

# Model napovedi odjema električne energije

## Matematika z računalnikom 2023/24

Karolina Šavli

Maj 2024

### 1 Uvod

Cilj projektne naloge je sestaviti model, ki bo napovedal odjem električne energije za celotni naslenji dan (za naslednjih 24 ur).

#### 1.1 Teoretična osnova

V projektni nalogi se bomo ukvarjali s časovno vrsto  $\{X_t\}_t$ , torej nekim zaporedjem podatkov. Celotna analiza in napovedovanje sta izvedena v programskem jeziku Python. Zaradi lažjega razumevanja vsebine si najprej pogledjmo nekaj pomembnih konceptov v okviru časovnih vrst.

Pomemben cilj pri sestavljanju modelov je transformacija časovne vrste v stacionarno. Časovna vrsta je stacionarna, če ima podobne statistične značilnosti kot vsaka časovno zamaknjena časovna vrsta. Ne smemo torej imeti trenda ali sezone. Stacionarno lahko preverimo z različnimi formalnimi testi, kot so na primer Augmented Dickey-Fuller (ADF) test, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test in Phillips-Perron test. [3, 1] Pogosti pristopi, da pridemo do stacionarne časovne vrste so logaritmiranje časovne vrste, diferenciranje, sezonsko diferenciranje, ... Pri tem si pogosto pomagamo z obravnavo avtokorelacijske funkcije (ACF) in parcialno avtokorelacijske funkcije (PACF). Ti funkciji sta pomembni, saj opisujeta korelacijo med podatkovnimi točkami v časovni vrsti in prejšnjimi vrednostmi časovne vrste (za različne dolžine zamika). Iz teh funkciji lahko razberemo tudi, če imamo sezonsko komponento. ACF meri linearno razmerje med opazovanjem in njegovimi prejšnjimi opazovanji pri različnih zamikih. PACF pa neposredno meri linearno razmerje med opazovanjem in njegovimi prejšnjimi opazovanji ob določenem zamiku, pri čemer so izključeni prispevki iz vmesnih zamikov.

Ker bomo opravka s časovno vrsto, katere lastnosti se s časom ne spreminjajo, bomo obravnavali model ARMA (ang. *autoregressive moving average process*), ki velja za enega najbolj zanimivih modelov, ki predpostavlja stacionarno vrsto.

Ukvarjali se bomo z modeli družine ARMA in GARCH.

## 2 Osnovna analiza časovne vrste

Podjetje GEN-I je v obliki excel razpredelnice pripravilo tabelo podatkov, sestavljeno iz sedmih stolpcev:

- `DateTimeStartUTC`: univerzalni koordinirani čas,
- `DateTimeStartCET`: srednjeevropski čas,
- `Odjem ACT`: neto odjem električne energije v kWh,
- `Temperatura ACT`: dejanska temperatura,
- `Temperatura FC`: s strani GEN-I napovedana temperatura,
- `Sevanje ACT`: dejansko sevanje in
- `Sevanje FC`: s strani GEN-I napovedano sevanje.

Slika 1: Tabela podatkov, 2021-2024

	<code>DateTimeStartUTC</code>	<code>DateTimeStartCET</code>	<code>Odjem ACT</code>	<code>Temperatura ACT</code>	<code>Temperatura FC</code>	<code>Sevanje ACT</code>	<code>Sevanje FC</code>
0	31.10.2021 23:00	2021-01-11 00:00:00	0.000010	3.60	5.300	0.0	0.0
1	31.10.2021 23:15	2021-01-11 00:15:00	0.000009	3.60	5.300	0.0	0.0
2	31.10.2021 23:30	2021-01-11 00:30:00	0.000009	3.60	5.300	0.0	0.0
3	31.10.2021 23:45	2021-01-11 00:45:00	0.000009	3.60	5.300	0.0	0.0
4	1.11.2021 00:00	2021-01-11 01:00:00	0.000008	3.45	5.300	0.0	0.0
...	...	...	...	...	...	...	...
80059	12.02.2024 21:45	2024-12-02 22:45:00	0.000012	8.00	3.375	0.0	0.0
80060	12.02.2024 22:00	2024-12-02 23:00:00	0.000012	7.90	3.200	0.0	0.0
80061	12.02.2024 22:15	2024-12-02 23:15:00	0.000011	7.80	3.300	0.0	0.0
80062	12.02.2024 22:30	2024-12-02 23:30:00	0.000011	7.15	3.400	0.0	0.0
80063	12.02.2024 22:45	2024-12-02 23:45:00	0.000009	6.50	3.500	0.0	0.0

V analizi sem uporabljala vse stolpce, razen stolpca `DateTimeStartUTC`, saj je v okviru časa bolj relevanten stolpec `DateTimeStartCET`. Podatki so podani za obdobje od 1. novembra 2021 do 12. februarja 2024, na vsakih 15 minut. Tabela ima vsega skupaj 80064 vrstic.

### 2.1 Odjem električne energije

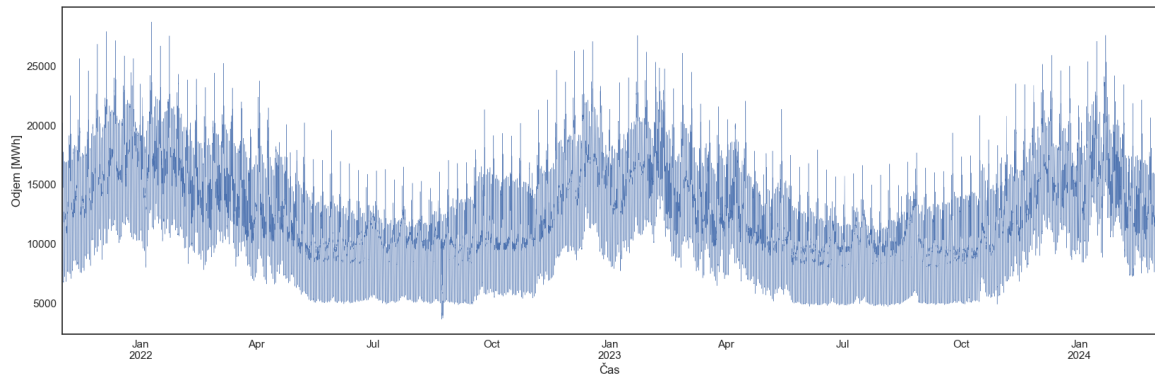
Slika 2 prikazuje odjem električne energije za obdobje od 1. novembra 2021 do 12. februarja 2024. Opazna je sezonskost; odjem je znatno večji jeseni in pozimi, zaradi povečane uporabe energije za ogrevanje in razsvetljevo (saj se število ur dnevne svetlobe podaljša).

S Tabele 1 preberemo, da je povprečna poraba električne energije gospodinjskih odjemalcev okrog 12240,53 MWh, minimalna dosežena vrednost je 3629,32 MWh, maksimalna pa 28736,80 MWh. Vrednosti varirajo za okrog 4167,98 MWh.

Tabela 1: Opisne statistike porabe električne energije, 2021-2024

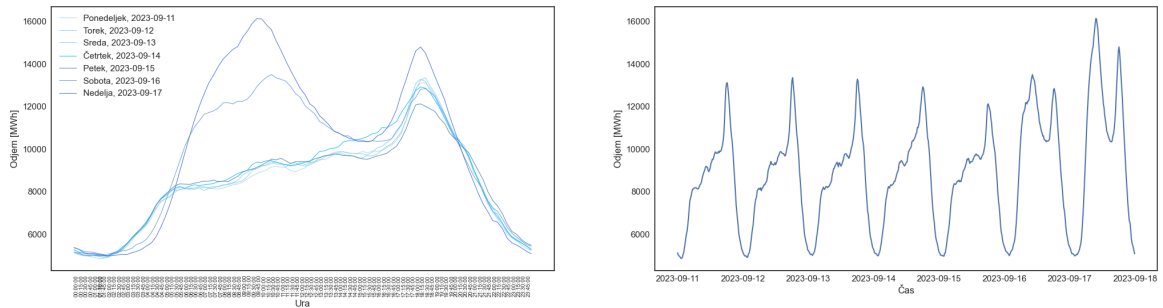
	Min	Max	Povprečje	Mediana	Standardni odklon
Odjem [MWh]	3629,32	28736,80	12240,53	11708,50	4167,98

Slika 2: Odjem električne energije, 2021-2024



Odjem električne energije si pogledjmo še na ravni tedna. Opazimo, da je od ponedeljka do petka odjem najmanjši ponoči in se večja vse do viška okrog 18 ure, nato pa hitro pade. V soboto in nedeljo pa je prvi višek porabe dopoldne, drugi pa okrog 18 ure, kar je opazno s Slike 3, ki prikazuje odjem električne energije v drugem tednu septembra 2023. Imamo torej sezonskost na dnevni ravni.

Slika 3: Odjem električne energije po urah, drugi teden septembra 2023

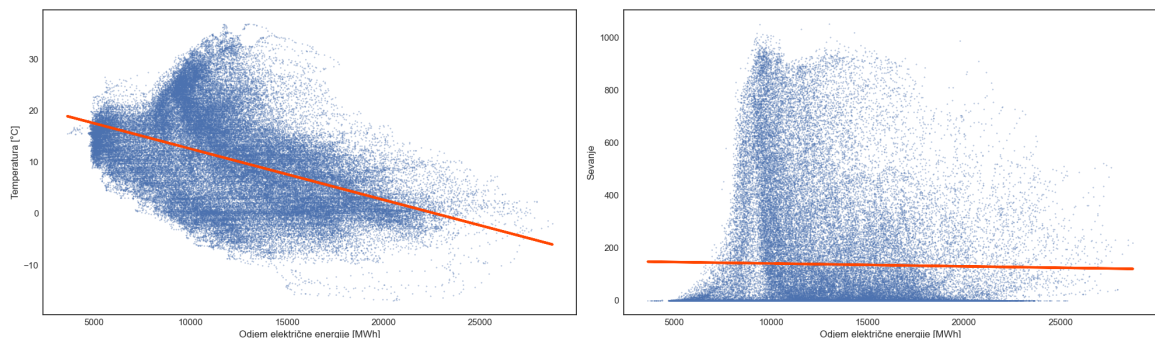


Zaključimo lahko, da je naša časovna vrsta visokofrekvenčna, ima sezonsko komponentno ter njeno povprečje ni konstantno.

## 2.2 Povezava med odjemom in temperaturo ter sevanjem

Smislno se zdi, da obstaja povezava med odjemom električne energije ter temperaturo in sevanjem. S pomočjo Slike 4 ugotovimo, da nižja temperatura pomeni višji odjem in obratno. Povezave med sevanjem in odjemom PA ni opaziti; sklepam da zato, ker ne analiziramo samooskrbnih odjemalcev (tisti ki imajo svojo sončno elektrarno).

Slika 4: Povezava med odjemom in temperaturo ter sevanjem, 2021-2024



V nadaljevanju bomo v upanju boljše napovedi v naše modele kot eksogeni spremenljivki vključili tudi temperaturo in sevanje.

### 3 Napredna analiza in izbira modela

#### 3.1 Izbira družine modelov

ARIMA je model, ki pri napovedovanju zajame vzorce, trende in sezonskost podatkov s kombinacijo preteklih vrednosti, razlik in napak. Ker pa imamo v podatkih očitno sezonskost, lahko ARIMO nadgradimo v SARIMO, ki dodatno upošteva še sezone vzorce. Ta družina modelov ima težavo predvsem pri napovedih, ko ima časovna vrsta skozi čas spremembo variance. Da bo naše napovedovanje bolj učinkovito, bomo SARIMA model nadgradili v model SARIMA-GARCH, saj se model GARCH uporablja ravno za modeliranje združevanja volatilitnosti v podatke časovnih vrst. [2]

Naš model bo oblike:

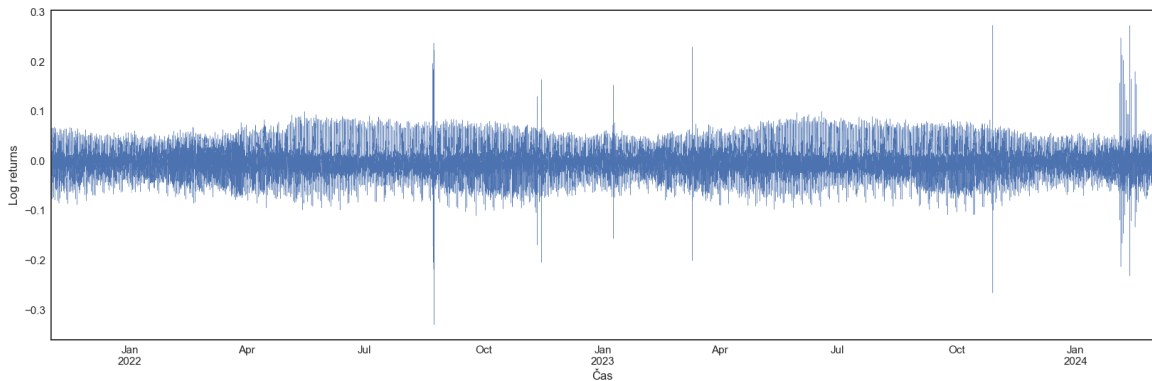
TUKAJ BO FORMULA

#### 3.2 Odstranitev sezonskosti in pridobitev stacionarnosti

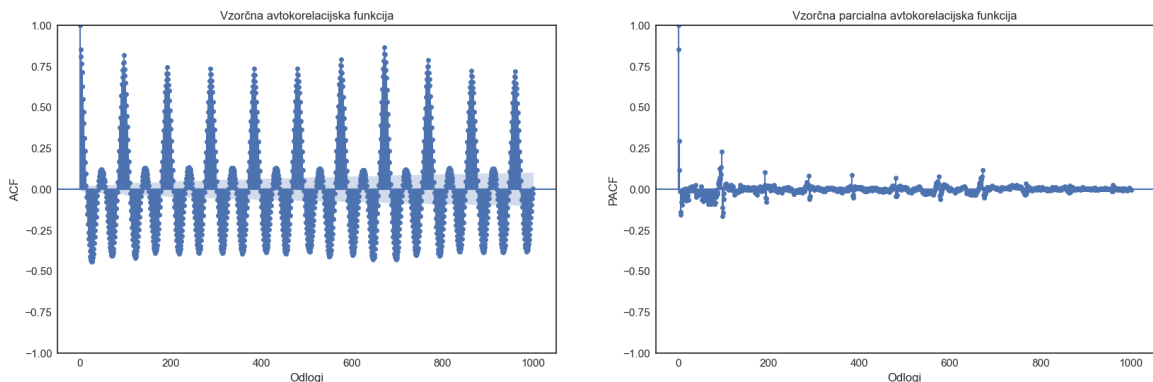
Da bomo lahko indentificirali potencialne modele, moramo najprej originalno časovno vrsto (Slika 2) narediti stacionarno.

Ker so naši podatki volatilni najprej naredimo **logaritmične donose** (ang. *log returns*). Če z  $W_t$  označimo originalno časovno vrsto odjema električne energije, so logaritmični donosi definirani kot časovna vrsta  $Y_t = \ln\left(\frac{W_t}{W_{t-1}}\right)$ . Slednja časovna vrsta je prikazana na Slika 5.

Slika 5: Logaritmični donosi odjema električne energije, 2021-2024



Slika 6: Vzorčna avtokorelacijska in parcialna avtokorelacija funkcija časovne vrste logaritmičnih donosov

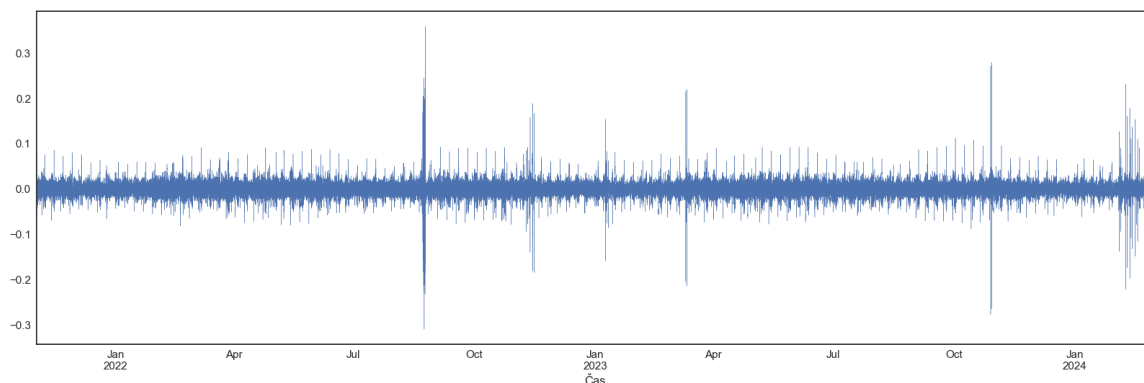


Slika 6 prikazuje ACF (vzorčno avtokorelacijo) in PACF (vzorčna parcialna avtokorelacija) logaritmičnih donosov do odloga 1000 (to je malo manj kot dva tedna). Z grafov je očitno, da časovna vrsta

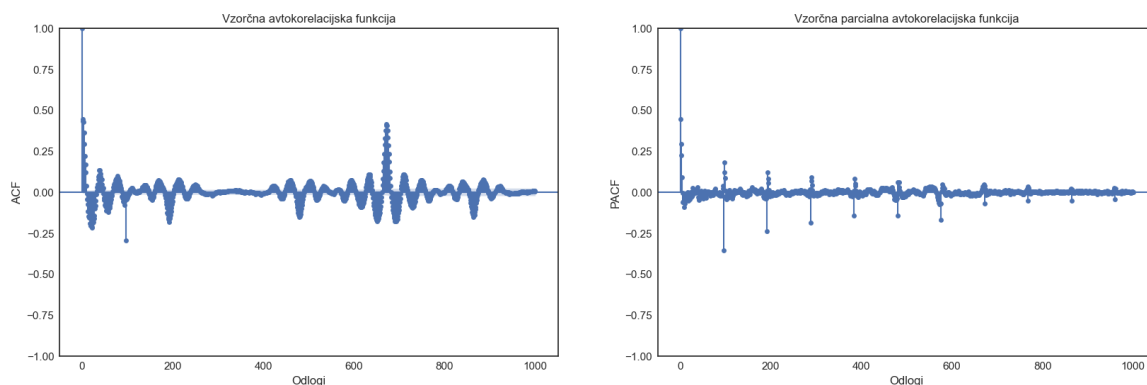
ni stacionarna. Opazna je sezonska komponentna, in sicer 96, kar je ravno en dan <sup>1</sup>.

Da dosežemo stacionarnost, se je potrebno sezonskosti znebiti, zato bomo podatke **sezonsko diferencirali**. Nova časovna vrsta bo  $Z_t = Y_t - Y_{t-96}$ . Prikazana je na Slika 7, njeni ACF in PACF pa na Slika 8.

Slika 7: Časovna vrsta po sezonskem diferenciranju, 2021-2024

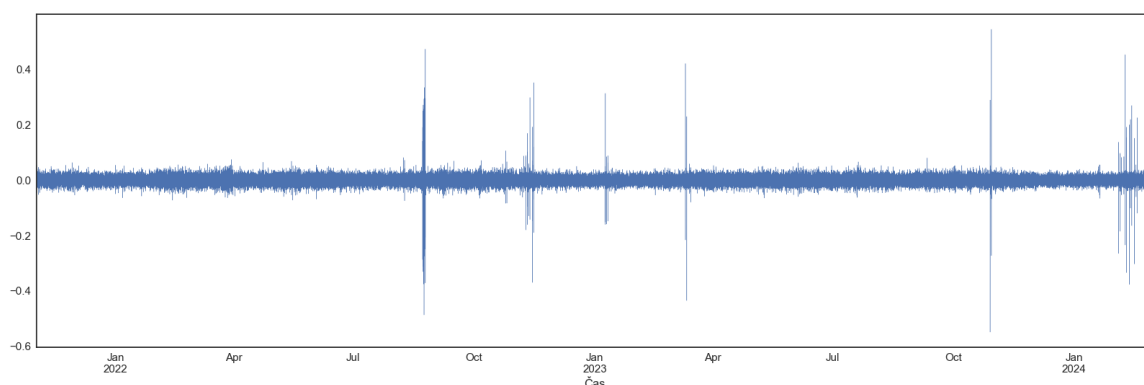


Slika 8: Vzorčna avtokorelacijska in parcialna avtokorelacija funkcija časovne vrste po sezonskem diferenciranju



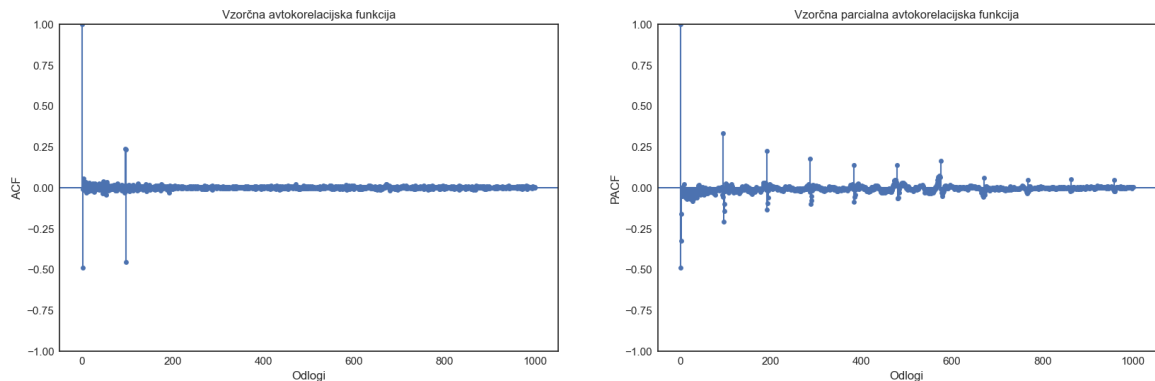
Na grafu ACF opazimo ponavljanje vzorca, kar pomeni, da časovna vrsta  $Z_t$  še kar ni stacionarna. Še enkrat jo diferenciramo in dobimo vrsto  $X_t = Z_t - Z_{t-1}$ . Prikazana je na Slika 9, njeni ACF in PACF pa na Slika 10.

Slika 9: Časovna vrsta po prvem nesezonskem diferenciranju, 2021-2024

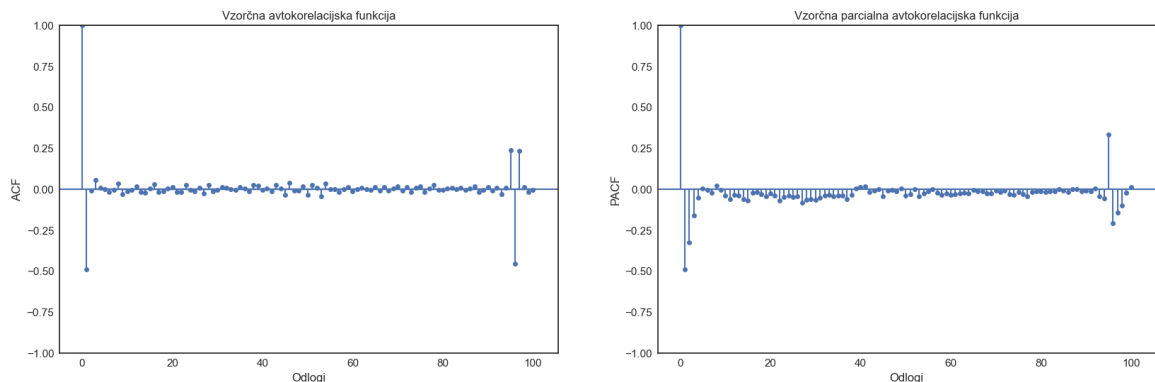


<sup>1</sup>Podatki so podani na 15 minut in  $15 \text{ min} \cdot 96 = 1440 \text{ min}$ , kar je ravno en dan.

Slika 10: Vzorčna avtokorelacijska in parcialna avtokorelacija funkcija časovne vrste po prvem nesezonskem diferenciranju



Slika 11: Vzorčna avtokorelacijska in parcialna avtokorelacija funkcija časovne vrste po prvem nesezonskem diferenciranju, odlogi do 100



Dobljena časovna vrsta  $X_t$  na pogladi ACF in PACF (Slika 10) zgleda stacionarno in tudi formalni testi ADF, KPSS in Phillips-Perron potrdijo njeno stacionarnost. Časovna vrsta je torej primerna za izbero parametrov modela SARIMA.

### 3.3 Identifikacija modela SARIMA

Na podlagi ACF in PACF (Slika 10) stacionarne časovne vrste  $X_t$ , izberimo parametre modela SARIMA(p, d, q)(P, D, Q)[S]. Parametre lahko razdelimo na nesezonske (p, d, q) in sezonske (P, D, Q, S) [4]:

- p ... koliko korakov nazaj bomo gledali (preberemo iz grafa PACF)
- d ... kolikokrat smo diferencirali
- q ... število zgodovinskih napak, ki vplivajo na trenutno vrednost (preberemo iz grafa ACF)
- P ... red sezonskega AR(P)
- D ... kolikokrat smo sezonsko diferencirali
- Q ... red sezonskega MA(Q)
- S ... perioda

Ker smo časovno vrsto enkrat sezonsko diferencirali bo  $D = 1$  in ker smo jo enkrat navadno diferencirali bo  $d = 1$ . Perioda je enaka 96, torej je  $S = 96$ .

Za določitev sezonskih parametrov  $P$  in  $Q$  gledamo korelacije pri odlogih, ki so večkratniki periode  $S$ . Pri PACF visoko korelacijo opazimo predvsem pri 96 in 192, nato pa se z vsako dodatno periodo manjša. Parameter  $P$  je torej 1 ali več; predlagala bi 1, 2 ali 3. Pri ACF pa je občutna korelacija zgolj pri 96, zato izberemo  $Q = 1$ .

Za lažjo določitev nesezonskih parametrov  $p$  in  $q$  bomo gledali ACF in PACF do prve periode (torej do odloga 96). Slednje je prikazano na Slika 11. Tako iz PACF, kot tudi iz ACF, je opazna večja korelacija pri prvih nekaj urah in v uri tik pred periodo. V modelu bomo zato zagotovo vključili prvih nekaj ur, saj le-te kažejo močan vpliv.

Izbrala sem nekaj potencialnih modelov SARIMA, kjer sem kot eksogeno spremenljivko podala temperaturo in sevanje. Modele sem trenirala na 75 % podatkov, to je od 1. novembra 2021 do 4. avgusta 2023. Na preostanku podatkov pa sem izvajala teste napovedi. Vrednosti kriterija AIC za posamezne modele so prikazane v Tabeli 2. Najmanjšo vrednostjo kriterija AIC ima model SARIMA(4, 1, 5)(0, 1, 0)[96].

S Tabele 2 je razvidno, da sem za sezonska parametra  $P$  in  $Q$  v vseh primerih izbrala vrednost 0, čeprav sklepam, da bi bile bolj optimalne višje vrednosti. Te modele sem poskusila zagnati, vendar moj računalnik ni vrnil nobenega rezultata, ker očitno ni dovolj zmogljiv.

Tabela 2: Vrednost kriterija AIC za izbrane modele iz družine SARIMA

Model	AIC
SARIMA(1,1,0)(0,1,0)[96]	-338588,046
SARIMA(0,1,1)(0,1,0)[96]	-346718,519
SARIMA(1,1,1)(0,1,0)[96]	-346768,583
SARIMA(2,1,1)(0,1,0)[96]	-347555,507
SARIMA(3,1,2)(0,1,0)[96]	-347573,394
SARIMA(4,1,3)(0,1,0)[96]	-278981,698
SARIMA(5,1,4)(0,1,0)[96]	-345287,413
SARIMA(5,1,5)(0,1,0)[96]	-345310,105
SARIMA(4,1,5)(0,1,0)[96]	<b>-347619,304</b>
SARIMA(6,1,5)(0,1,0)[96]	-345325,842
SARIMA(6,1,6)(0,1,0)[96]	-345342,972
SARIMA(5,1,6)(0,1,0)[96]	-345351,794

Izbran imamo model SARIMA, v naslednjem koraku pa na podoben način izberimo še model GARCH.

### 3.4 Izbira modela SARIMA-GARCH

Model GARCH( $p$ ,  $q$ ) bomo konstruirali na rezidualih izbranega modela SARIMA. Preizkusili bomo 6 kombinacij parametrov  $p$  in  $q$ , in sicer (1, 1), (1, 2), (2, 1), (2, 2), (1, 3) ter (3, 1). Na podlagi vrednosti kriterija AIC se odločimo za model SARIMA(4, 1, 5)(0, 1, 0)[96]-GARCH(1, 3), saj je iz Tabele 3 razvidno, da ima najmanjši AIC.

Tabela 3: Vrednost kriterija AIC za izbrane modele SARIMA(4, 1, 5)(0, 1, 0)[96]-GARCH( $p$ ,  $q$ )

SARIMA-GARCH( $p$ , $q$ )	AIC
(0, 0)	-347619,304
(1, 1)	-365754,448
(1, 2)	-366056,218
(2, 1)	-365434,577
(2, 2)	-365843,673
(1, 3)	<b>-366134,335</b>
(3, 1)	-365252,576

### 3.5 Izbran model

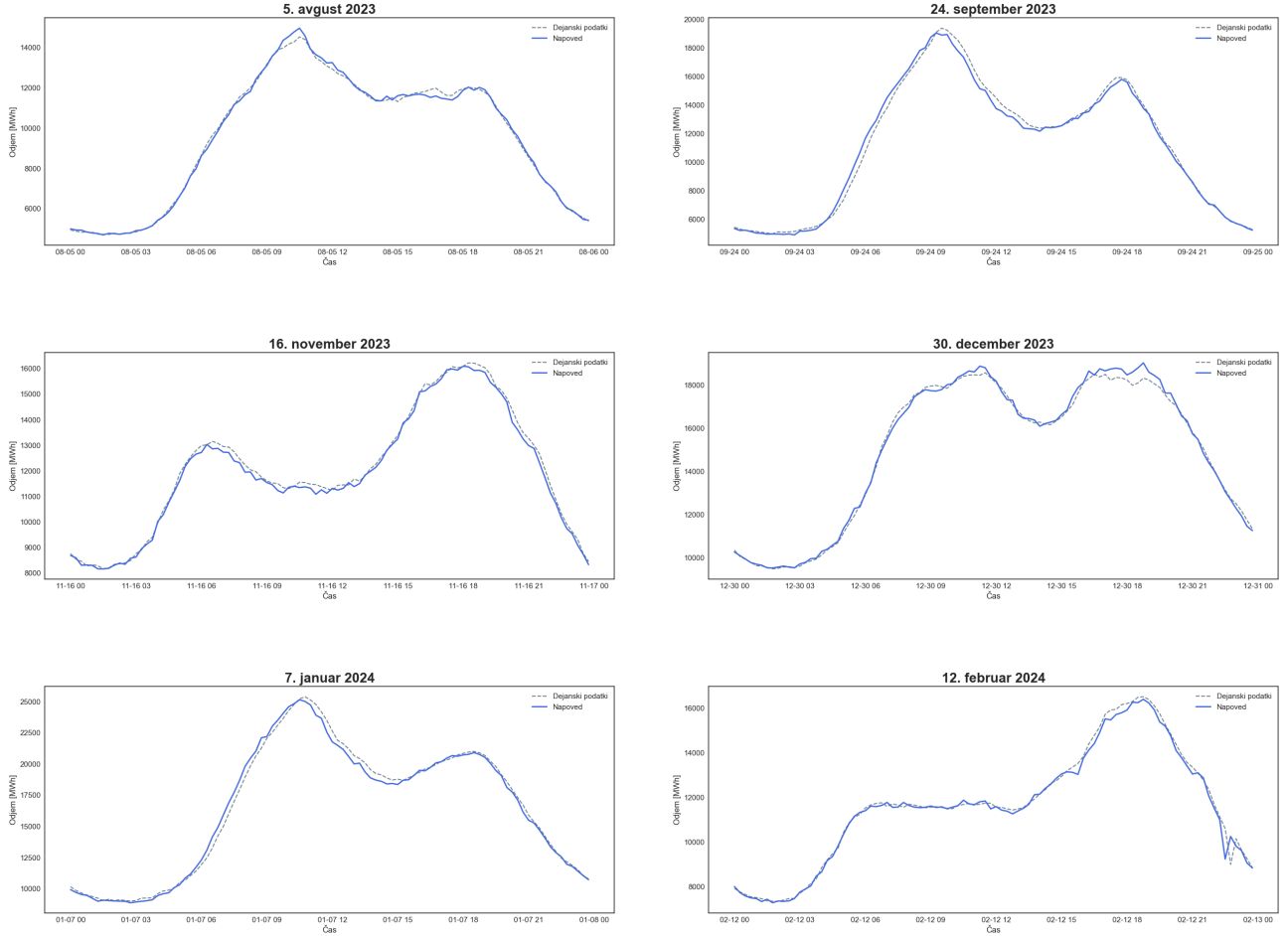
Na podlagi vrednosti kriterija AIC se odločimo za model SARIMA(4, 1, 5)(0, 1, 0)[96]-GARCH(1, 3). V matematičnem jeziku je oblike:



## 4 Testiranje izbranega modela SARIMA-GARCH

Izbran model SARIMA-GARCH uporabimo za napovedovanje odjema električne energije za različne dni v testnem obdobju (od 5. avgusta 2023 do 12. februarja 2024). Izbrala sem si 6 različnih dni, narisala njihove napovedi ter jih prikazala na Sliki 12.

Slika 12: Napovedi odjema električne energije za izbrane datume



Prileganje napovedi dejanski vrednosti se zdi zelo dobra, vseeno pa si pogledjmo napaki RMSE (ang. *Root-mean-square deviation*) in MAPE (ang. *Mean absolute percentage error*), ki ju izračunamo po formulah:

$$RMSE = \sqrt{\frac{1}{96} \sum_{t=1}^{96} (W_t - \hat{W}_t)^2} \quad \text{in} \quad MAPE = \frac{1}{96} \sum_{t=1}^{96} \frac{W_t - \hat{W}_t}{W_t},$$

kjer je  $W_t$  dejanska vrednost odjema,  $\hat{W}_t$  pa napovedana.

Vrednosti napak RMSE in MAPE za izbrane napovedi so zapisane v Tabeli 4 in v povprečju znašajo RMSE 261,926, MAPE pa 1,478 %.

Napaki RMSE in MAPE sem izračunala še posebej za zadnji teden septembra 2023 in prvi teden februarja 2024. Rezultati so prikazani v Tabeli

Tabela 4: Vrednost napake RMSE in MAPE za izbrane datume

Datum	RMSE	MAPE
5.avg.23	156,600	1,104
24.sep.23	365,372	2,392
16.nov.23	176,901	1,199
30.dec.23	224,123	1,046
7.jan.23	399,578	1,839
12.feb.23	248,981	1,287
Povprečje	261,926	1,478

## 5 Zaključek

## Literatura

- [1] K. Brilej. *Napovedovanje cen in povpraševanja po električni energiji: magistrsko delo*. PhD thesis, 2021.
- [2] T. Dierckx. ARIMA-GARCH forecasting with Python. *Medium*, Sept. 2020.
- [3] V. K. G. Statistical Tests to Check Stationarity in Time Series. *Medium*, May 2024.
- [4] B. Köseoğlu. Guide to Time Series Analysis with Python — 4: ARIMA and SARIMA. *Medium*, Jan. 2024.