

PROJEKT

AUTOR: Dana Kozáková

14.12.2024

Predmet: PřF:M9121 Časové řady I

OBSAH

1	Dáta – zdroj a obsah.....	2
1.1	Dáta o glykémii	2
2	Preskúmanie dát.....	3
2.1	Graf časového radu	3
2.2	ACF a PACF.....	5
2.3	Philips-Perronov test na potvrdenie stacionarity	5
3	Fitovanie modelov	6
3.1	AR(1) model.....	6
3.2	AR(2) model.....	8
3.3	SARMA(1,0)(0,1) ₂₄	10
3.4	SARMA(1,0)(1,0) ₂₄	12
3.5	Navrhnutý model – SARMA(2,3)(1,1) ₂₄	14
4	Výsledky – interpretácia	16
5	Záver	17

Obrázok 1	Uzavretá slučka - Closed loop	2
Obrázok 2	Vstupné dáta	3
Obrázok 3	Dáta upravené pre spracovanie	3

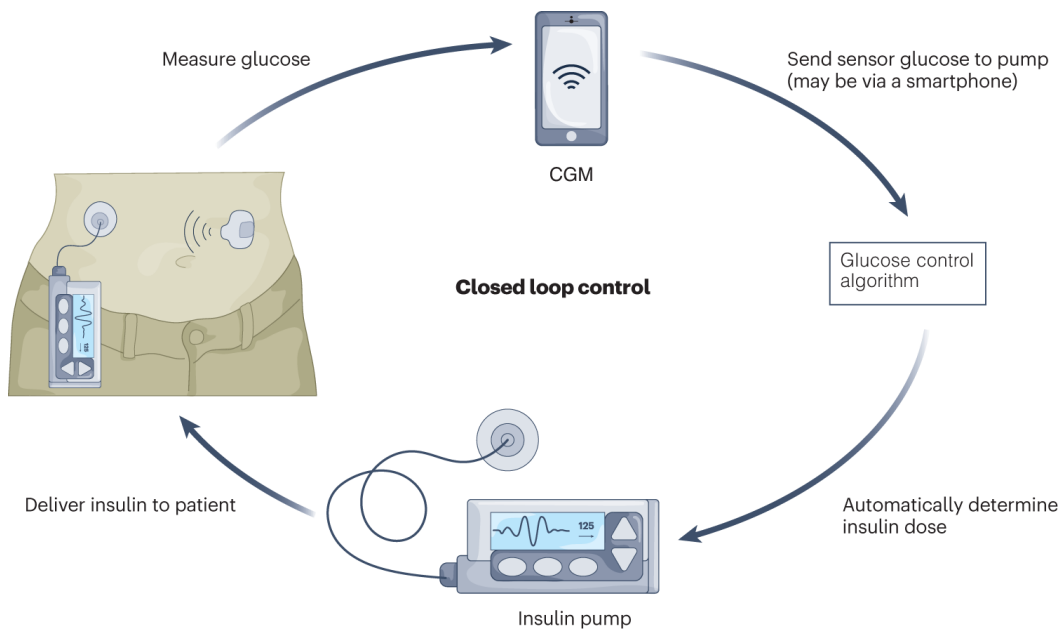
Graf 1	Glykémia pacienta počas 180 dní - časový rad	4
Graf 2	Glykémia pacienta počas 7 dní – výber z časového radu	4
Graf 3	ACF a PACF pre časový rad.....	5
Graf 4	Reziduá pre ARIMA(1,0,0)	6
Graf 5	Porovnanie predpovedaných a skutočných hodnôt pre AR(1)	7
Graf 6	Reziduá pre ARIMA(2,0,0) v porovnaní s ARIMA(1,0,0)	8
Graf 7	Porovnanie predpovedaných a skutočných hodnôt pre AR(2)	9
Graf 8	Reziduá pre SARIMA(1,0,0)(0, 0, 1) ₂₄ v porovnaní s ARIMA(2,0,0)	10
Graf 9	Porovnanie predpovedaných a skutočných hodnôt pre SARIMA(1, 0, 0)(0, 0, 1) ₂₄	11
Graf 10	Reziduá pre SARIMA(1,0,0)(1, 0, 0) ₂₄ v porovnaní s predchádzajúcim modelom	12
Graf 11	Porovnanie predpovedaných a skutočných hodnôt pre SARIMA(1, 0, 0)(1, 0, 0) ₂₄	13
Graf 12	Reziduá pre navrhnutý SARIMA(2,0,3)(1, 0, 1) ₂₄ v porovnaní s predchádzajúcim modelom	15
Graf 13	Porovnanie predpovedaných a skutočných hodnôt pre SARIMA(2, 0, 3)(1, 0, 1) ₂₄	15

1 Dáta – zdroj a obsah

1.1 Dáta o glykémii

Zdroj: Klinický manažment a nastavenie parametrov kontrolovanej uzavretej slučky v systéme Control-IQ:
Výsledky 6-mesačnej multicentrickej randomizovanej klinickej štúdie

- údaje o 168 diabetikoch, u ktorých sledovali hladinu glykémie
- cieľ štúdie = potvrdiť efektívnosť uzavretého slučky v systéme podávania inzulínu



Obrázok 1 Uzavretá slučka - Closed loop

Dáta pre projekt:

- výber náhodného jedného náhodného pacienta z tejto štúdie a skúmanie jedného časového radu = jeho glykémií počas 180 dní

	PtID	RecID	DataDtTm	CGMValue	HighLowIndicator	DataDtTm_adjusted
	<int>	<int>	<chr>	<int>	<int>	<chr>
1	22	1	2018-05-30 16:44:14	83	0	
2	22	2	2018-05-30 16:59:15	117	0	
3	22	3	2018-05-30 16:49:16	98	0	

Obrázok 2 Vstupné dáta

Príprava dát:

- filter = výber vhodného pacienta (dostatok dát, bez NA)
- agregácia = priemerná glykémia za hodinu
- prevod jednotky: z dg/l na mmol/l

	date_hour	CGM_mmol
	<dtm>	<dbl>
1	2018-05-30 16:00:00	5.68
2	2018-05-30 17:00:00	5.05
3	2018-05-30 18:00:00	4.63
4	2018-05-30 19:00:00	5.88
5	2018-05-30 20:00:00	4.94
6	2018-05-30 21:00:00	7.96

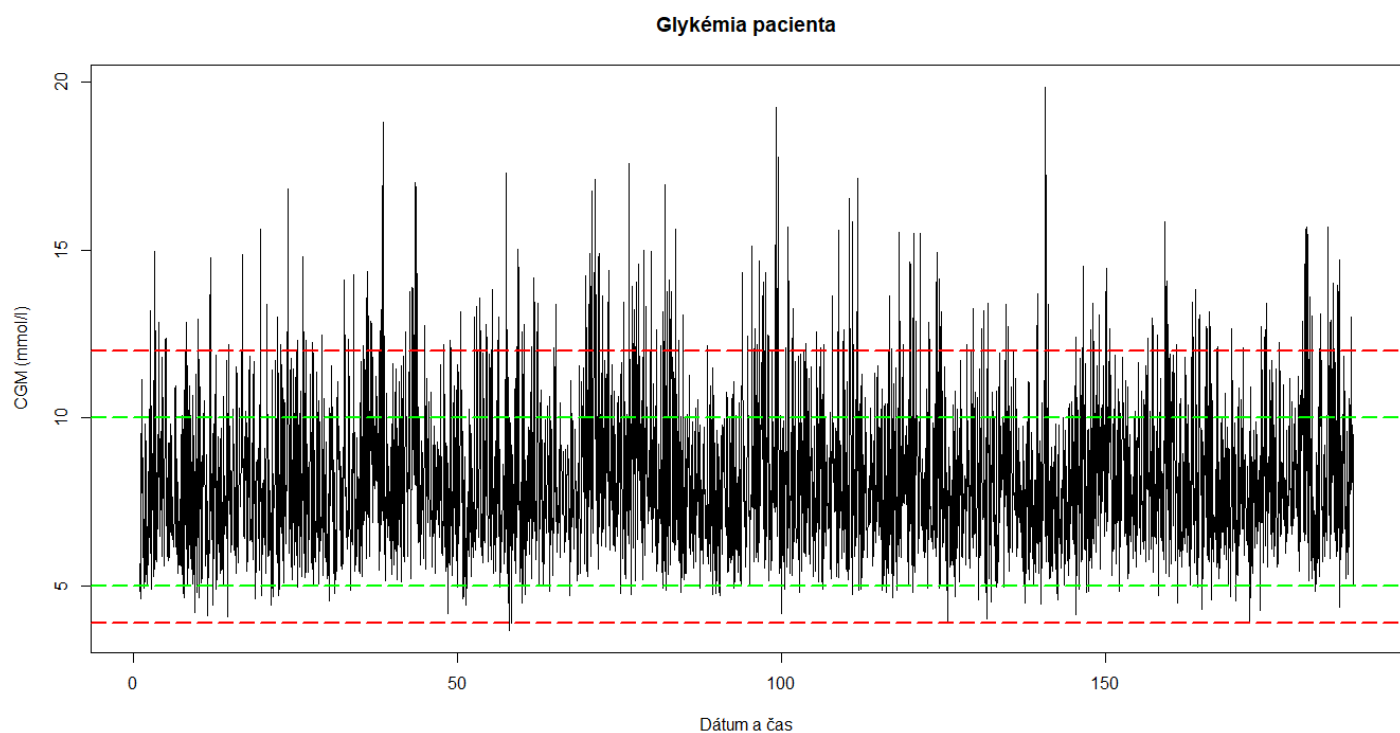
Obrázok 3 Dáta upravené pre spracovanie

2 Preskúmanie dát

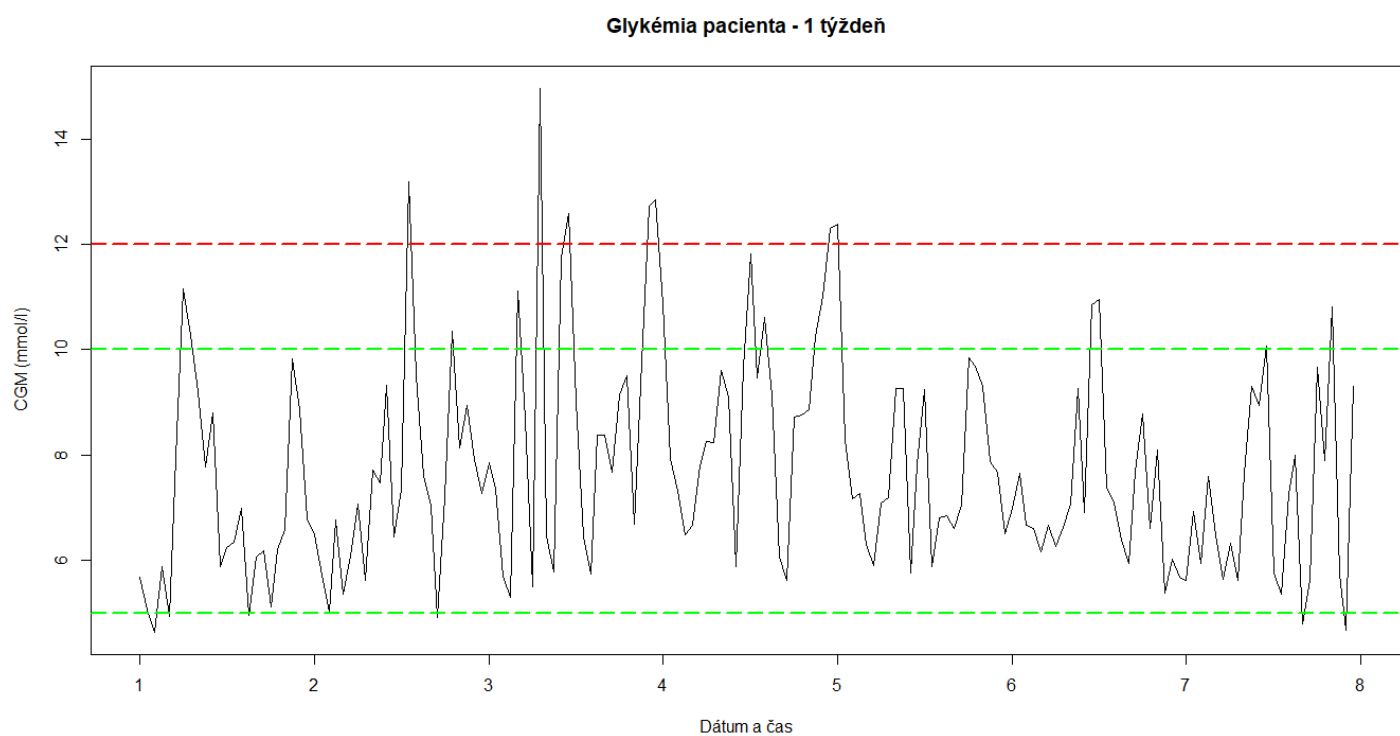
2.1 Graf časového radu

Glykémia jedného pacienta počas 180 dní – priemerná hodnota glykémie v mmol/l každú hodinu. Vyznačené sú dve úrovne:

- ideálna: glykémia od 5 do 10 mmol
- akceptovateľná: od 3,9 do 12 mmol

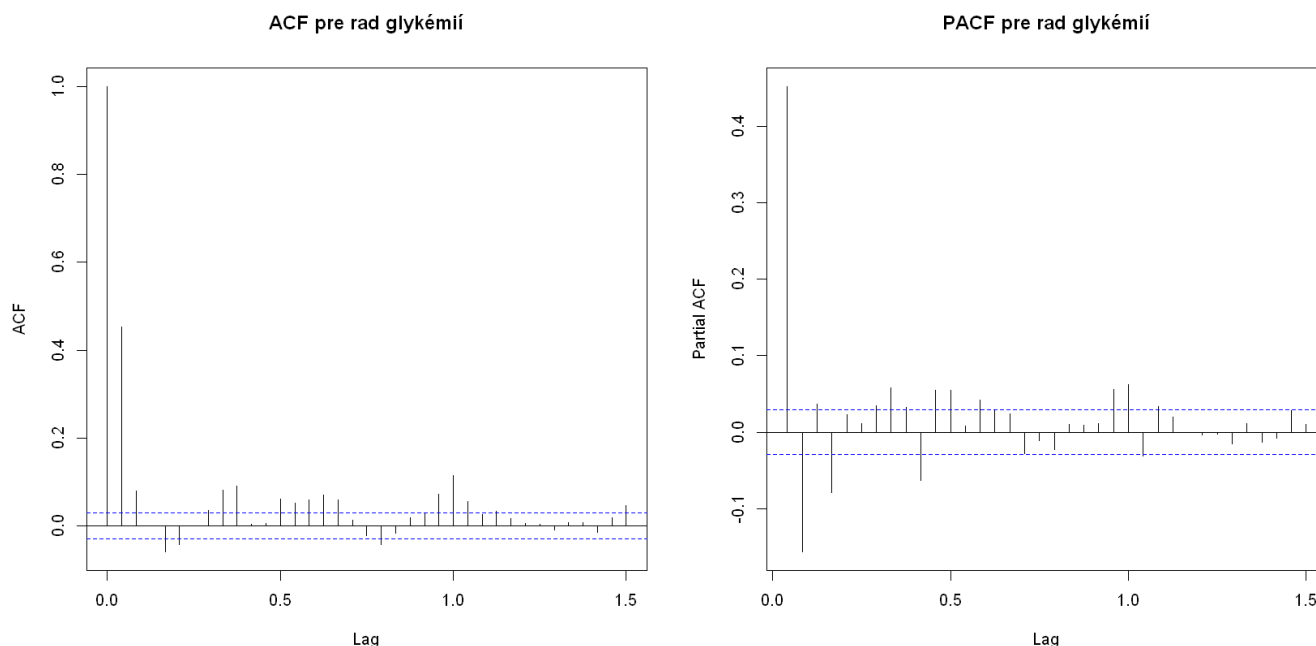


Graf 1 Glykémia pacienta počas 180 dní - časový rad



Graf 2 Glykémia pacienta počas 7 dní – výber z časového radu

2.2 ACF a PACF



Graf 3 ACF a PACF pre časový rad

- ACF klesá veľmi rýchlo, je tam veľmi malý akoby sínusový tvar, možno zvážiť MA zložku. Výraznejšiu autokoreláciu, okrem prvej, vidíme ešte v 1, teda po 24 hodinách
- v PACF vidíme veľmi výrazný prvý lag, možno ešte druhý, ostatné už veľmi nevyčnievajú, čiže vhodný by bol možno AR(1) model, možno AR(2)
- keďže autokorelácia klesá rýchlo, nie je potrebné diferencovať

2.3 Philips-Perronov test na potvrdenie stacionarity

- z dát to zatiaľ vyzerá, že dáta sú stacionárne
- overenie P-P testom:

Phillips-Perron Unit Root Test

Dickey-Fuller = -39.6, Truncation lag parameter = 10, p-value = 0.01

- p-hodnota testu je 0,01. Teda zamietame nulovú hypotézu, môžeme vylúčiť jednotkový koreň => časový rad je stacionárny, t.j. jednotlivé merania majú konštantný priemer, konštantný rozptyl a autokorelácia závisí len od časového posunu (lagu), nie od konkrétneho času

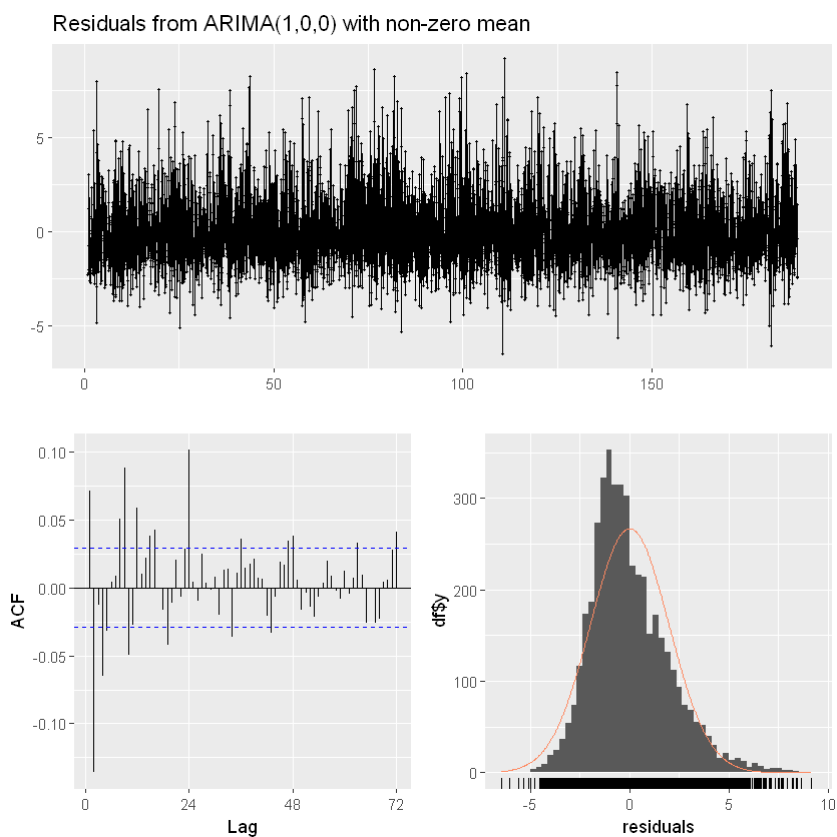
3 Fitovanie modelov

Časový rad je stacionárny, možno aplikovať SARIMA modely. Začnem najjednoduchším: na základe ACF a PACF skúsim AR(1) model.

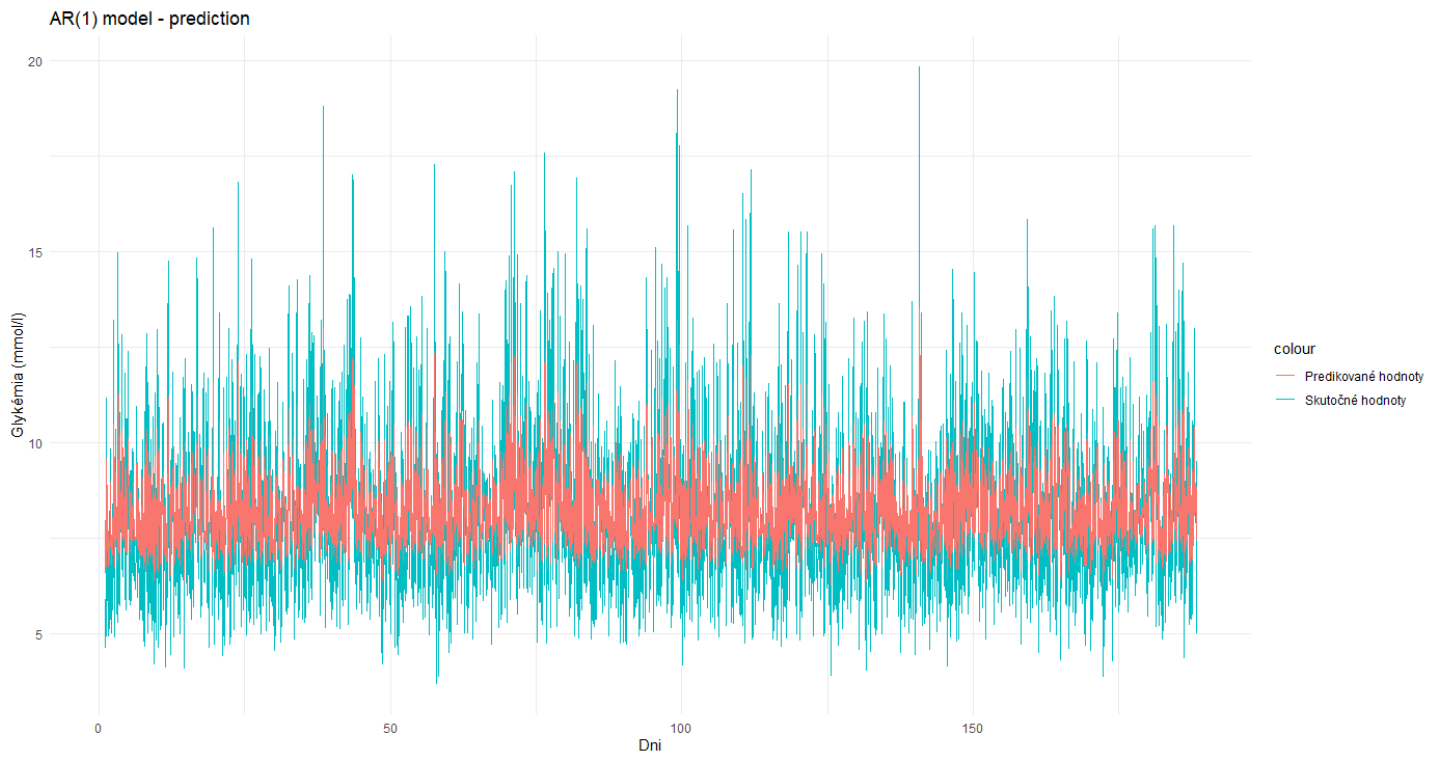
3.1 AR(1) model

`mean(ts_data) = 8.24`

	AR(1)
intercept	8.2461
ar1	0.4525
Ljun-Box test	2.2e-16
Q*	303.33
AIC	18 927.49



Graf 4 Reziduá pre ARIMA(1,0,0)

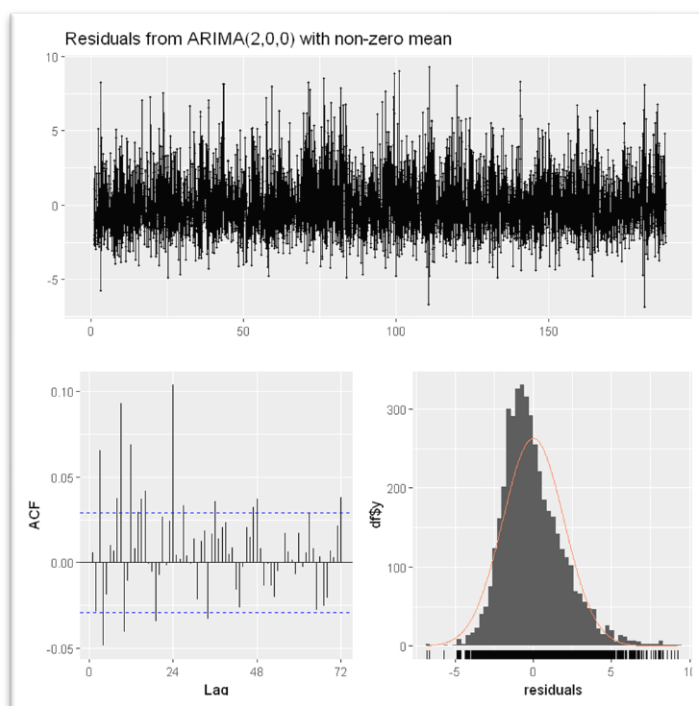
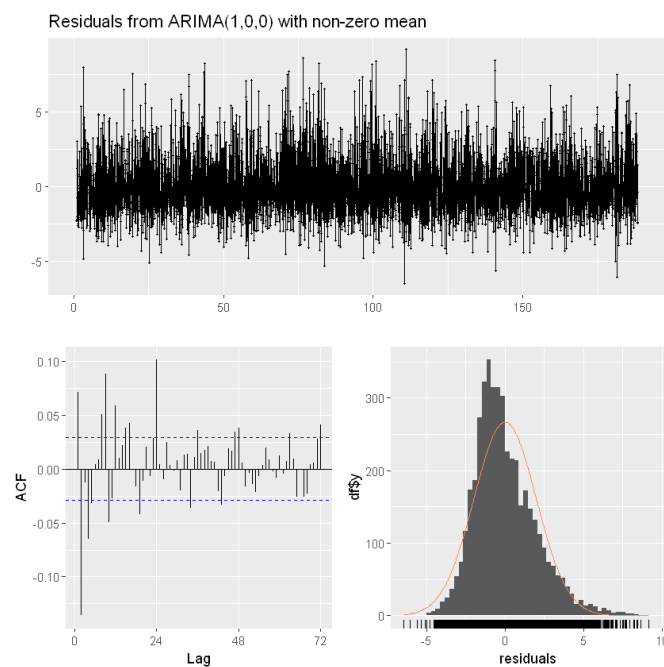


Graf 5 Porovnanie predpovedaných a skutočných hodnôt pre AR(1)

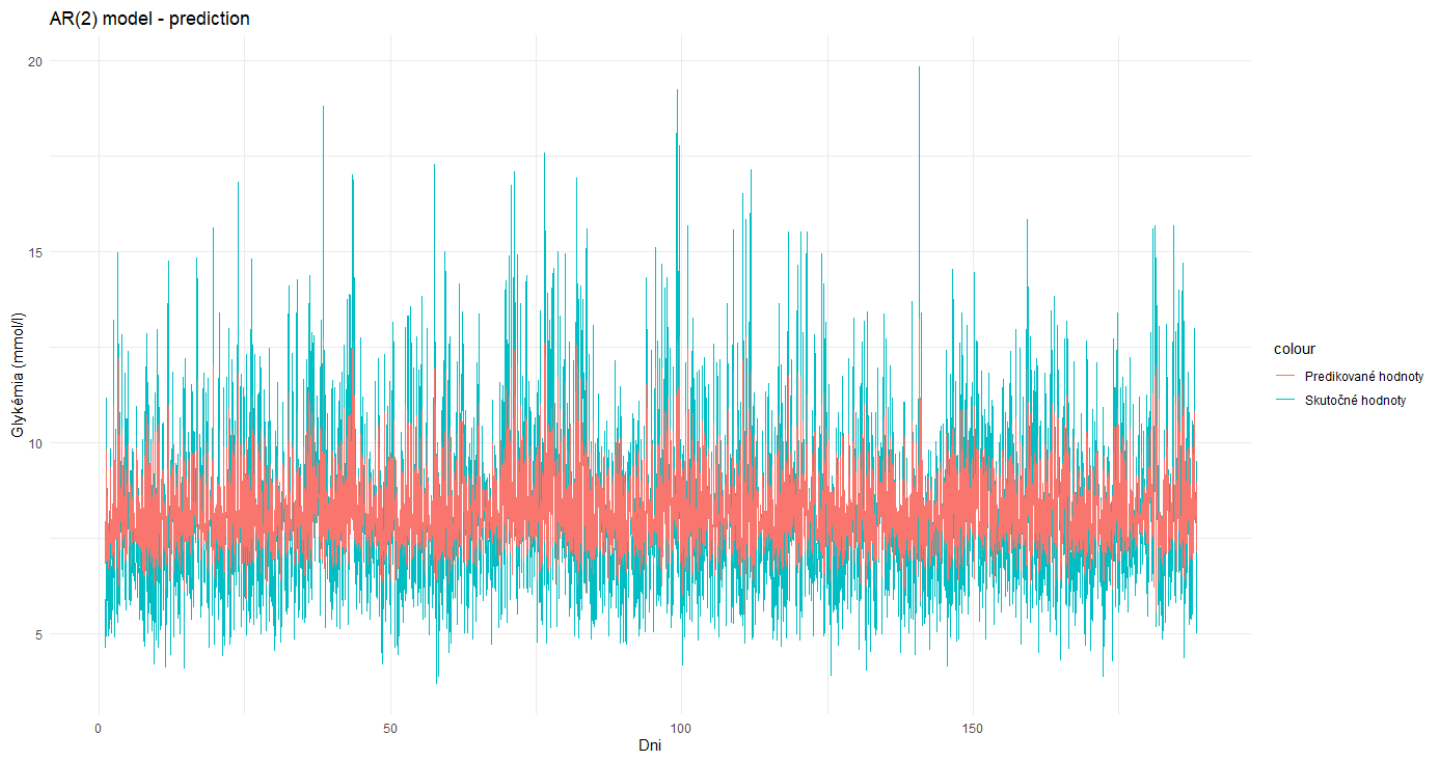
3.2 AR(2) model

`mean(ts_data) = 8.24`

	AR(1)	AR(2)
intercept	8.2461	8.2462
ar1	0.4525	0.5233
ar2		-0.1568
Q*	333.03	236.6
Ljun-Box test	2.2e-16	2.2e-16
AIC	18 927.49	18 817.62



Graf 6 Reziduá pre ARIMA(2,0,0) v porovnaní s ARIMA(1,0,0)



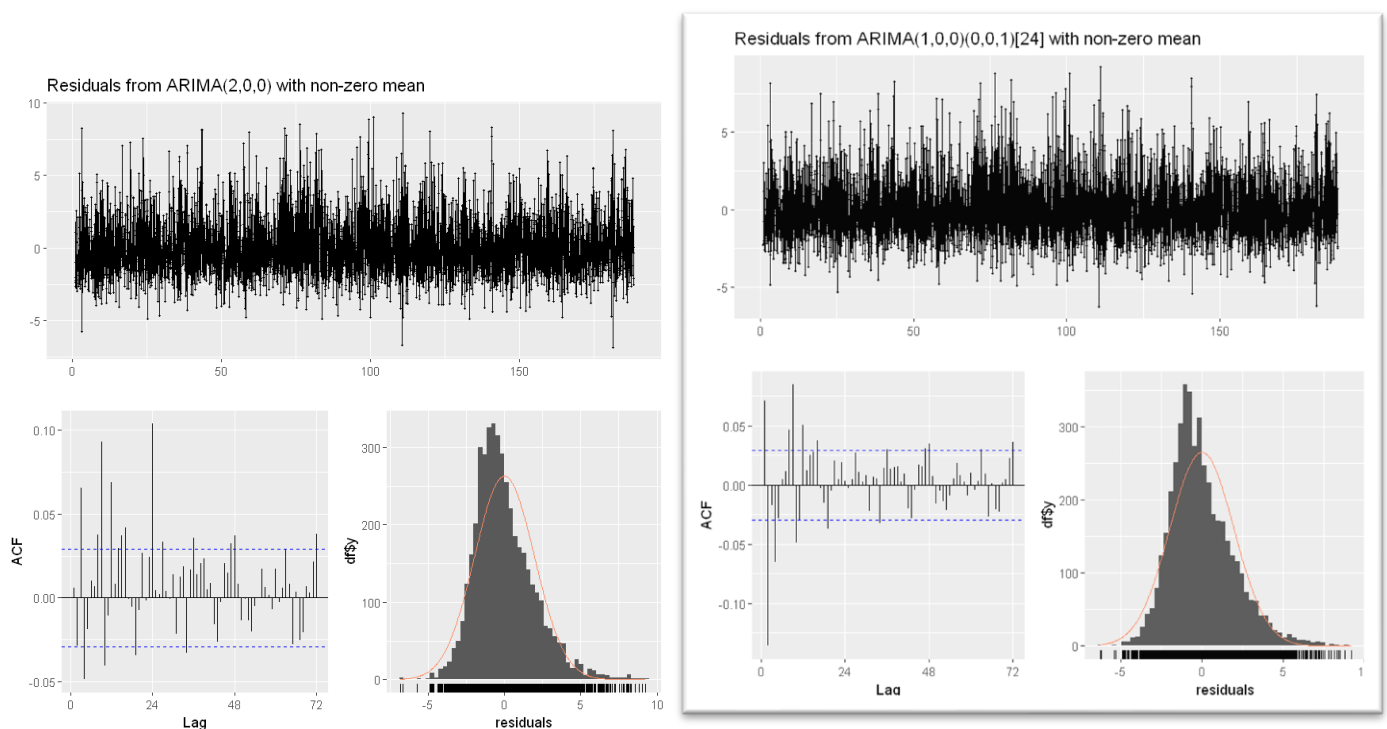
Graf 7 Porovnanie predpovedaných a skutočných hodnôt pre AR(2)

3.3 SARMA(1,0)(0,1)24

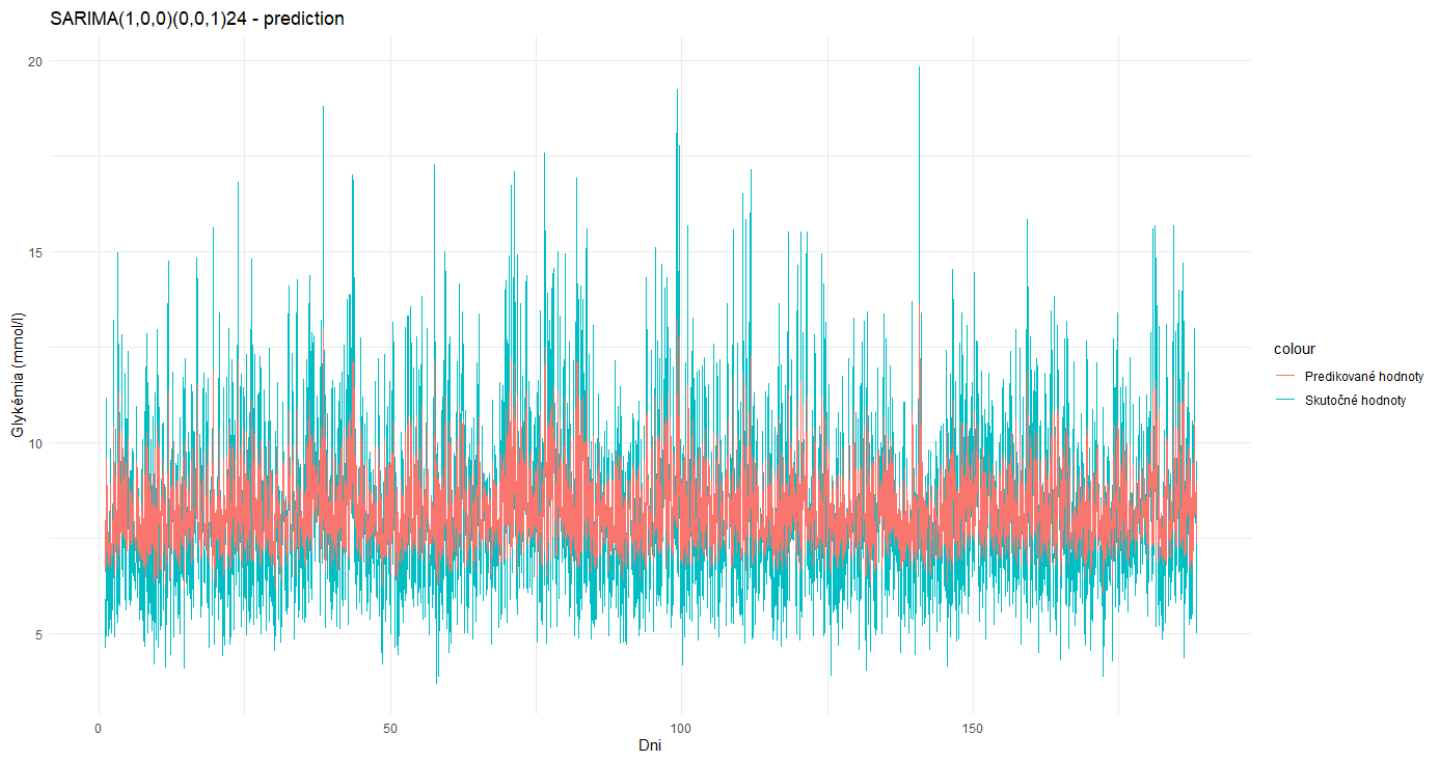
Predchádzajúce pokusy zatiaľ nepriniesli úplne uspokojivé výsledky. Na základe znalosti domény a aj z ACF grafu časové radu možno predpokladať:

- autokoreláciu v rámci dňa – najmä hodinu-dve dozadu AR(1) alebo AR(2) (výrazný lag1 v PACF)
- sezónnosť 1 deň, t.j. 24 hodín – teda MA(1) pre $s = 24$ pre (mierne vyšší lag24 v ACF)
- $\text{mean}(\text{ts_data}) = 8.24$

	AR(1)	AR(2)	SARMA(1,0)(0,1)24
intercept	8.2461	8.2462	8,2461
ar1	0.4525	0.5233	0,4503
ar2		-0.1568	
sma1			0,0966
Q*	333.03	236.6	261.56
Ljun-Box test	2.2e-16	2.2e-16	2.2e-16
AIC	18 927.49	18 817.62	



Graf 8 Reziduá pre SARIMA(1,0,0)(0, 0, 1)24 v porovnaní s ARIMA(2,0,0)



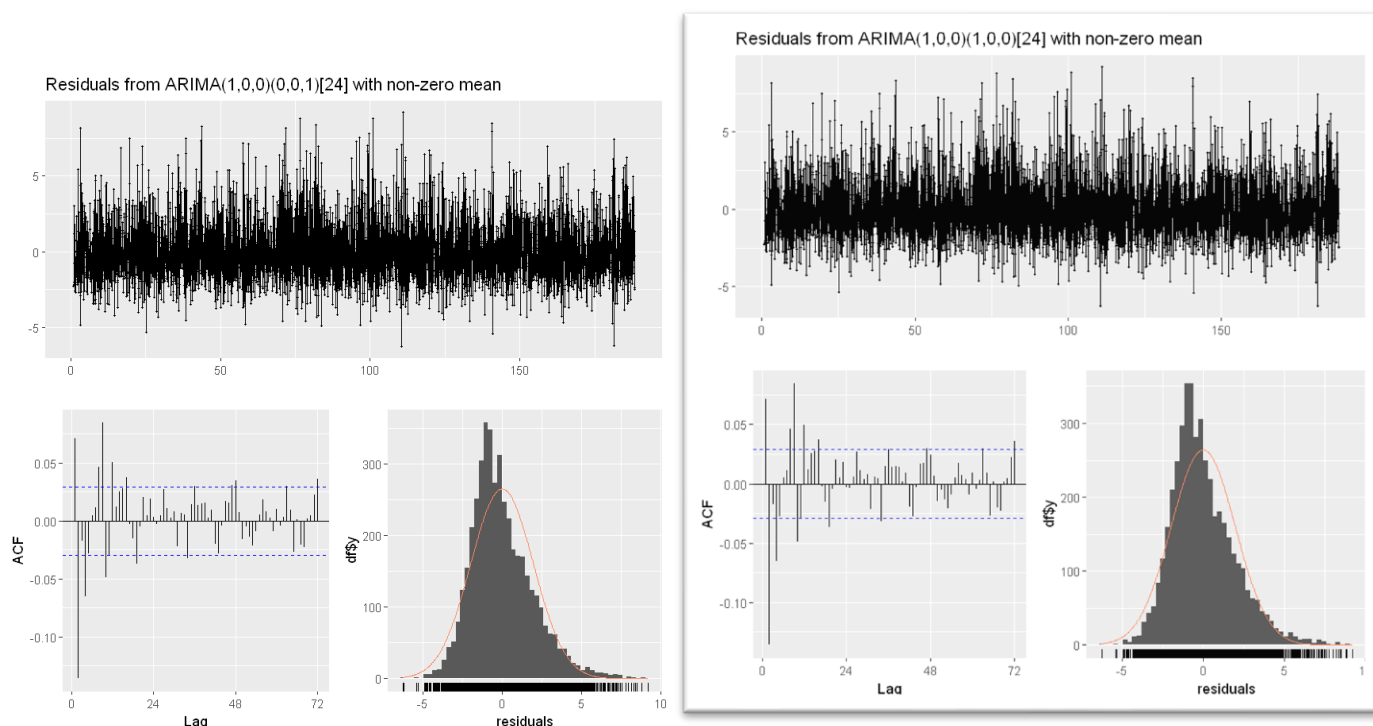
Graf 9 Porovnanie predpovedaných a skutočných hodnôt pre SARIMA(1, 0, 0)(0, 0, 1)24

3.4 SARMA(1,0)(1,0)24

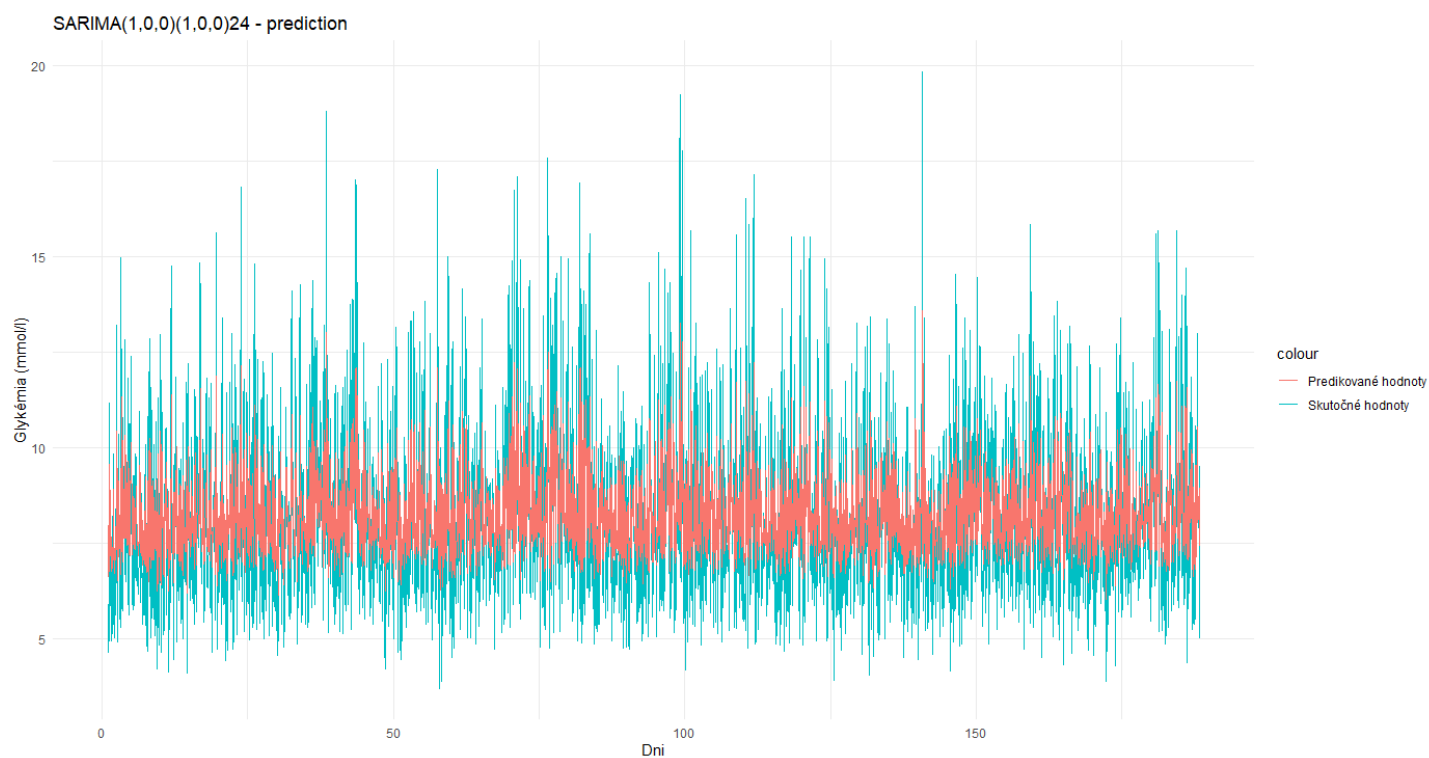
Stále tie výsledky nie sú dramaticky lepšie od predchádzajúcich modelov. Vyskúšam ešte takúto jednoduchú kombináciu parametrov:

- AR(1) – autokoreláciu v rámci dňa (hodinu-dve dozadu)
- AR(1) – pre sezónnu autokoreláciu medzi dňami
- $\text{mean}(\text{ts_data}) = 8.24$

	AR(1)	AR(2)	SARMA(1,0)(0,1)24	SARMA(1,0)(1,0)24
intercept	8.2461	8.2462	8,2461	8,2446
ar1	0.4525	0.5233	0,4503	0,4499
ar2		-0.1568		
sar1				0,1021
sma1			0,0966	
Q*	333.03	236.6	261.56	256.76
Ljun-Box test	2.2e-16	2.2e-16	2.2e-16	2.2e-16
AIC	18 927.49	18 817.62	18 885.25	18 882.58



Graf 10 Reziuá pre SARIMA(1,0,0)(1,0,0)24 v porovnaní s predchádzajúcim modelom



Graf 11 Porovnanie predpovedaných a skutočných hodnôt pre SARIMA(1, 0, 0)(1, 0, 0)24

3.5 Navrhnutý model – SARMA(2,3)(1,1)₂₄

Po viacerých manuálnych pokusoch vyskúšať rôzne koeficienty modelu som použila automatický návrh koeficientov pre model SARIMA.

Táto funkcia navrhla takýto model:

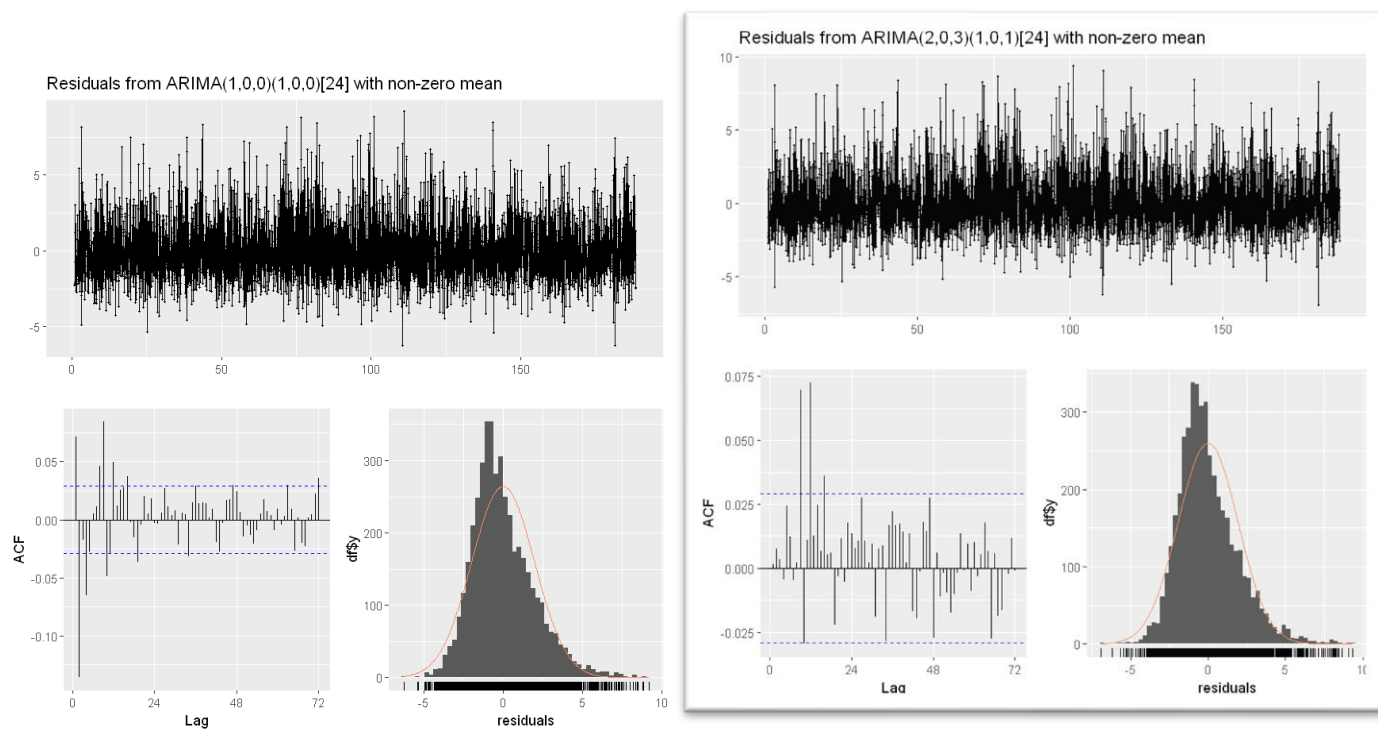
- ARIMA(2,0,3)(1,0,1)₂₄ with non-zero mean
- vidím, že tento model nenavrhne diferencovanie $d = 0$ aj $D = 0$. To sedí s výsledkom P-P testu, ktorý potvrdil stacionaritu radu.

Navrhnuté koeficienty som použila a porovnávala výsledky s predchádzajúcimi modelmi.

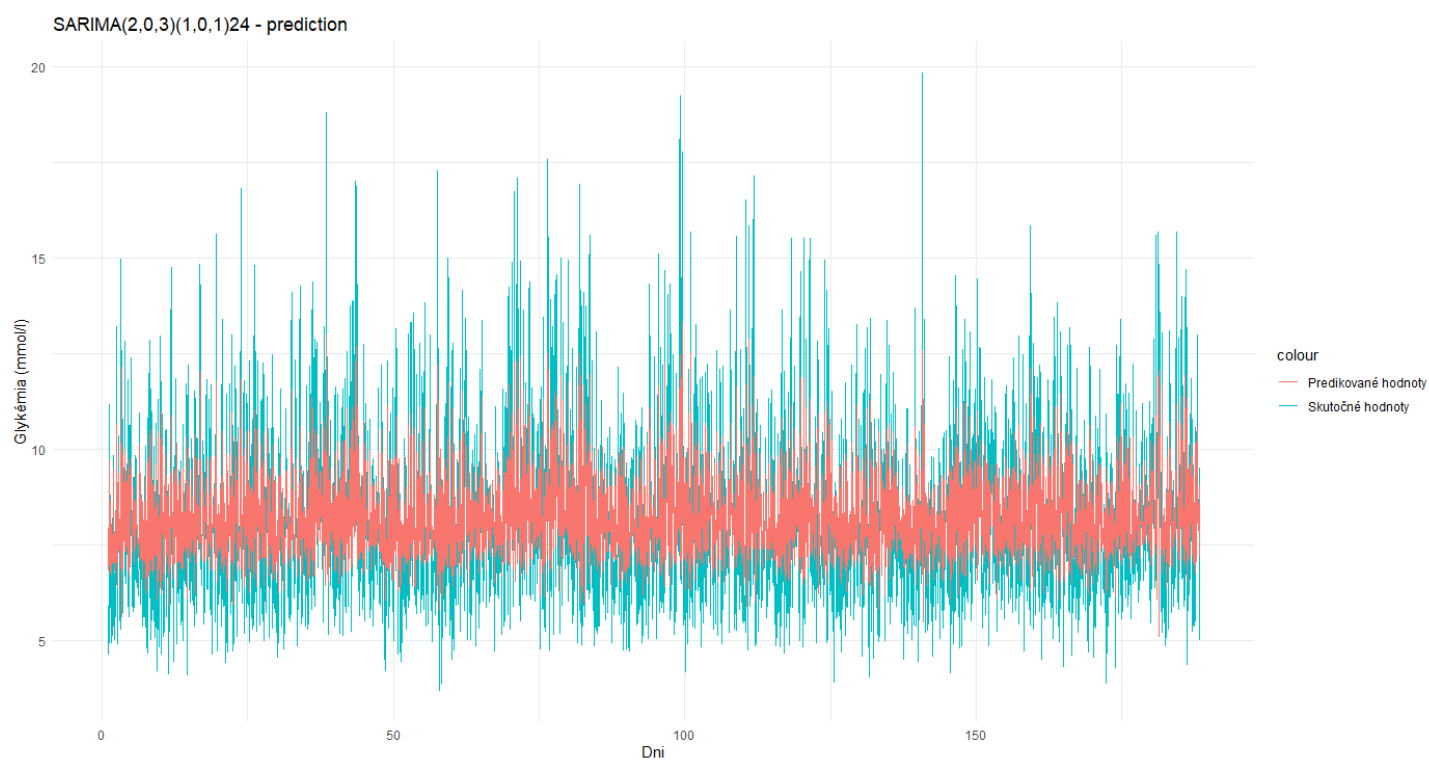
Pre porovnanie: priemer časového radu = 8.24

	AR(1)	AR(2)	SARMA(1,0)(0,1) ₂₄	SARMA(1,0)(1,0) ₂₄	SARMA(2,3)(1,1) ₂₄
intercept	8.25	8.25	8,25	8,24	8,24
ar1	0.45	0.52	0,45	0,45	1,10
ar2		-0.16			-0,81
ma1					-0,57
ma2					0,31
ma3					0,36
sar1				0,10	0,64
sma1			0,10		-0,55
Q*	333.03	236.6	261.56	256.76	100.53
Ljun-Box test	2.2e-16	2.2e-16	2.2e-16	2.2e-16	6.553e-07
AIC	18 927.49	18 817.62	18 885.25	18 882.58	18723.38

Vyzerá to, že tento model je lepší ako predchádzajúci. Štatistika Q^* je oveľa nižšia (100.53 oproti 261.56), teda menej nezachytenej autokorelácie v reziduách.



Graf 12 Reziduá pre navrhnutý SARIMA(2,0,3)(1, 0, 1)24 v porovnaní s predchádzajúcim modelom



Graf 13 Porovnanie predpovedaných a skutočných hodnôt pre SARIMA(2, 0, 3)(1, 0, 1)24

4 Výsledky – interpretácia

Stacionarita modelu v kontexte domény implikuje, že priemerná glykémia sa v čase nemení. Teda v reči diabetu sa priemerná kompenzácia ani nezlepšuje, ani nezhoršuje.

Koeficienty sarima modelu¹:

- intercept = 8,24: priemerná hodnota glykémie
- **ar1** = koeficient autoregresie pre lag 1: **109,8%** zmeny glykémie v prechádzajúcej hodine sa prejaví aj v aktuálnej hodnote
- **ar2** = koeficient autoregresie pre lag 2: **-80,74%** zmeny glykémie pred 2 hodinami sa prejaví v aktuálnej hodnote s v opačnom smere
- **ma1** = koeficient MA pre lag 1: **-57,34%** z chyby predikcie (krátkodobej fluktuácie) v predch. hodine
- **ma2** = koeficient MA pre lag 2: **+31,16%** z chyby predikcie pred 2 hodinami
- **ma3** = koeficient MA pre lag 2: **+35,71%** z chyby predikcie pred 3 hodinami
- **sar1** = sezónny koeficient autoregresie pre lag 24: **63,73%** zmeny glykémie pred 24 hodinami sa prejaví na aktuálnej glykémii
- **sma1** = sezónny koeficient MA pre ho ma pre lag 24; **-55,18%** z chyby predikcie pred 24 hodinami sa prejaví na aktuálnej glykémii

intercept	8,24
ar1	1,0980
ar2	-0,8074
ma1	-0,5734
ma2	0,3116
ma3	0,3571
sar1	0,6373
sma1	-0,5518

Z porovnaní skutočných a predikovaných hodnôt vidíme, že model zachytáva celkový trend a sezónnosť v dátach, ale predikované hodnoty majú menší rozptyl ako skutočné data. Vyzerá to, že model nedokáže zachytiť všetky fluktuácie a variabilitu v dátach.

Toto zistenie zrejme dosť dobre koreluje s realitou bežného života diabetikov, že krivka glykémie dosť variuje a dosiahnuť vyrovnané hodnoty je pomerne ťažké. SARIMA model je asi na tento rad dosť zjednodušený, keďže neberie do úvahy rôzne iné faktory, ktoré vplývajú na glykémiu, ako napríklad pohyb, jedlo, stres apod. A najmä zrejme vplyv samotného uzavretého okruhu, ktorý je nastavený v zásade vždy tak, že pri poklese glykémie pod stanovenú hranicu znižuje podávanie inzulínu a pri výstupe glykémie nad stanovenú hranicu naopak inzulín pridáva. Reaguje teda práve na tú variabilitu dát.

¹ v porovnácej tabuľke boli uvedené zaokrúhlené hodnoty na 2 desatinné miesta

Rád MA procesu: Pôvodne som očakávala len AR model, resp. sezónny AR model, ale závislosti glykémii v čase sú zrejme zložitejšie. Aplikovaním na bežnú skúsenosť môžem potvrdiť, že glykémia v čase môže niekedy reagovať aj napríklad na prekonanú hypoglykémiu, či hyperkorekciu a prudký pokles glykémie práve tzv. kontraregulačným mechanizmom aj niekoľko hodín po. Tá závislosť v čase je teda komplexnejšia a MA proces ju zrejme dokáže zachytiť lepšie ako samotný AR proces. MA proces až do tretieho rádu, teda vplyv krátkodobých fluktuácií až 3-4 hodiny dáva v tomto kontexte zmysel.

5 Záver

Výsledok o stacionarite dát ma mierne prekvapil. To vlastne znamená, že kompenzácia sa po nastavení na uzavretý okruh u tohto pacienta nezlepšila. Laicky by som očakávala zlepšenie. Taktiež v grafe nevidno dopad intervencií = zmenu nastavení parametrov. To by sa malo prejaviť zmenou priemernej hodnoty alebo variability. Žiaden takýto zásah v grafe nevidno.

Pôvodný predpoklad o závislosti glykémie na glykémii hodinu dozadu sa potvrdil. Taktiež sa potvrdila istá sezónnosť (každých 24 hodín). Je pre mňa zaujímavé, že sa ukázala aj istá náhodnosť / fluktuácia hodnôt počas dňa, ktorý model dokázal zachytiť, hoci nie úplne.