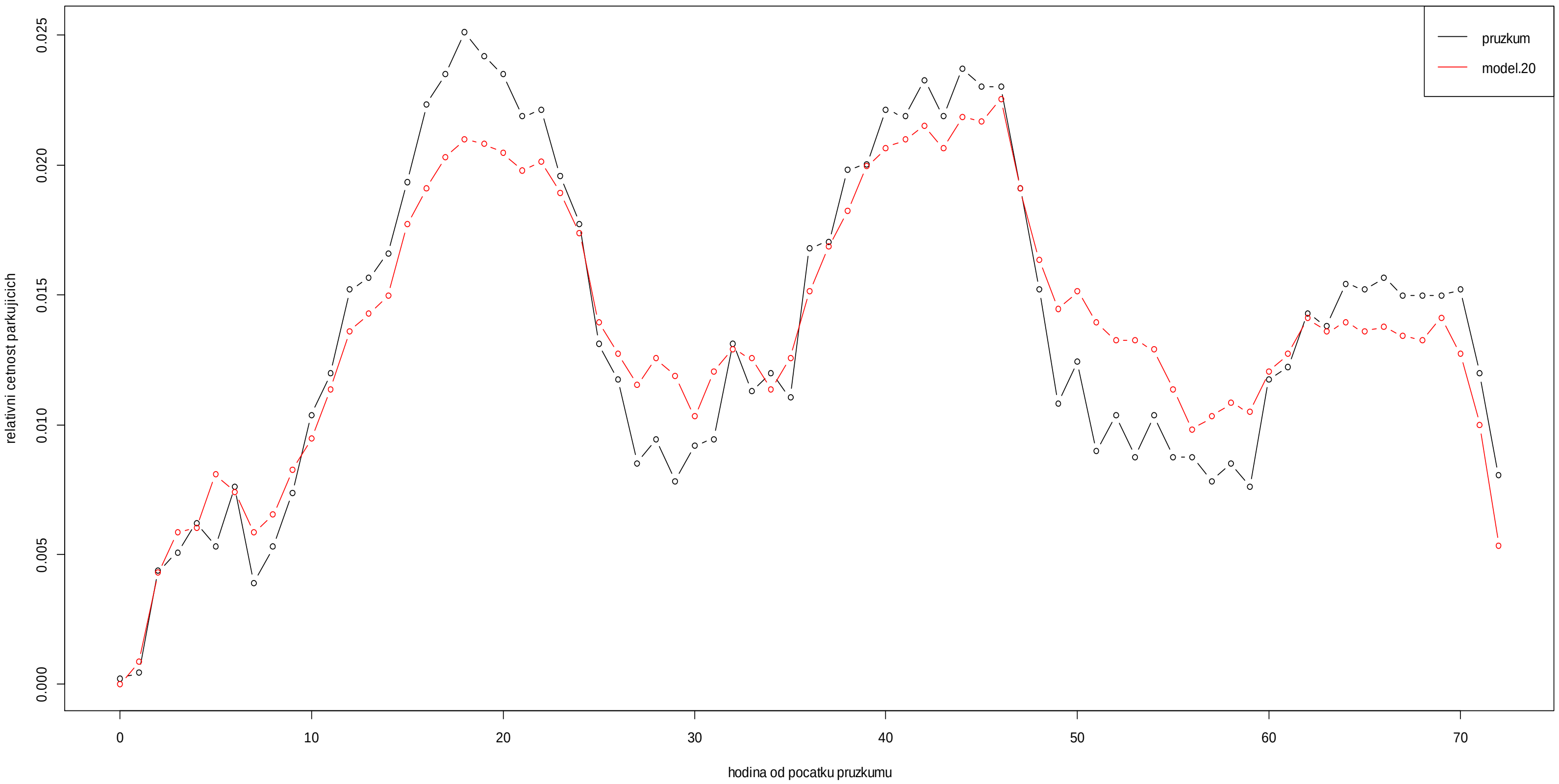
Vezmeme-li manuální kampaň za „zlatý standard“, máme:
total misclassification=0.0368
FN 0.0209
FP 0.3042</td><td colspan="2">Model.20</td></tr>
<tr><td>Neparkuje</td><td>Parkuje</td></tr>
<tr><td rowspan="2">Manuální kampaň</td><td>Neparkuje</td><td>0.9241</td><td>0.0197</td></tr>
<tr><td>Parkuje</td><td>0.0171</td><td>0.0391</td></tr>
</table>

Vztah mezi manuálním sčítáním a proxy

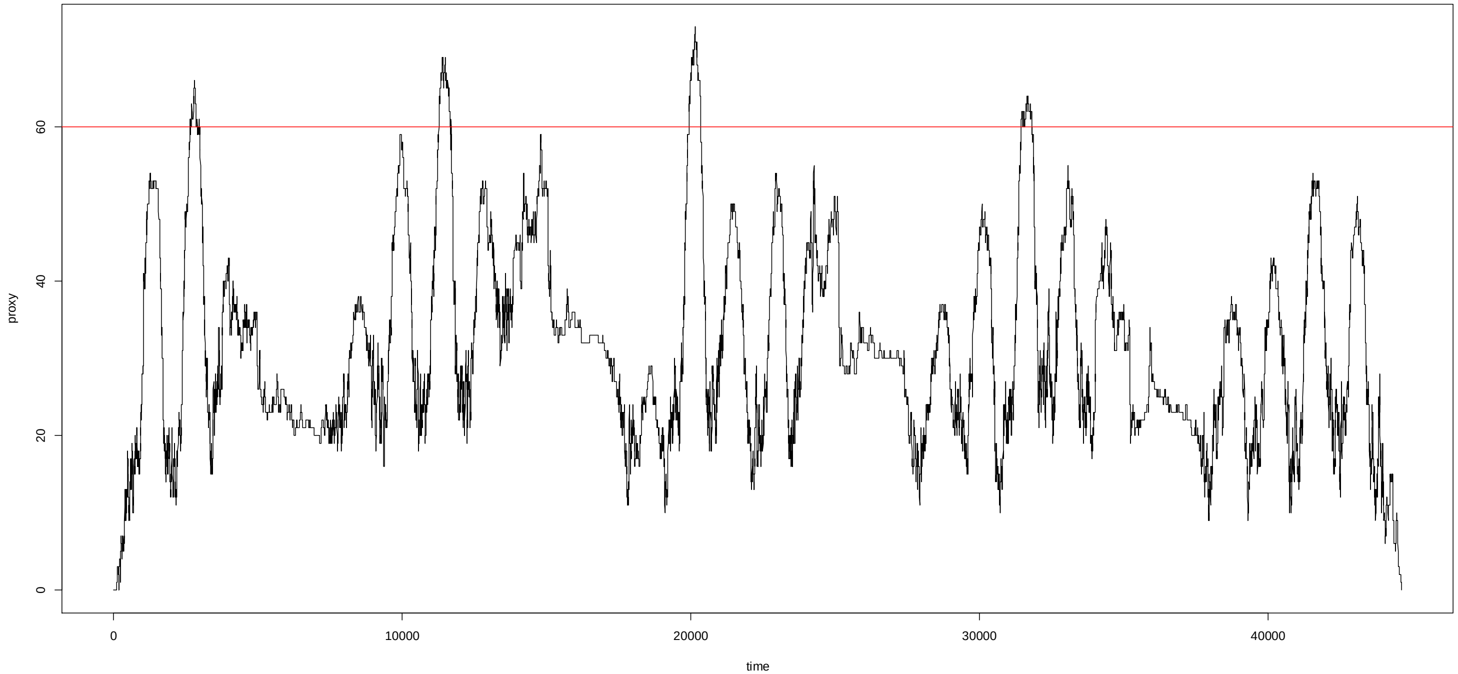


Časově-proměnlivá kalibrace a různá intenzita krátkodobého parkování v průběhu dne

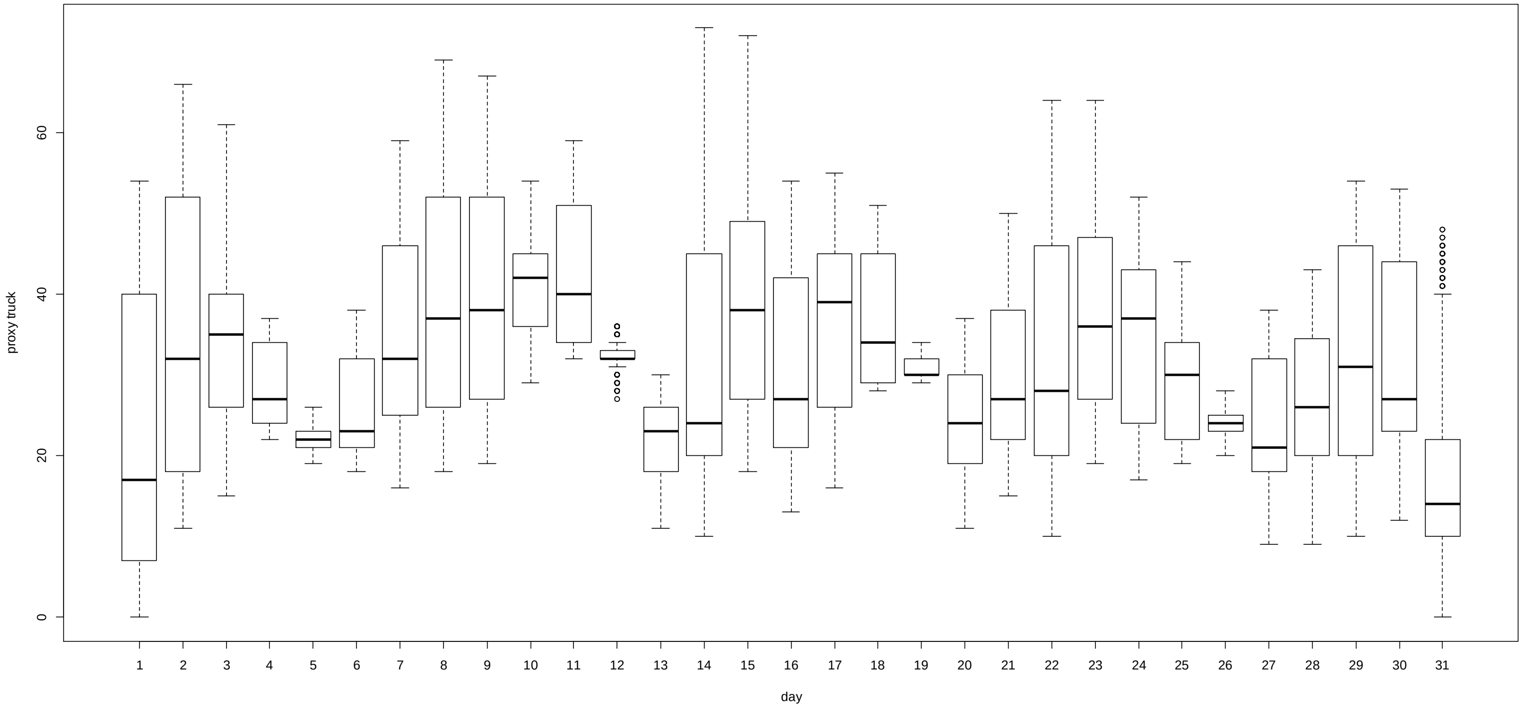
(výsledky manuálního sčítání)



Počty parkujících odvozené z proxy proměnné a nominální kapacita parkoviště



Proxy pro jednotlivý úsek, chování v různých kalendářních dnech



Statické modelování, I

- Poissonovský model

  GLM třída

  výhoda IRWLS algoritmu

  zohlednění heteroskedasticity

  periodicita v intenzitách

- Overdispersion

  quasi-poisson GLM

  negativně-binomický model

Statické modelování, II

$P_{jt} \sim Poi(\mu_{jt})$ log jako link

Wliv typu dne a hodiny:

- Additive (on the log scale)

$$\log(\mu_{jt}) = \alpha_j + \beta_{j,d(t)} + \gamma_{j,h(t)}$$

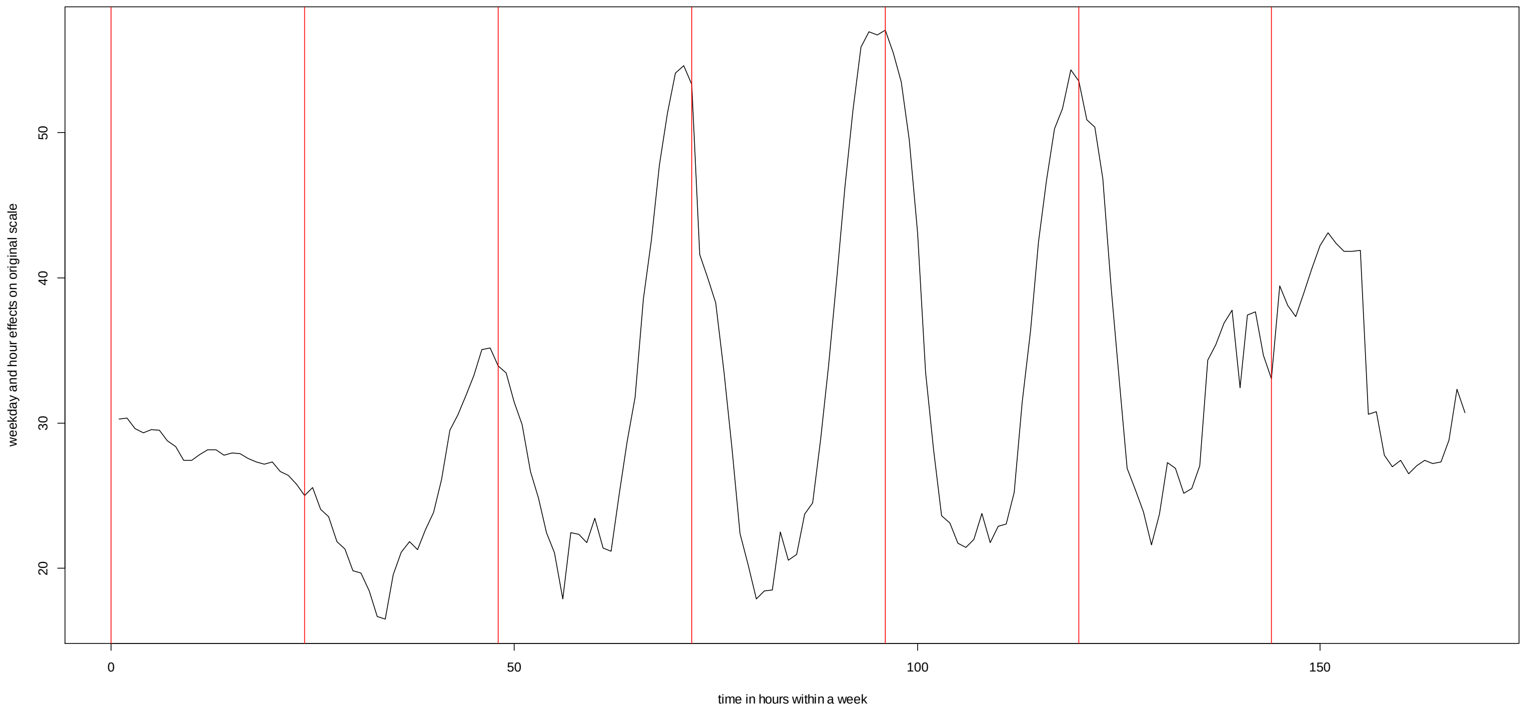
- Interaction (non-multiplicative on original scale)

$$\log(\mu_{jt}) = \alpha_j + \beta_{j,d(t)} + \gamma_{j,h(t)} + \delta_{j,dh(t)}$$

reparametrizace:

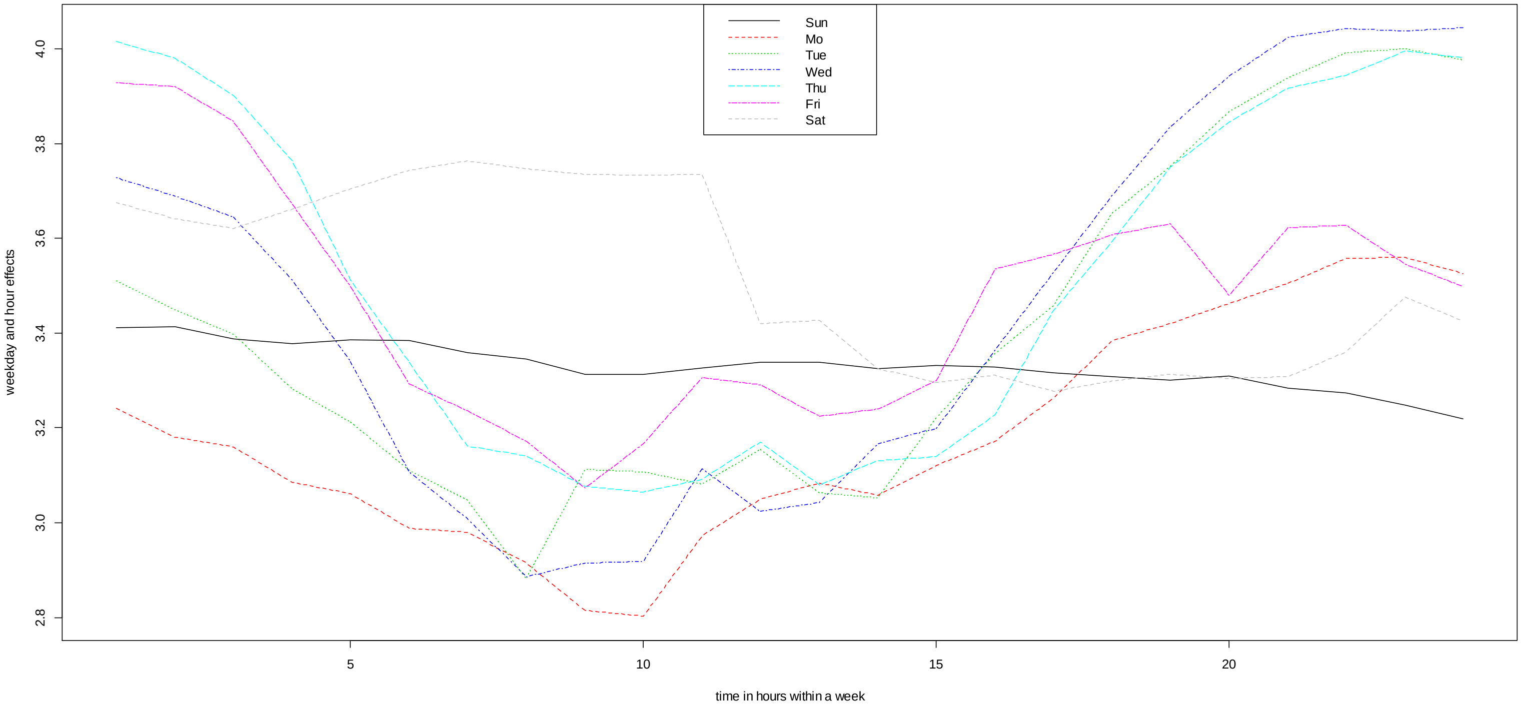
$$\log(\mu_{jt}) = \omega_{j,dh(t)}$$

Zpětná transformace na původní škálu
(# of vozidel)



Trajektorie intenzity parkování pro různé dny v týdnu

(na log škále)



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

Modely s jediným lagem

(jednoduchá alternativa k „distributed lags“ přístupu)

$$P_{jt} \sim Poi(\mu_{jt})$$

- Lag1

$$\log(\mu_{jt}) = \omega_{j,h} + \beta.\log(P_{j,t-1})$$

$\hat{\beta} = 1.025\,(0.004)$

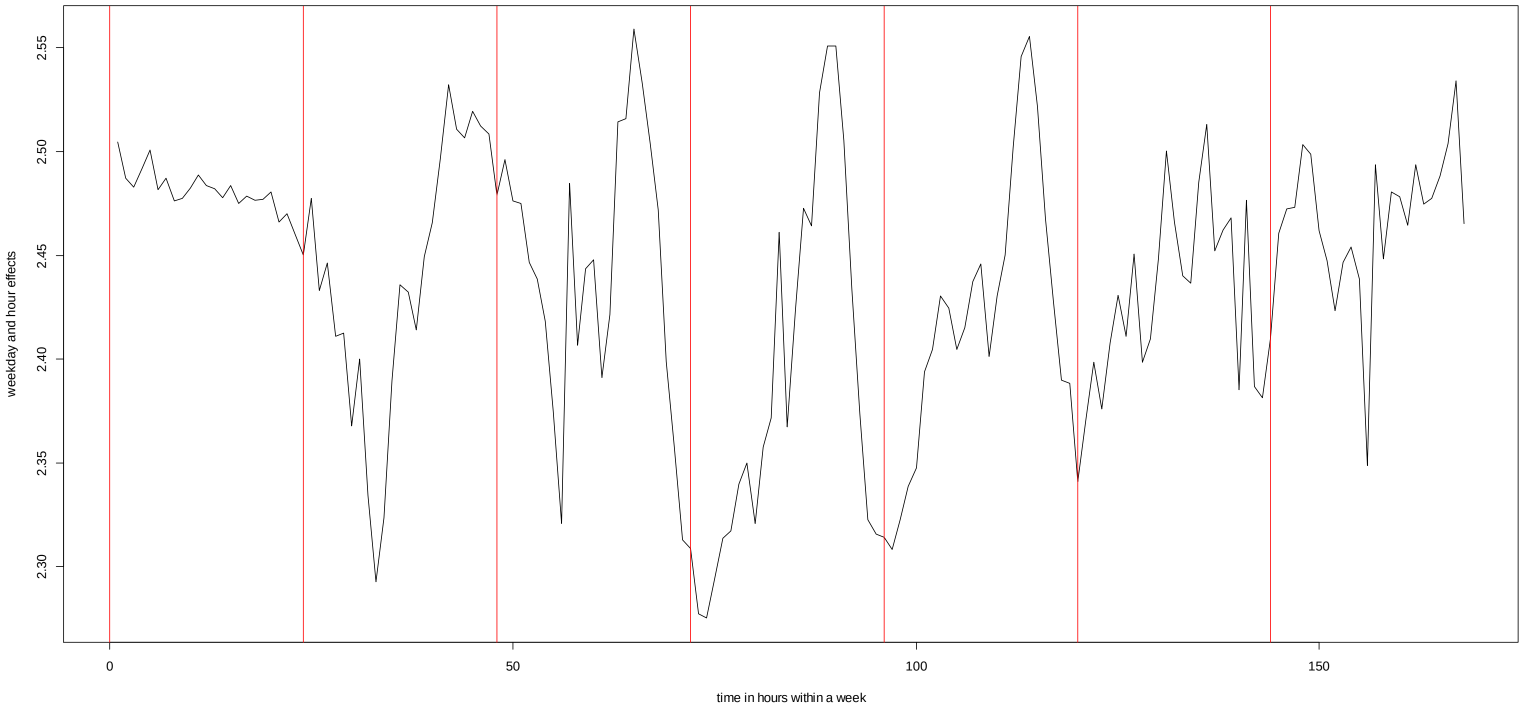
- Lag20 (minut)

$$\log(\mu_{jt}) = \omega_{j,h} + \beta.\log(P_{j,t-20})$$

$\hat{\beta} = 0.950\ (0.004)$

Lag20 model, statická část

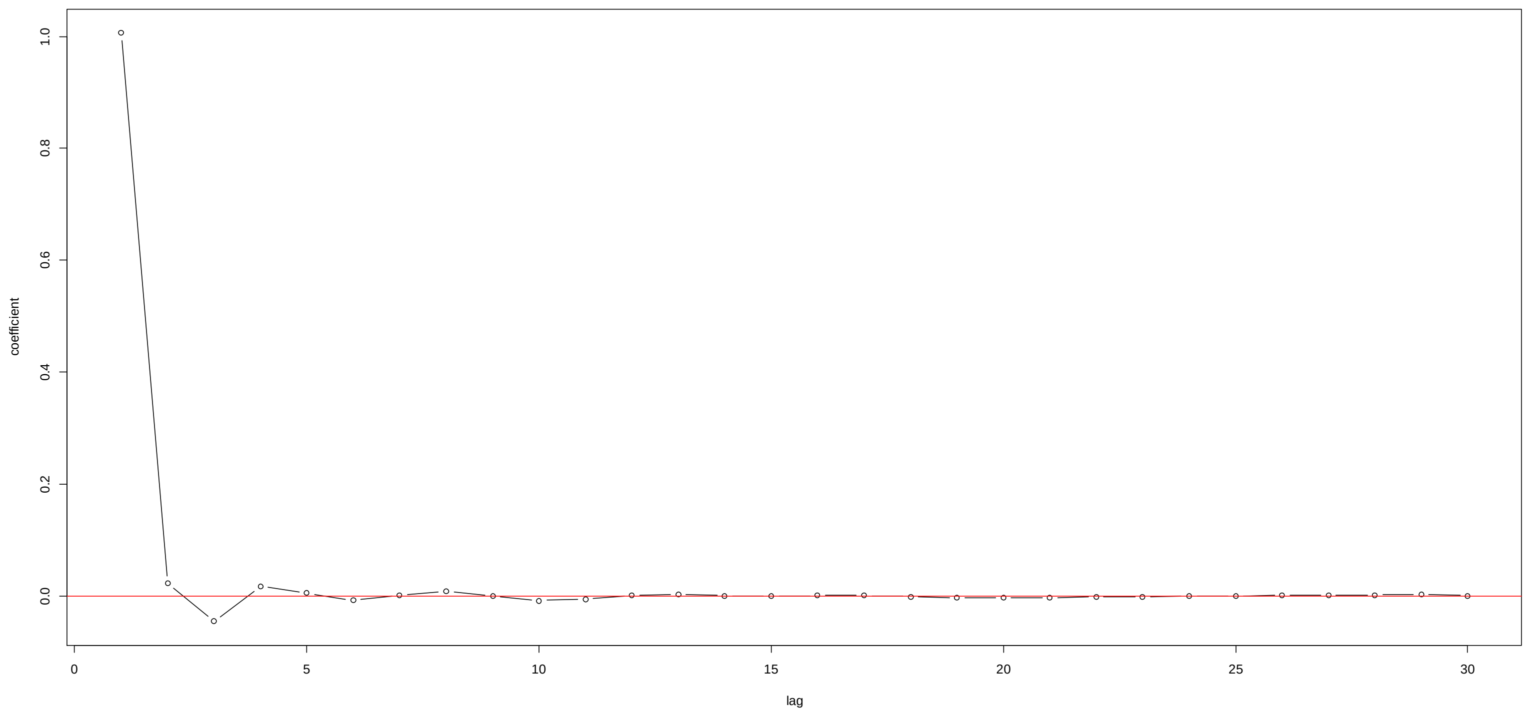
(ne-ortogonalita, přítomnost dynamické části deformuje tvar týdenní periodicity oproti plně statickému modelu)



Použití delší historie?

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

Almon model, (time invariant) filter



Časově-proměnlivý koeficient pro pevně zvolený lag?

- Předchozí dynamické modely měly jeden nebo více lagů s pevným (časově ne-proměnlivým) koeficientem

  dynamická část pak sestává ze specifikace časově- neproměnlivého (byť nelineárního) filtru

- Lze ale uvažovat o konstrukci časově-proměnlivého filtru ve smyslu časově-proměnlivého koeficientu pro pevně daný lag Např. koeficient pro lag 1 či lag 20 se mění s časem (např. dle dne v týdnu a/nebo denní hodiny.

- Ideálně, den*hodina interakce jak pro statickou tak dynamickou část

  Pak máme časově-proměnlivé absolutní členy i směrnice

  Ale: kolinearita jejich odhadů má tendenci likvidovat případné vylepšení vnesené obecnějším/flexibilnějším modelem

Kompromis

- Jediný lag s časově-proměnlivým koeficientem (model bez čistě statické části)

- Odpovídá modelu s trojnou interakcí

  (na log škále máme den v týdnu*hodina*lagovaný počet, bez marginálních členů či nižších interakcí)

- Roughness penalizace nutná pro regularizaci koeficientů popisujících trajektorie

Regularizovaný model s interakcí

- Poisson

$$P_{jt} \sim Poi(\mu_{jt})$$

- Linearní prediktor na (log) link škále

$$\log(\mu_{jt}) = \beta_{jh} . \log(P_{j,t-20})$$

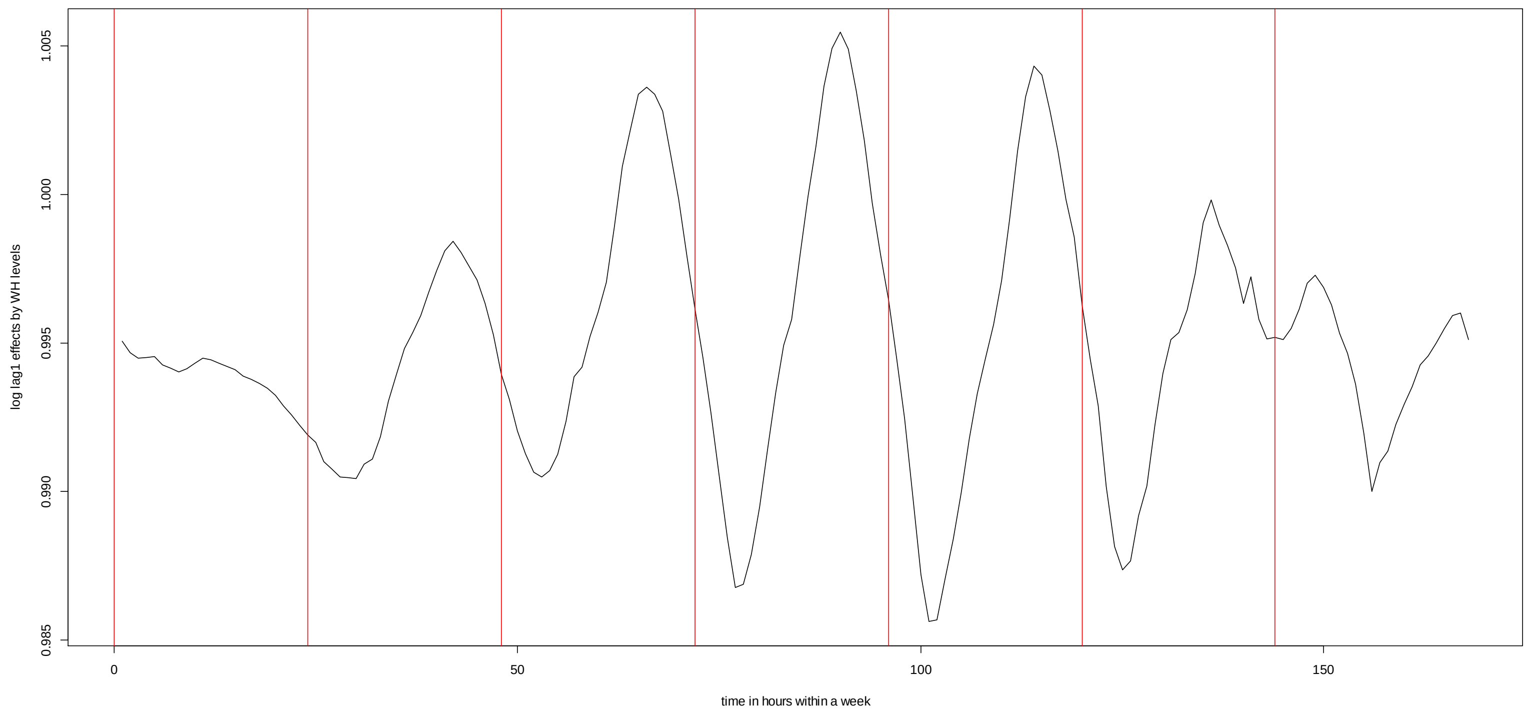
(časově-nehomogenní Markov Poisson model)

- Roughness penalizace koeficientů

$$L(\underline{\beta}) + \lambda.(\underline{\beta}.H)'.(\underline{\beta}.H)$$

s $\underline{\beta} = \begin{pmatrix} \beta_{j1} \\ \vdots \\ \beta_{jH} \end{pmatrix}$ a $\begin{pmatrix} 1 & 0 & \dots & -1 \\ -1 & 1 & \dots & 0 \\ \vdots & & \dots & \vdots \\ 0 & \dots & -1 & 1 \end{pmatrix}$

Lag20, změny koeficientu uvnitř týdne



Testování (prakticky motivovaných) hypotéz

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

- Lineární prediktor expandován na

$$\log(\mu_{jt}) = \beta_{jh}.\log(P_{j,t-20}) + \delta.\log(P_{j-1,t-20})$$

- Přídavek členu odpovídajícímu lag20 opředchozícho parkoviště lze snadno testovat (LRT)

$$\hat{\delta} = 0.419,\ \ (0.001),\ \ p - value < 0.0001$$