Predviđanje utroška energije uporabom vremenskih nizova

Ostrunić, Matija

Master's thesis / Diplomski rad

2020

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture / Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:235:054216

Rights / Prava: Attribution 3.0 Unported/Imenovanje 3.0

Download date / Datum preuzimanja: **2024-03-29**

Repository / Repozitorij:

Repository of Faculty of Mechanical Engineering and Naval Architecture University of Zagreb



SVEUČILIŠTE U ZAGREBU
FAKULTET STROJARSTVA I BRODOGRADNJE

DIPLOMSKI RAD

Matija Ostrunić

Zagreb, 2020.

SVEUČILIŠTE U ZAGREBU
FAKULTET STROJARSTVA I BRODOGRADNJE

DIPLOMSKI RAD

Mentor:

Doc. dr. sc. Hrvoje Cajner

Student:

Matija Ostrunić

Zagreb, 2020.

IZJAVA

Izjavljujem da sam ovaj rad izradio samostalno koristeći znanja stečena tijekom studija i navedenu literaturu.

Matija Ostrunić

ZAHVALA

Zahvaljujem svom mentoru Doc. dr. sc. Hrvoju Cajneru na pruženoj stručnoj pomoći tijekom izrade ovog rada.

Posebno zahvaljujem svojoj obitelji, roditeljima Željku i Gordani,
bratu Iliji i djedu Franji
na bezuvjetnoj podršci i razumijevanju tijekom mojeg obrazovanja.

Također, zahvaljujem svim svojim prijateljima
koji su moj neizostavan oslonac i podrška.

Matija Ostrunić

SVEUČILIŠTE U ZAGREBU
FAKULTET STROJARSTVA I BRODOGRADNJE
Središnje povjerenstvo za završne i diplomske ispite
Povjerenstvo za diplomske radove studija strojarstva za smjerove:
proizvodno inženjerstvo, računalno inženjerstvo, industrijsko inženjerstvo i menadžment, inženjerstvo materijala te mehatronika i robotika

Sveučilište u Zagrebu Fakultet strojarstva i brodogradnje	
Datum:	Prilog:
Klasa: 602 - 04 / 20 - 6 / 3	
Ur. broj: 15 - 1703 - 20 -	

DIPLOMSKI ZADATAK

Student: **MATIJA OSTRUNIĆ** Mat. br.: 0035199832

Naslov rada na hrvatskom jeziku: **Predviđanje utroška energije uporabom vremenskih nizova**

Naslov rada na engleskom jeziku: **Prediction of energy consumption using time series models**

Opis zadatka:

Pod pojmom vremenskih nizova podrazumijeva se realizacija nekoga mjerenja ili eksperimenta u vremenskoj domeni tj. određenom vremenskom periodu. Analiza vremenskih nizova primjenom različitih modela baziranih na pokretnim prosjecima daje mogućnost kako lakšeg razumijevanja procesa tako i predviđanja budućih vrijednosti. Aplikacija samih modela, posebice ARMA i ARIMA modela, podrazumijeva da je proces opisan u vremenskoj domeni te da sadrži dovoljnu količinu podataka koji opisuju sve ključne značajke. U sustavima proizvodnje i isporuke energije važno je poznavati potrebe za zahtijevanom količinom energije u nekom budućem vremenskom periodu kako bi se odredili optimalni parametri proizvodnje i dobave. Uporabom historijskih podataka iz baze podataka o utrošku energije na pojedinim objektima te apliciranjem postojećih modela moguće je predvidjeti zahtjeve za energijom. Temeljem takve predikcije moguće je preciznije odrediti optimalne parametre proizvodnje.

U radu je potrebno:
a) Dati pregled metoda i tehnika za modeliranje vremenskih nizova.
b) Prikupiti podatke iz realnog sustava opskrbe energijom.
c) Analizirati podatke dobivene iz realnog sustava koristeći se univarijatnim i multivarijatnim statističkim metodama.
d) Modelirati vremenski niz za potrebe predikcije budućih vrijednosti uzimajući u obzir varijabilnost i specifičnost procesa definiranih utjecajnim varijablama.
e) Dati prijedlog modela za predikciju i korekciju parametara u sustavu regulacije isporuke energije.

U radu je potrebno navesti korištenu literaturu i eventualno dobivenu pomoć.

Zadatak zadan:	Rok predaje rada:	Predviđeni datum obrane:
5. ožujka 2020.	7. svibnja 2020.	11. svibnja do 15. svibnja 2020.

Zadatak zadao: Predsjednica Povjerenstva:

doc. dr. sc. Hrvoje Cajner prof. dr. sc. Biserka Runje

SADRŽAJ

SADRŽAJ I
POPIS SLIKA III
POPIS TABLICA V
POPIS KRATICA VI
SAŽETAK VII
SUMMARY VIII
1. UVOD 1
2. ANALIZA I PREDVIĐANJE VREMENSKIH SERIJA 2
2.1. Analiza vremenskih serija 3
2.2. Proces (postupak) predviđanja vremenskih serija 5
2.2.1. Definiranje problema 5
2.2.2. Prikupljanje podataka 5
2.2.3. Analiza podataka 6
2.2.4. Odabir modela 7
2.2.5. Validacija modela 8
2.2.6. Verifikacija modela – kontrola rada modela 8
3. STATISTIČKE METODE PREDVIĐANJA VREMENSKIH SERIJA 9
3.1. Pojmovi potrebni za razumijevanje metoda predviđanja vremenskih serija 9
3.2. Metode eksponencijalnog zaglađivanja 10
3.2.1. Metoda jednostavnog eksponencijalnog zaglađivanja 11
3.2.2. Holt – Winters metoda 12
3.2.2.1. Aditivna Holt – Winters metoda 12
3.2.2.2. Multiplikativna Holt – Winters metoda 13
3.3. ARIMA metode 13
3.3.1. Autoregresivni (AR) model 14
3.3.2. Model pomičnog prosjeka (MA) 15
3.3.3. Integrirana (I) komponenta 16
3.3.3.1. Diferencijacija 17
3.3.3.2. Autokorelacijske funkcije 19
3.3.4. ARMA, ARIMA i SARIMA model 21
3.3.4.1. Odabir parametara p i q – AIC kriterij 23
3.4. Mjere točnosti modela 23

4. DEFINIRANJE PROBLEMA PREDVIĐANJA 25
5. ANALIZA REALNOG SUSTAVA I PODATAKA 27
5.1. Realni sustav – svojstva i obilježja 27
5.2. Prikupljanje podataka iz realnog sustava 28
5.3. Analiza podataka iz realnog sustava 29
5.3.1. Usporedba lokacija (građevina) 29
5.3.1.1. Usporedba troškova po energentima 30
5.3.1.2. Usporedba geometrijskih podataka lokacija (objekata) 32
5.4. Analiza vremenske serije – lokacija L1 / toplinska energija 35
5.4.1. Dekompozicija vremenske serije 37
5.4.2. Analiza varijance – ANOVA 39
5.4.3. Korelacijska analiza 42
6. PREDIKCIJSKI MODELI 44
6.1. Model eksponencijalnog zaglađivanja 44
6.1.1. Izrada modela eksponencijalnog zaglađivanja 44
6.1.2. Validacija modela 47
6.2. SARIMA model 49
6.2.1. Izrada SARIMA modela 49
6.2.2. Validacija modela 54
6.2.2.1. Provjera rezidualnih vrijednosti modela 54
6.2.2.2. Valdiacija modela – testni skup 56
6.2.3. Analiza rezultata i prijedlog modela 57
7. ZAKLJUČAK 59
LITERATURA 60

POPIS SLIKA

Slika 2.1 Grafički prikaz vremenske serije [1] 2
Slika 2.2 Dekompozicija vremenske serije [2] 4
Slika 2.3 Proces modeliranja predviđanja 5
Slika 2.4. Primjeri različitih obrazaca ponašanja (pojava) [4] 7
Slika 2.5. Različiti prilagođeni modeli [5] 7
Slika 2.6 Podjela podataka na skupove 8
Slika 3.1 Prikaz rezidualnih vrijednosti 10
Slika 3.2 Graf odnosa vremena i faktora težine za jedan podatak (mjerenje) 11
Slika 3.3 Usporedba AR (1) i AR (2) modela [4] 14
Slika 3.4 Usporedba MA (1) i MA (2) modela [4] 16
Slika 3.5 Stacionarnost vremenske serije i ARIMA modeli 17
Slika 3.6 Grafički prikaz vremenske serije prije i nakon diferencijacije 18
Slika 3.7 Korelogrami – ACF i PACF [7] 20
Slika 3.8 Veza između autokorelacijskih funkcija i parametara p i q 20
Slika 3.9 Komponenta sezonalnosti u vremenskoj seriji [8] 22
Slika 4.1 Hodogram aktivnosti u procesu predviđanja budućih stanja 26
Slika 5.1 Shema realnog sustava 27
Slika 5.2 Mjerni uređaji za praćenje potrošnje energije i drugih parametara 28
Slika 5.3 Grafovi ukupnog troška energenata po lokacijama 30
Slika 5.4 Potrošnja toplinske i električne energije tijekom jedne godine (L1) 31
Slika 5.5 Broj etaža i ukupna bruto površina po lokacijama 32
Slika 5.6 Struktura potrošača električne energije i toplinske energije na lokaciji L1 33
Slika 5.7 Odnos *UBP* i A_k po lokacijama 34
Slika 5.8 Usporedba vremenske serije s obzirom na frekvenciju mjerenja 36
Slika 5.9 Dekompozicija vremenske serije (mjesečna potrošnja) 38
Slika 5.10 Provjera normalnosti podataka 39
Slika 5.11 Razlika između godina po potrošnji toplinske energije – bez perioda lipanj-rujan 40
Slika 5.12 Razlika između godina po potrošnji toplinske energije – s periodom lipanj-rujan 41
Slika 5.13 Vremenske serije potrošnje toplinske energije i vanjske temperature 42
Slika 5.14 Korelacijska analiza – rezultati 43
Slika 6.1 Model: $\alpha = \beta = \gamma = 0{,}1$ 45
Slika 6.2 Model: $\alpha = 0{,}9$; $\beta = 0{,}05$; $\gamma = 0{,}05$ 46

Slika 6.3 Validacija optimiziranog modela 47
Slika 6.4 Analiza apsolutnih grešaka modela 48
Slika 6.5 Diferencijacija mjerenja – uklanjanje trenda i sezonalnosti 49
Slika 6.6 Usporedba ACF dijagrama 50
Slika 6.7 Usporedba PACF dijagrama 51
Slika 6.8 Optimizacija parametara modela 52
Slika 6.9 SARIMA model $(3, 1, 3)(0, 1, 1)_{12}$ 53
Slika 6.10 Provjera autokorelacijskih funkcija rezidualnih vrijednosti 54
Slika 6.11 Provjera normalnosti rezidualnih vrijednosti 55
Slika 6.12 Validacija SARIMA modela 56
Slika 6.13 Analiza apsolutnih grešaka modela 56
Slika 6.14 Usporedba rezultata kreiranih modela 58

POPIS TABLICA

Tablica 3.1. Diferencijacija vremenske serije - podaci 18
Tablica 3.2. Popis parametara ARIMA modela 22
Tablica 4.1 Definicija problema i cilja predviđanja 25
Tablica 5.1 Faze analize podataka iz realnog sustava 29
Tablica 5.2 Popis energenata i mjernih jedinica 29
Tablica 5.3 Definicije korištenih fizikalnih veličina [11] 32
Tablica 5.4 Odabrana kombinacija za analizu i predikcijski model 35
Tablica 5.5 Osnovni podaci odabrane vremenske serije 35
Tablica 6.1 Skup podataka (mjerenja) za učenje modela 44
Tablica 6.2 Rezultati predviđanja potrošnje toplinske energije 57

POPIS KRATICA

Kratica	**Značenje kratice**
ACF	Funkcija autokorelacije (eng. *Autocorrelation Function*)
AIC	Akaikeov informacijski kriterij (eng. *Akaike's Information Criterion*)
ANOVA	Analiza varijanci (eng. *Analysis of Variance*)
AR	Autoregresivni model (eng. *Autoregressive model*)
ARIMA	Autoregresivna integrirana metoda pomičnih prosjeka (eng. *Autoregressive Integrated Moving Average*)
ETL	Ekstrakcija, transformacija, učitavanja (eng. *Extract, Transform, Load*)
MA	Model pomičnih prosjeka (eng. *Moving Average model*)
MAE	Prosječno apsolutno odstupanje (eng. *Mean Absolute Error*)
MAPE	Prosječno postotno apsolutno odstupanje (eng. *Mean Absolute Percentage Error*)
PACF	Funkcija parcijalne autokorelacije (eng. *Partial Autocorrelation Function*)
RMSE	Korijen srednjeg kvadratnog odstupanja (eng. *Root Mean Square Error*)
SARIMA	Sezonalna autoregresivna integrirana metoda pomičnih prosjeka (eng. *Seasonal Autoregressive Integrated Moving Average*)
UBP	Ukupna bruto površina

SAŽETAK

Tema ovog rada je predviđanje utroška energije uporabom vremenskih nizova (serija). Vremenski niz ili serija jest slijed izmjerenih vrijednosti promatrane pojave realiziran u vremenskoj domeni tj. određenom vremenskom periodu. Cilj rada je izraditi predikcijski model pomoću kojeg će se moći predvidjeti buduća potrošnja energije, te iskoristiti tako generirano znanje za donošenje ispravnih odluka.

U prvom dijelu rada ukratko su opisani fundamentalni pojmovi vezani uz vremenske nizove, te sam proces nastanka predikcijskog modela temeljenog na vremenskom nizu. Nakon upoznavanja s vremenskim nizovima, dan je prikaz najčešće korištenih statističkih metoda za predviđanje budućih stanja vremenskih nizova. U nastavku rada precizno je definiran glavni cilj predviđanja. U glavnom dijelu rada analiziran je realni sustav i podaci dobiveni iz njega. Nakon analize podataka slijedi izrada i usporedba predikcijskih modela. Na kraju rada je iznesena usporedba dobivenih rezultata te su prezentirani zaključci.

Ključne riječi: vremenski niz, predviđanje, utrošak energije, ARIMA metoda, metoda eksponencijalnog zaglađivanja

SUMMARY

The scope of this paper is prediction of energy consumption using time series. A time series is a sequence of measured values of an observed phenomenon realized in time domain. The aim of the paper is to create a model that can predict energy consumption and leverage generated knowledge.

The first part of the paper briefly describes the fundamental concepts related to time series and the process of creating a prediction model based on them. After introduction, the most commonly used statistical methods are presented. The following section precisely defines main objective of prediction. Afterwards, physical system and data obtained from it are analyzed. Data analysis is followed by the creation and comparison of prediction models. At the end of the paper, an analysis of the obtained results is presented. Ultimately, conclusions and learnings are shared.

Key words: time series, forecasting, energy consumption, ARIMA, exponential smoothing method

1. UVOD

Potreba za predviđanjem utkana je u ljudsko biće od samih početaka. Ljudi su već tisućama godina fascinirani pojmom predviđanja koji se ponekad smatrao božanskim nadahnućem, a ponekad crnom magijom. Danas znamo da iza svakog predviđanja stoje određeni modeli i zakonitosti iz kojih su ona nastala, te je sigurno da je neka stanja, odnosno događaje, teže predvidjeti od drugih jer imaju više utjecajnih faktora (veću kompleksnost) ili jednostavno jer ne postoji adekvatan set podataka koji je potreban kako bi se generiralo znanje – predviđanje. Općenito, predviđanje je proces donošenja zaključaka (predviđanja) o budućnosti na temelju podataka iz prošlosti i sadašnjosti kombinirano sa znanjem o mogućim budućim događajima koji bi mogli imati utjecaj na predviđanje. Naravno, cilj je dobiti što veću točnost predviđanja.

U današnjem svijetu, napredak tehnologije i suvremeni računalni sustavi zajedno su olakšali prikupljanje i akumuliranje podataka složene strukture u različitim područjima znanosti i industrije. Upravo zbog toga nastaju skupovi podataka (baze podataka) u različitim područjima koji se kasnije koriste za analizu i donošenje zaključaka – predviđanja.

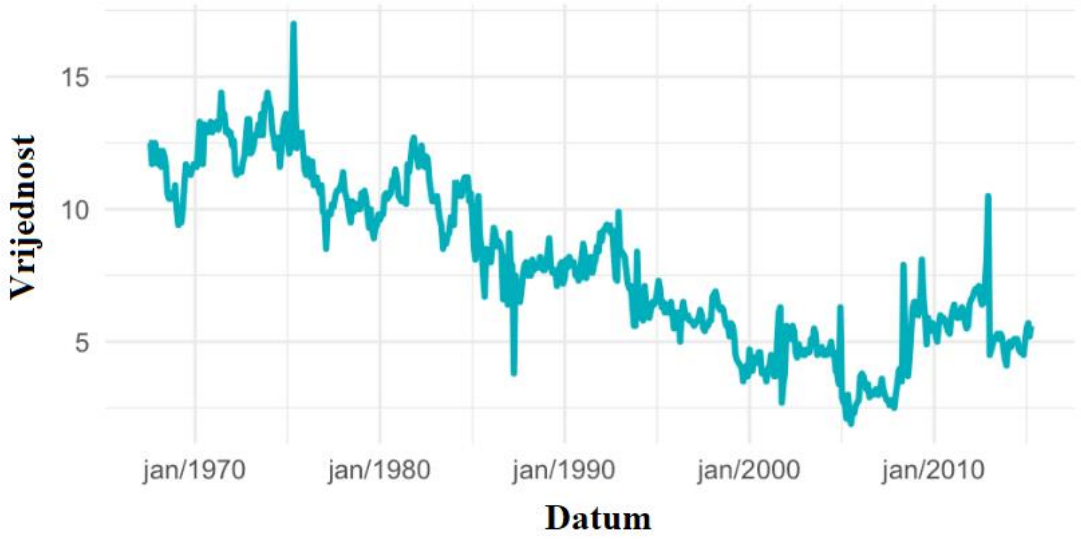
Tržišna utrka nameće tvrtkama razvijanje i implementaciju statističkih modela koji predviđaju buduća stanja. Tako je uspjeh moderne tvrtke usko koreliran s robusnošću i točnosti modela za procjenu, odnosno prognozu budućih stajna. Predviđanje je jedan od glavnih zadataka u poslovanju koji pomaže pri donošenju odluka o proizvodnji, transportu, potrebnoj radnoj snazi (osoblju) i mnogim drugim aktivnostima, te služi kao potpora za dugoročno strateško planiranje.

Analiza vremenski orijentiranih podataka i predviđanje budućih vrijednosti vremenskih serija (eng. *time series*) jedan je od najvažnijih problema s kojima se analitičari susreću u mnogim poljima; od financija i ekonomije do upravljanja proizvodnim i neproizvodnim procesima (operacijama). Većina problema s predviđanjem uključuje upotrebu podataka vremenskih serija koji predstavljaju kronološki niz opažanja promatrane pojave (varijabla koja je u fokusu interesa). U današnjem svijetu uspješna analiza i predviđanje vremenskih serija zahtijeva interakciju analitičara s računalnim softverom jer tehnike i algoritmi jednostavno nisu prikladni za manualno računanje.

U ovom radu bit će prikazan pregled tehnika (algoritama) za predviđanje budućih stanja pomoću vremenski serija kao i sam proces analize istih i generiranih rezultata.

2. ANALIZA I PREDVIĐANJE VREMENSKIH SERIJA

Predviđanje je važan problem koji obuhvaća različita područja kao što su mnogobrojne industrije, poslovanje, ekonomija, znanost, medicina, politika i financije. Većina problema s predviđanjem uključuje upotrebu podataka vremenskih serija. Vremenska serija je kronološki niz opažanja pojave (varijable) koja je u fokusu interesa. Slika 2.1 grafički je prikaz vremenske serije.



Slika 2.1 Grafički prikaz vremenske serije [1]

Metode predviđanja uvelike ovise o podacima koji su dostupni. Prema tome metode predviđanja mogu se podijeliti u dvije grupe:

- kvalitativne metode:
  - podaci nisu dostupni ili nisu relevantni za modeliranje predviđanja,
  - uvelike ovise o iskustvu analitičara (znanja),
  - npr. Delphi metoda,
- kvantitativne metode:
  - numerički podaci su dostupni,
  - pretpostavka da će se obrasci ponašanja nastaviti u budućnosti,
  - npr. Holt-Winters metoda.

2.1. Analiza vremenskih serija

Vremenska serija je slijed podataka promatrane varijable koji su odijeljeni jednakim vremenskim intervalima te kronološki poredani u vremenskom periodu. Podaci vremenskih serija mogu prikazivati različite obrasce u vremenu, a najčešće se vremenski niz može razložiti na nekoliko komponenti, od kojih svaka predstavlja kategoriju „ponašanja“ promatrane pojave.

Primarni zadatak analize vremenskih serija je prikaz, opis i objašnjenje serije podataka, odnosno vrijednosti promatrane varijable. Prvo korak u procesu analize podataka vremenskih serija jest grafički prikaz istih. Grafički prikazi omogućuju vizualizaciju mnogih značajki (osobina) podataka, uključujući obrasce, neuobičajene vrijednosti (eng. *outliers*), promjene tijekom vremena i odnose između zabilježenih podataka (mjerenja).

Dekompozicija vremenske serije vrši se na sljedeće komponente.

- **Trend**

Trend postoji kada je prisutno dugoročno povećanje ili smanjenje vrijednosti promatrane varijable, odnosno trend je uzlazna ili silazna promjena vrijednosti promatrane varijable u vremenu. Trend može biti linearnog, eksponencijalnog, nelinearnog ili nekog drugog oblika.

- **Sezonalnost**

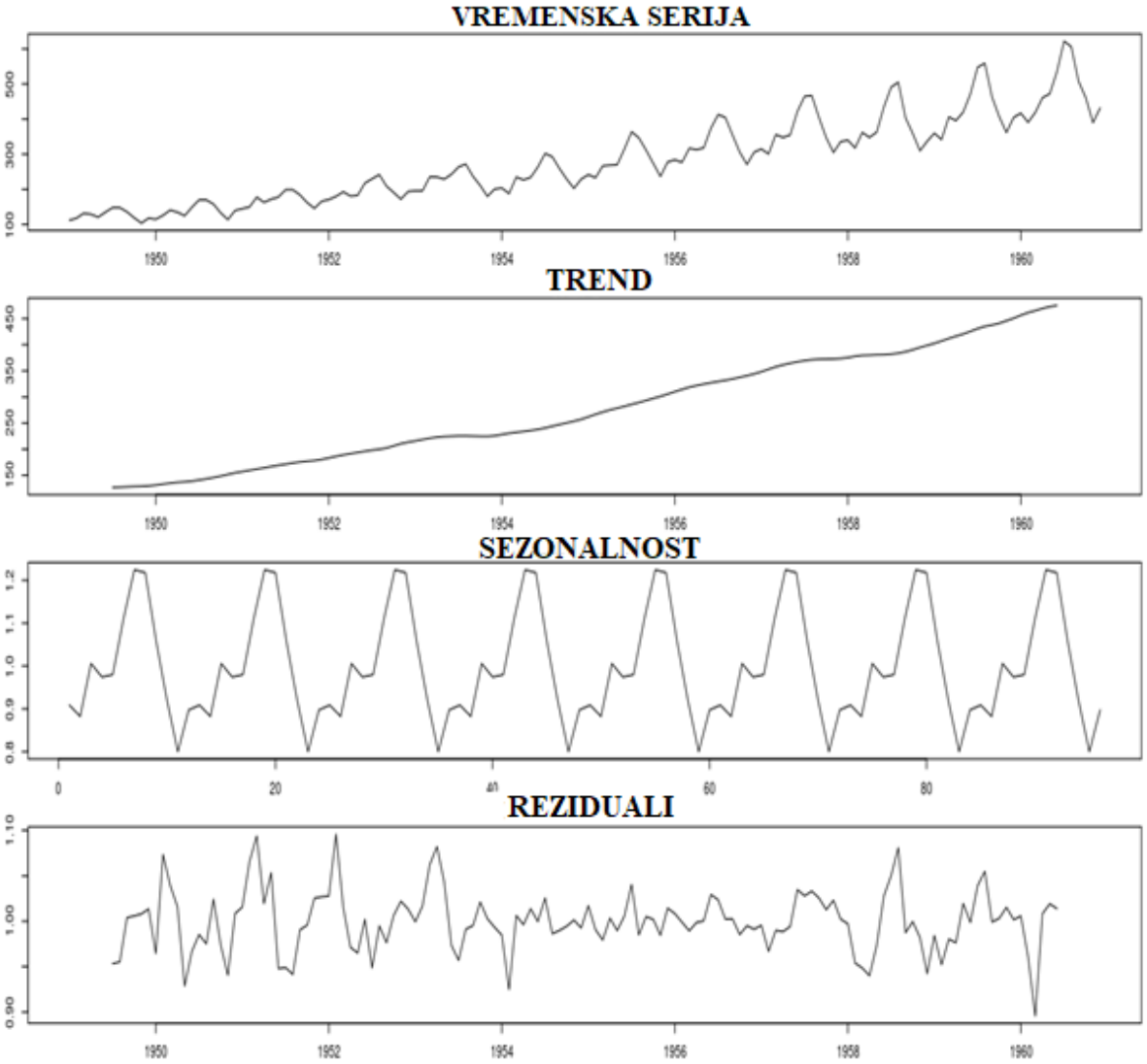
Sezonalnost se javlja kada vremenski niz utječe na vrijednost promatrane varijable. Definira se kao pravilno odstupanje vrijednosti vremenske serije koje se opetovano ponavlja unutar nekog redovitog razdoblja. Sezonalnost je uvijek fiksna i poznata. Sezonski faktor može biti određen specifičnim vremenskim periodom kao što su određeni mjeseci u godini, dani u tjednu, sati u danu i slično.

- **Ciklus**

Pravilno odstupanje vrijednosti vremenske serije koje se događa unutar vremena duljeg od godinu dana. Ciklusi nisu fiksne frekvencije, te se po tome razlikuju od sezonalnosti. Ovisi o specifičnim faktorima koji se javljaju u vremenu (npr. na neku ekonomsko – financijsku pojavu može utjecati pojava recesije).

- **Rezidualna komponenta (šum)**

Nepredvidljive, nasumične promjene u vremenskoj seriji unutar kojih ne postoji pravilnost. Najčešće stvara najveći problem pri modeliranju predviđanja.



Slika 2.2 Dekompozicija vremenske serije [2]

Slika 2.2 prikazuje dekompoziciju vremenske serije. Prikazani set podataka vezan je uz broj putnika internacionalne aviokompanije od 1949. do 1960. godine. Iz samog grafičkog prikaza vidljivo je da postoji uzlazni (pozitivni) trend i sezonalnost.

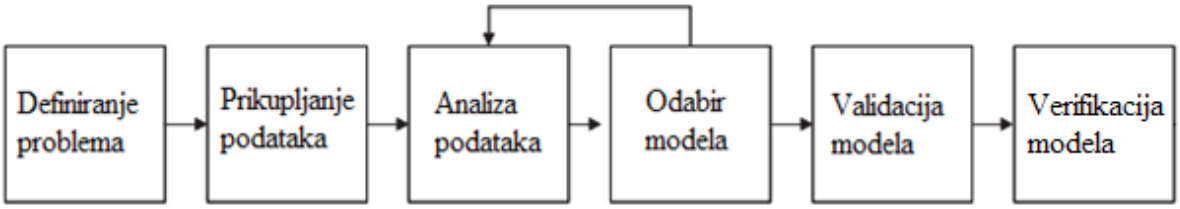
Bitno je napomenuti da dekompozicija može biti različita s obzirom na način na koji se vršila tj. s obzirom na odnos među komponentama. Tako razlikujemo dvije najučestalije dekompozicije vremenskih serija:

- aditivna – nastaje zbrajanjem komponenti,
- multiplikativna – nastaje množenjem komponenti.

2.2. Proces (postupak) predviđanja vremenskih serija

Proces je niz povezanih aktivnosti koje imaju zajednički cilj, a to je najčešće pretvaranje jedenog ili više ulaza u jedan ili više izlaza. Prema [3] aktivnosti u procesu predviđanja su:

1.) definicija problema,
2.) prikupljanje podataka,
3.) analiza podataka,
4.) odabir modela,
5.) validacija modela,
6.) verifikacija modela.



Slika 2.3 Proces modeliranja predviđanja

Slika 2.3 predstavlja shematski prikaz procesa modeliranja predviđanja, a detaljnije će biti objašnjen u nastavku.

2.2.1. Definiranje problema

Definicija problema često predstavlja najteži dio predviđanja. Pažljivo definiranje problema zahtijeva razumijevanje načina na koji će predviđanje biti korišteno i kako se funkcija predviđanja uklapa u širu sliku promatranog problema. Analitičar za definiciju problema potrebne informacije prikuplja komunikacijom sa svima koji su uključeni u prikupljanje podataka, održavanje baza podataka, te budućim korisnicima koji će predviđanja nastala iz kreiranog modela koristiti za buduće planiranje.

2.2.2. Prikupljanje podataka

Prikupljanje podataka predstavlja proces prikupljanja relevantnih povijesnih podataka koji potencijalno imaju utjecaj na varijablu ili varijable koje se predviđaju.

Uvijek su potrebne najmanje dvije vrste podataka:

- statistički podaci – kvantitativni podaci,
- informacije (znanje) osoba koje su zadužene za prikupljanje podataka i budućih korisnika modela za predviđanje.

Frekvencija prikupljanja podataka najčešće je u jednako raspoređenim vremenskim intervalima što je tipično za većinu vremenskih serija. Tako se prikupljanje podataka najčešće vrši u sljedećim intervalima:

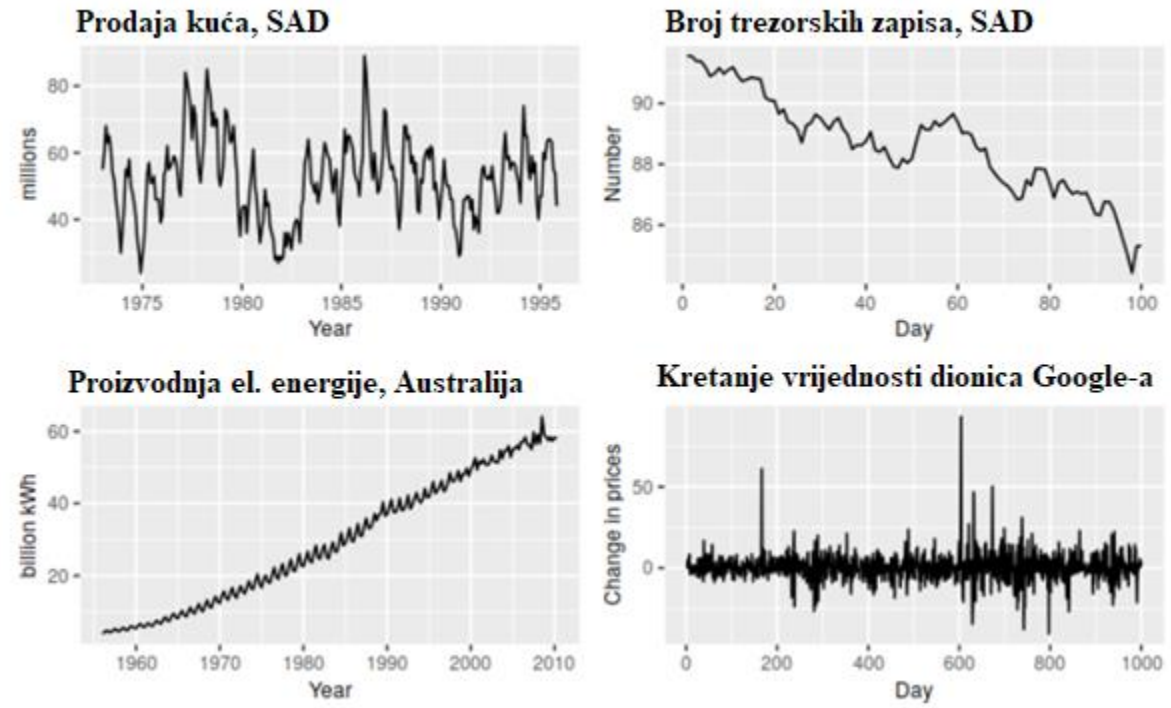
- dnevnom,
- tjednom,
- mjesečnom,
- godišnjem.

Naravno, intervali prikupljanja podataka vremenskih serija mogu biti proizvoljni, a odabir najčešće ovisi o svrsi prikupljanja podataka (kratkoročno ili dugoročno predviđanje). Frekvencija prikupljanja podataka ima direktan utjecaj na bazu podataka koja će biti dostupna pri modeliranju predviđanja ,a samim time i na kreirani model.

2.2.3. Analiza podataka

Analiza podataka važan je preliminarni korak pri odabiru modela predviđanja koji će se koristiti. Dijagrami tj. grafovi podataka vremenske serije trebaju se prikazati i vizualno pregledati kako bi se prepoznale pojave poput trendova, sezonalnosti ili drugih komponenti. Slika 2.4 prikazuje obrasce ponašanja različitih pojava koji se mogu otkriti preliminarnom analizom podataka. U preliminarnu analizu podataka također bi trebalo uključiti deskriptivnu statistiku jer ona daje generalnu sliku o podacima i njihovim svojstvima.

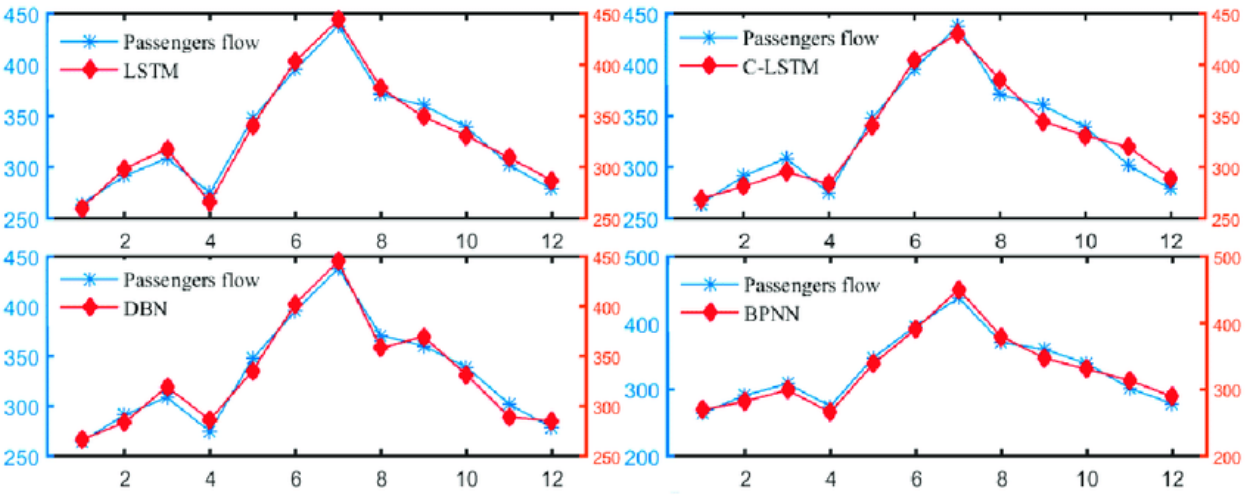
Svrha preliminarne analize je stjecanje osjećaja o podacima koji su dostupni i koji će se koristiti za modeliranje. Također, vrlo je bitna jer se vrlo lako mogu otkriti različiti obrasci kao što su trendovi i sezonalnost. Informacije prikupljene preliminarnom analizom obično će sugerirati početne vrste kvantitativnih metoda i modela predviđanja.



Slika 2.4. Primjeri različitih obrazaca ponašanja (pojava) [4]

2.2.4. Odabir modela

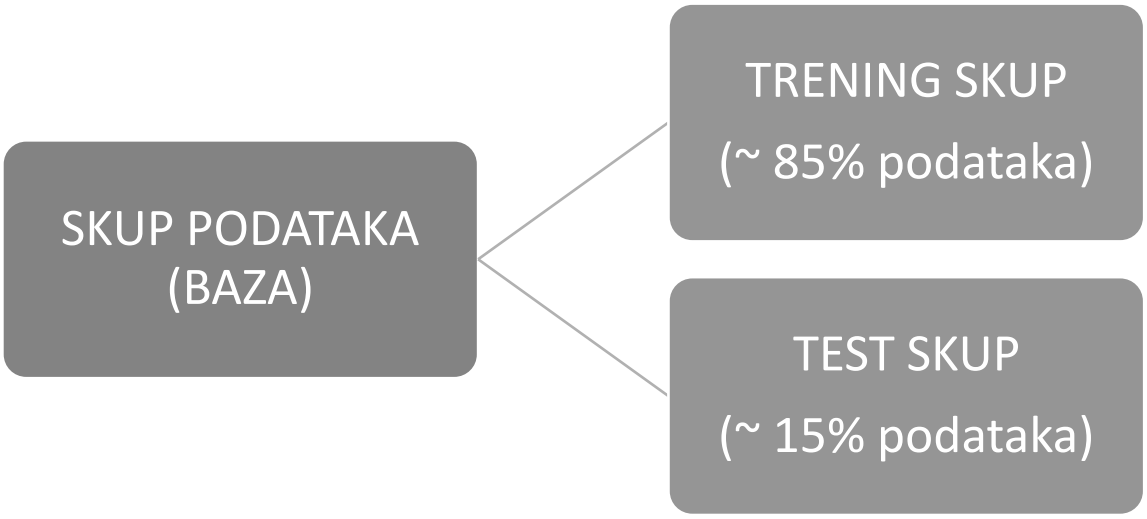
Za odabir modela uobičajeno je usporediti više modela. Svaki je model umjetni entitet koji se temelji na skupu pretpostavki i obično uključuje jedan ili više parametara koji se moraju procijeniti pomoću poznatih povijesnih podataka. S obzirom na dostupne podatke, vezu između zavisnih i nezavisnih varijabli, te samu točnost modela, odabire se najbolji koji će se koristiti za predviđanje. Prikaz prilagođenih krivulja (eng. *fitting curves*) različitih modela prikazan je na sljedećoj slici (Slika 2.5).



Slika 2.5. Različiti prilagođeni modeli [5]

2.2.5. Validacija modela

Validacija modela obuhvaća evaluaciju modela koji će se koristiti za predviđanje budućih stanja kako bi se dobio uvid u točnost istog. Najzastupljenija metoda validacije je korištenje testnog skupa podataka. Slika 2.6 prikazuje najčešću podjelu podataka na trening i test skup.



Slika 2.6 Podjela podataka na skupove

Skup podataka (baza podataka) dijeli se na trening skup koji sadrži najčešće 85% podataka i test skup koji sadrži 15% podataka. Trening skup se koristi za prilagodbu i kreaciju modela za predviđanje, dok se testni skup koristi za testiranje kreiranog modela. Testni skup predstavlja skup „novih“ podataka kojim se simuliraju uvjeti za testiranje i provjeru kreiranog modela.

2.2.6. Verifikacija modela – kontrola rada modela

Nakon odabira modela i procjene njegovih parametara, model se koristi za predviđanje budućih stanja. Učinkovitost modela može se ispravno procijeniti tek nakon što podaci o promatranom procesu postanu dostupni tijekom vremena. Postoje brojne metode s kojima se procjenjuje točnost generiranog predviđanja. Naravno, rezultate je potrebno pratiti (kontrolirati) kako bi se verificirala točnost modela u eksploataciji, odnosno poduzeli potrebni koraci optimizacije ili promjene modela ukoliko prvotno odabrani ne zadovoljava postavljene kriterije (ciljeve).

3. STATISTIČKE METODE PREDVIĐANJA VREMENSKIH SERIJA

Predviđanje vremenskih serija vrši se s ciljem predviđanja, odnosno prognoze budućih vrijednosti vremenske serije. Postoje različiti modeli koji se koriste u tu svrhu, a sam proces nastanka detaljnije je objašnjen u drugom poglavlju. Matematičko modeliranje je proces pronalaženja relacija tj. odnosa među komponentama modela koji precizno određuju stanje neke promatrane pojave tj. stanje pojave istraživanja. Kod analize vremenskih serija pojavljuje se i statističko modeliranje koje je specifičniji oblik matematičkog modeliranja u kojem se javljaju slučajne varijable koje utječu na sami model i točnost istog.

Podaci vremenskih serija zahtijevaju obradu koja uključuje „čišćenje“ i određeni oblik transformacije istih. Također, vrlo je bitno odrediti koliki vremenski raspon će se predviđati jer postoje kratkoročna (eng. *short term*) i dugoročna (eng. *long term*) predviđanja. Prije navedeni koraci direktno utječu na odabir i parametre modela.

Eksponencijalno zaglađivanje i ARIMA (eng. *autoregressive integrated moving average*) modeli dva su najčešće korištena pristupa predviđanju vremenskih serija i pružaju komplementarne pristupe problemu. Dok se eksponencijalni modeli zaglađivanja temelje na trendu i sezonalnosti podataka, ARIMA modeli imaju za cilj opisati i objasniti autokorelacije u podacima. U nastavku će biti prikazani model eksponencijalnog zaglađivanja i ARIMA model.

3.1. Pojmovi potrebni za razumijevanje metoda predviđanja vremenskih serija

Prije prelaska na samu strukturu modela, objasnit će se pojmovi koji su potrebni za razumijevanje kasnije prikazanih modela.

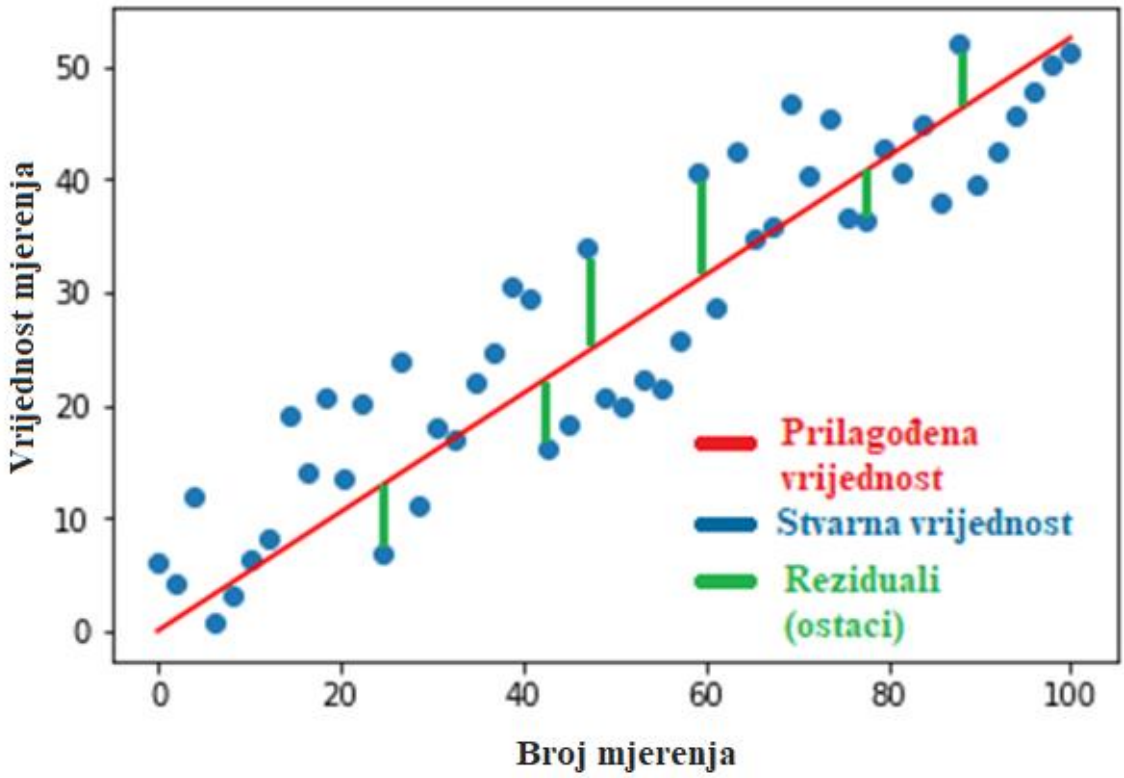
- **Prilagođene vrijednosti (eng. *fitted values*)**

Svaki podatak iz vremenske serije može se „predvidjeti“ pomoću svih ostalih opažanja. Te vrijednosti nazivaju se prilagođene vrijednosti i označavaju se $\hat{y}_{t-1|t}$ te one predstavljaju „predviđanje“ varijable y_t koje se temelji na podacima $y_1, \ldots, y_{t-1}$.

Zapravo, prilagođene vrijednosti nisu isto što i predviđanje jer se bilo koji parametar uključen u metodu prilagodbe („predviđanja“) procjenjuju korištenjem svih dostupnih podataka iz vremenske serije, uključujući i buduća opažanja (npr. y_{t+1}). Iz tog su se razloga pri objašnjavanju ovog pojma koristili navodnici uz riječ „predviđanje“.

- **Reziduali (ostaci)**

Rezudali odnosno ostaci su za modele vremenskih serija jednaki razlici između stvarnog opažanja (podatka iz vremenske serije) i odgovarajuće prilagođene vrijednosti.



Slika 3.1 Prikaz rezidualnih vrijednosti

Slika 3.1 prikazuje grafički što su rezidualne vrijednosti, odnosno prije spomenuta razlika stvarne vrijednosti i prilagođene vrijednosti. Rezidualne vrijednosti mogu poprimiti pozitivan i negativan predznak što je vidljivo iz prije spomenute slike.

3.2. Metode eksponencijalnog zaglađivanja

Metoda eksponencijalnog zaglađivanja razvijena je krajem 1950-ih godina (Brown, 1959; Holt, 1957; Winters, 1960) i rezultirala je jednom od najuspješnijih metoda predviđanja. [4] Predviđanja izrađena korištenjem metode eksponencijalnog zaglađivanja ponderiraju se s obzirom na starost podatka (kada je podatak zabilježen u vremenu), pri čemu vrijednost faktora težine (eng. *weight*) raste eksponencijalno sa starošću podataka. Drugim riječima, što je novije opažanje, veći je i faktor težine. Za lakše razumijevanje, na sljedećoj slici prikazan je odnos vremena i faktora težine za izmjerene vrijednosti (podatke) u vremenu (Slika 3.2).



Slika 3.2 Graf odnosa vremena i faktora težine za jedan podatak (mjerenje)

U nastavku će biti prikazane sljedeće metode eksponencijalnog zaglađivanja:

- jednostavno eksponencijalno zaglađivanje,
- Holt – Winters metoda sezonalnosti.

3.2.1. Metoda jednostavnog eksponencijalnog zaglađivanja

Metoda jednostavnog eksponencijalnog zaglađivanja (eng. *simple exponential smoothing*) prikladna je za kreiranje modela za predviđanje temeljenog na vremenskim serijama koje nemaju:

- izražen trend i
- sezonalnost.

Upravo zbog isključivanja ove dvije komponente ova metoda naziva se jednostavnom. Slika 3.2 prikazuje pravilo na kojem se sama metoda zasniva. Formula za predviđanje je sljedeća [4]:

$$\hat{y}_{T+1|T} = \alpha \cdot y_T + \alpha \cdot (1-\alpha) \cdot y_{T-1} + \alpha(1-\alpha)^2 \cdot y_{T-2} + \cdots \tag{3.1}$$

gdje su:

$\hat{y}_{T+1|T}$ – predviđanje za diskretno vrijeme $T + 1$,

α – faktor zaglađivanja koji može poprimiti vrijednosti od 0 do 1,

T – diskretno vrijeme u vremenskoj seriji.

3.2.2. Holt – Winters metoda

Holt – Winters metoda proširenje je metode jednostavnog eksponencijalnog zaglađivanja, ali s jednom bitnom razlikom, a to je da uključuje pojavu sezonalnosti i trenda pri predviđanju vremenskih serija.

U ovoj metodi uključene su tri jednadžbe zaglađivanja, a to su:

- jednadžba za stupanj (eng. *level*) – l_t,
- jednadžba za trend – b_t,
- jednadžba za sezonalnost – s_t.

Svaka od njih posjeduje svoj parametar zaglađivanja:

- parametar za jednadžbu stupanja – α,
- parametar za jednadžbu trenda – β,
- parametar za jednadžbu sezonalnosti – γ.

Postoje dvije varijacije ove metode koje se razlikuju s obzirom na komponentu sezonalnosti:

- aditivna i
- multiplikativna.

3.2.2.1. Aditivna Holt – Winters metoda

Aditivna metoda koristi se kada su sezonske varijacije (sezonska komponenta) približno konstantne kroz cijelu vremensku seriju, te je sezonska komponenta izražena u apsolutnom obliku. Jednadžbe kojima se opisuje ovaj model su sljedeće [4]:

$$\hat{y}_{t+h|t} = l_t + h \cdot b_t + s_{t+h-m\cdot(k+1)} \tag{3.2}$$

$$l_t = \alpha \cdot (y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{3.3}$$

$$b_t = \beta \cdot (l_t - l_{t-1}) + (1-\beta) \cdot b_{t-1} \tag{3.4}$$

$$s_t = \gamma \cdot (y_t - l_{t-1} - b_{t-1}) + (1-\gamma) \cdot s_{t-m} \tag{3.5}$$

gdje su:

m – frekvencija sezonalnosti (npr. za kvartalnu podjelu m = 4, za mjesečnu podjelu m = 12),

h – korak predviđanja,

k – faktor koji je jednak (h – 1) / m.

3.2.2.2. Multiplikativna Holt – Winters metoda

Multiplikativna metoda se koristi kada sezonske varijacije (sezonska komponenta) nisu konstante kroz cijeli period vremenske serije, te je sezonska komponenta izražena u relativnom obliku (u postocima). Jednadžbe kojima se opisuje ovaj model su [4]:

$$\hat{y}_{t+h|t} = (l_t + h \cdot b_t) \cdot s_{t+h-m\cdot(k+1)} \tag{3.6}$$

$$l_t = \alpha \cdot \frac{y_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{3.7}$$

$$b_t = \beta \cdot (l_t - l_{t-1}) + (1-\beta) \cdot b_{t-1} \tag{3.8}$$

$$s_t = \gamma \cdot \frac{y_t}{l_{t-1} + b_{t-1}} + (1-\gamma) \cdot s_{t-m} \tag{3.9}$$

Oznake koje se koriste identične su onima kod aditivne H. – W. metode.

3.3. ARIMA metode

ARIMA (eng. *autoregressive integrated moving average*) modeli pružaju drukčiji pristup predviđaju vremenskih serija od metode eksponencijalnog zaglađivanja. Naime, eksponencijalno zaglađivanje i ARIMA modeli daju komplementarne pristupe rješavanju problema. Dok se eksponencijalni modeli zaglađivanja temelje na opisu trenda i sezonalnosti podataka, ARIMA modeli temelje se na autokorelaciji i autoregresiji podataka.

U statistici je za analizu i prognozu vremenskih serija najčešće upotrebljavana ARIMA metoda, odnosno metoda bazirana na autoregresivnom modelu i modelu pokretnih prosjeka.

Model koji nastaje ARIMA metodom može sadržavati sljedeće modele i komponente:

- **Autoregresivni model (AR)** – dio modela koji koristi linearne kombinacije prije izmjerenih vrijednosti (engl. *lagged values*) kako bi predvidio buduće vrijednosti.
- **Integrirana komponenta (I)** – dio modela koji vrši diferencijaciju nad originalnom vremenskom serijom sa svrhom postizanja stacionarnosti iste.
- **Model pomičnog prosjeka (MA)** – dio modela koji koristi linearne kombinacije ostataka, odnosno rezidualnih vrijednosti serije kako bi predvidio buduće vrijednosti.

U nastavku će biti opisane prije navedene komponente ARIMA metode.

3.3.1. Autoregresivni (AR) model

Autoregresivni model (eng. *autoregressive model*) za predviđanje koristi linearne kombinacije prije izmjerenih vrijednosti vremenske serije te rezidualnu komponentu (bijeli šum) u trenutnom koraku.

Stoga se AR (*p*) odnosno, autoregresivni model reda *p* može zapisati pomoću formule [6]:

$$y_t = C + \sum_{i=1}^{p} \varphi_i \cdot y_{t-i} + \varepsilon_t \tag{3.10}$$

gdje su:

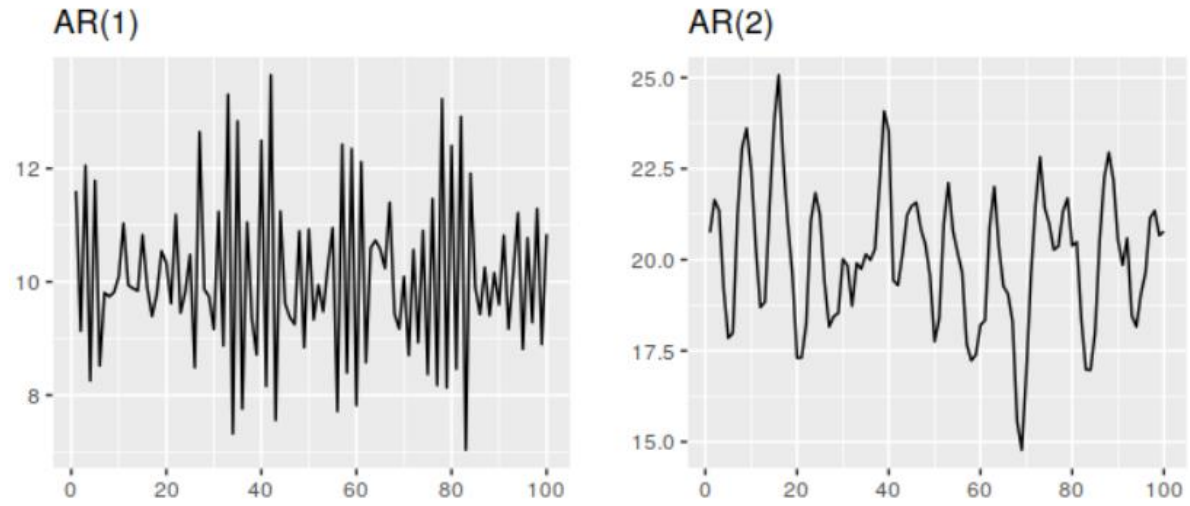
p – red autoregresivnog modela,

C – konstanta,

φ_i – parametri modela,

ε_t – rezidualna komponenta modela (bijeli šum).

Kod autoregresivnog modela promjena reda *p* i parametra φ_i mogu rezultirati različitim obrascima (eng. *patterns*) vremenskih serija, dok varijanca rezidualne komponente može utjecati samo na skalu (eng. *scale*) vremenske serije, ali ne i na obrasce.



Slika 3.3 Usporedba AR (1) i AR (2) modela [4]

Slika 3.3 prikazuje modele AR (1) i AR (2) iste vremenske serije, te se može primijetiti da se promjenom reda autoregresivnog modela promijenio i obrazac ponašanja vremenske serije.

Također, vidljiva je promjena skale na osi ordinata što ukazuje na promjenu rezudualnih komponenti.

Kako su modeli AR (1) i AR (2) najčešće upotrebljavani, bitno je napomenuti uvjete koje oni moraju zadovoljavati, a oni su:

AR(1):
- vremenska serija mora biti stacionarna
- $-1 < \varphi_1 < 1$

AR(2):
- vremenska serija mora biti stacionarna
- $-1 < \varphi_2 < 1$
- $\varphi_1 + \varphi_2 < 1$
- $\varphi_2 - \varphi_1 < 1$

3.3.2. Model pomičnog prosjeka (MA)

Model pomičnog prosjeka (eng. *moving average model*) za predviđanje koristi linearne kombinacije ostataka, odnosno rezidualnih vrijednosti.

Stoga se MA (q), odnosno model pomičnog prosjeka reda q može zapisati pomoću formule [6]:

$$y_t = C + \varepsilon_t + \sum_{i=1}^{q} \theta_i \cdot \varepsilon_{t-i} \tag{3.11}$$

gdje su:

q – red modela pomičnog prosjeka,

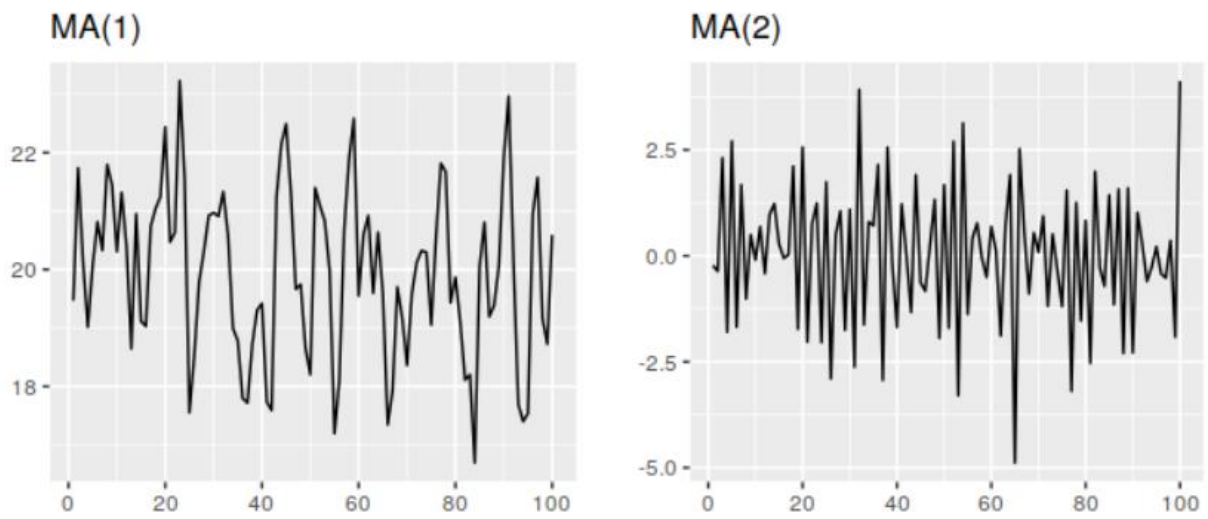
C – konstanta,

θ_i – parametri modela,

ε_t – rezidualna komponenta modela (bijeli šum).

Model pomičnog prosjeka kod ARIMA metode treba razlikovati od metode pomičnih prosjeka koja se koristi za zaglađivanje (eng. *moving average smoothing*) jer se prvotno spomenuta koristi za predviđanje budućih vrijednosti, dok se potonja koristi za zaglađivanje trenda, odnosno ciklusa prošlih vrijednosti.

Slično kao kod autoregresivnog modela, promjena reda q i parametra θ_i mogu rezultirati različitim obrascima (eng. *patterns*) vremenskih serija, dok varijanca rezidualne komponente može utjecati samo na skalu (eng. *sacle*) vremenske serije, ali ne i na obrasce.



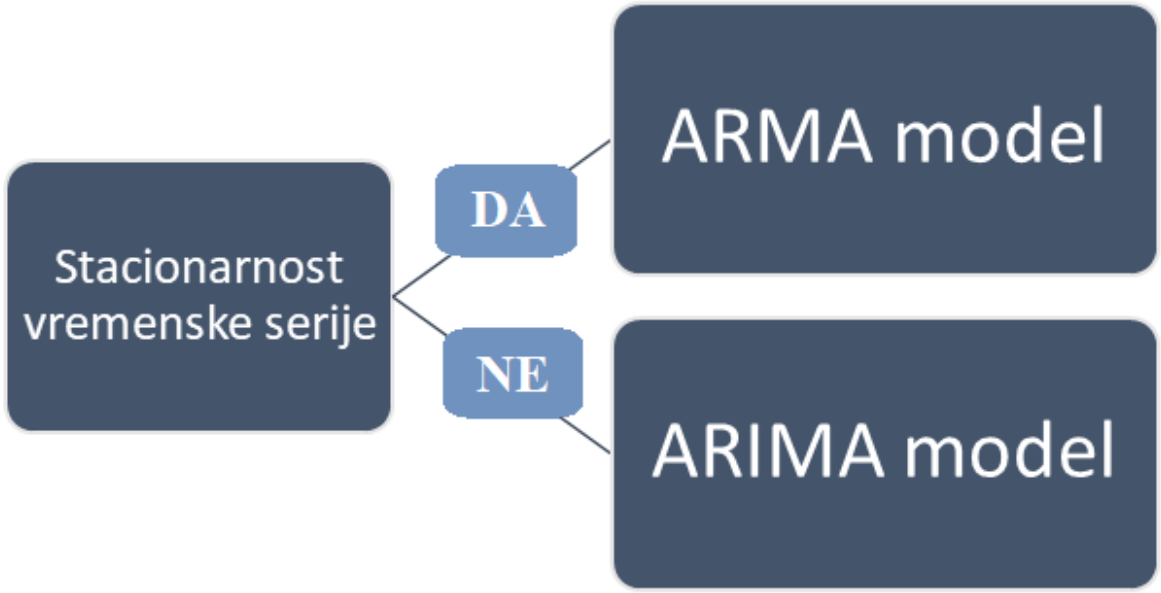
Slika 3.4 Usporedba MA (1) i MA (2) modela [4]

Slika 3.4 prikazuje modele MA (1) i MA (2) iste vremenske serije, te se može primijetiti da se promjenom reda modela pomičnih prosjeka promijenio izgled obrasca ponašanja vremenske serije. Također, vidljiva je promjena skale na osi ordinata zbog promjene rezudualnih komponenti. Bitno je napomenuti da za model pomičnog prosjeka vremenska serija mora biti stacionarna.

3.3.3. Integrirana (I) komponenta

Kod ARIMA metode integrirana komponenta se može, a i ne mora pojaviti. Hoće li ona biti uključena u model ovisi o stacionarnosti vremenske serije. Slika 3.5 prikazuje prije spomenuti uvjet i model kojim metoda rezultira ukoliko je uvjet zadovoljen ili ne. Ukoliko je uvjet stacionarnosti postignut, koristit će se ARMA metoda (integrirana komponenta neće biti prisutna), dok će se u suprotnom, ako uvijek stacionarnosti nije postignut, koristiti ARIMA metoda (integrirana komponenta će biti prisutna).

Stacionarna vremenska serija je onaj čija svojstva ne ovise o vremenu nastanka mjerenja (podataka). Dakle, vremenske serije s trendovima ili sa sezonalnošću nisu stacionarne - trend i sezonalnost utjecat će na vrijednosti podataka vremenske serije.



Slika 3.5 Stacionarnost vremenske serije i ARIMA modeli

Stacionarnost vremenske serije može se provjeriti na više načina:

- autokorelacijskim funkcijama:
  - autokorelacijska funkcija (eng. *autocorrelation function*, ACF),
  - parcijalna autokorelacijska funkcija (eng. *partial autocorrelation function*, PACF),
- Dickey Fuller testom,
- Ljung – Box testom.

Ukoliko vremenska serija nije stacionarna, potrebno je provesti diferencijaciju iste kako bi se postigla željena stacionarnost koja je potrebna za primjenu ARMA modela. Diferencijacija je postupak koji se provodi unutar integrirane komponente ARIMA modela.

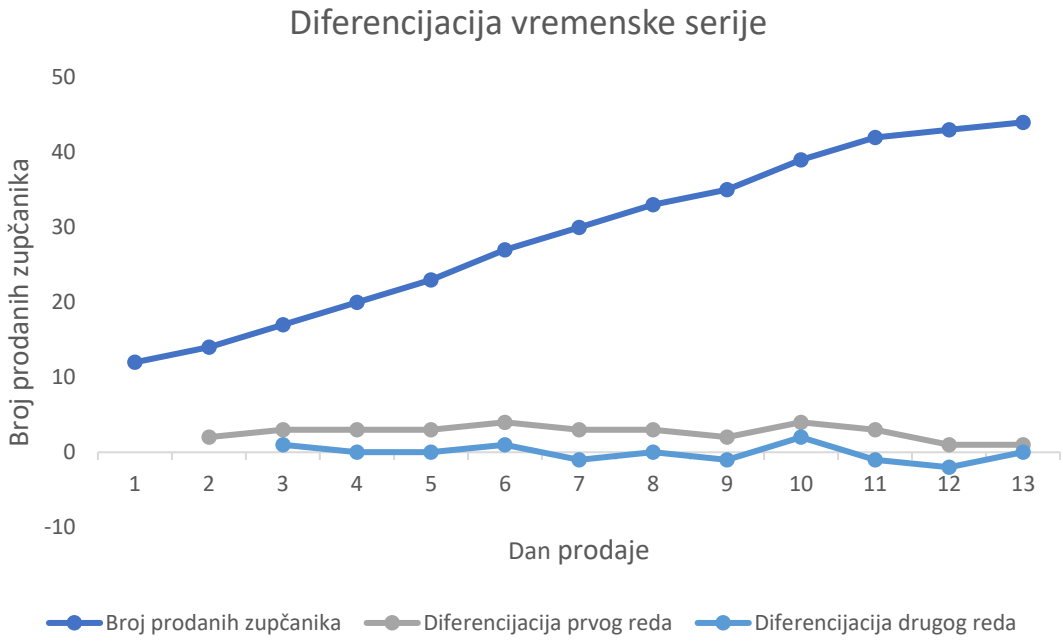
3.3.3.1. Diferencijacija

Diferencijacija je matematički postupak koji služi za postizanje stacionarnosti vremenskog niza. Stupanj diferencijacije označava se s *d*. Najjednostavnije rečeno, diferencijacija predstavlja razliku između neposredno izmjerenih vrijednosti, odnosno podataka unutar vremenske serije. Neke druge transformacije mogu pomoći pri „stabilizaciji“ varijance vremenske serije, međutim ne i pri „stabilizaciji“ srednje vrijednosti iste. Upravo je zbog toga diferencijacija vremenskih serija odličan alat za postizanje stacionarnosti. Diferencijacija vremenskog niza prodaje zupčanika prikazana je u sljedećoj tablici (Tablica 3.1).

Tablica 3.1. Diferencijacija vremenske serije - podaci

Dan	Broj prodanih zupčanika	Diferencijacija prvog reda	Diferencijacija drugog reda
1	12		
2	14	2	
3	17	3	1
4	20	3	0
5	23	3	0
6	27	4	1
7	30	3	-1
8	33	3	0
9	35	2	-1
10	39	4	2
11	42	3	-1
12	43	1	-2
13	44	1	0

Iz tablice podataka lako je uočljiv obrazac prema kojem se vrši diferencijacija. Svaka vrijednost jednaka je razlici trenutnog broja prodanih zupčanika u danu i dana koji mu je prethodio. Diferencijacija drugog reda vrši se na setu podataka nastao prvom diferencijacijom. Postupak može trajati dok se ne zadovolji uvjet stacionarnosti, međutim većim redom diferencijacije model postaje sve kompliciraniji, a samim time ugrožena je i točnost modela za predviđanje.



Slika 3.6 Grafički prikaz vremenske serije prije i nakon diferencijacije

Slika 3.6 grafički prikazuje vremensku seriju prodaje zupčanika prije i nakon diferencijacije prvog i drugog reda. Uočljivo je da broj prodanih zupčanika prati linearan trend rasta koji upućuje da vremenska serija nije stacionarna. Ovaj proces naziva se vizualnom kontrolom vremenske serije i njene stacionarnosti. Svakako bi se i statistički trebalo dokazati da model nije stacionaran. Da bi vremenska serija bila stacionaran, mora zadovoljavati sljedeće uvjete:

- srednja vrijednost (μ) vremenske serije je konstantna,
- standardna devijacija (σ) vremenske serije je konstantna,
- ne postoji sezonalnost unutar vremenske serije.

Nakon diferencijacije prvog i drugog reda, može se primijetiti da je trend gotovo pa nestao. Ovu tvrdnju bi također trebalo potkrijepiti relevantnim statističkim mjerama. Međutim, vidljivo je da diferencijacija podataka vremenske serije uvelike pomaže pri uklanjanju nestacionarnosti. Upravo je to ono što je preduvjet za korištenje ARMA modela – stacionarnost modela.

3.3.3.2. Autokorelacijske funkcije

Analiza autokorelacijskih funkcija važna je iz dva razloga:

- pomoću nje se može zaključivati o stacionarnosti vremenske serije,
- pomoću nje se može predložiti parametar *p* atuoregresivnog modela, odnosno parametar *q* modela pomičnih prosjeka.

Bitno je razlikovati korelaciju od autokorelacije. Stoga, korelacija je mjera povezanosti između dvije različite varijable, dok je autokorelacija mjera povezanosti između vrijednosti jedne varijable, odnosno između izmjerenih vrijednosti iste. U ovom slučaju spomenute vrijednosti tvore vremensku seriju. Grafovi autokorelacijskih funkcija nazivaju se korelogrami (eng. *correlograms*). Postoje različiti razredi korelacije, a formula za njihovo izračunavanje je sljedeća [4]:

$$r_k = \frac{\sum_{t=k+1}^{T}(y_t - \overline{y}) \cdot (y_{t-k} - \overline{y})}{\sum_{t=1}^{T}(y_t - \overline{y})^2} \tag{3.12}$$

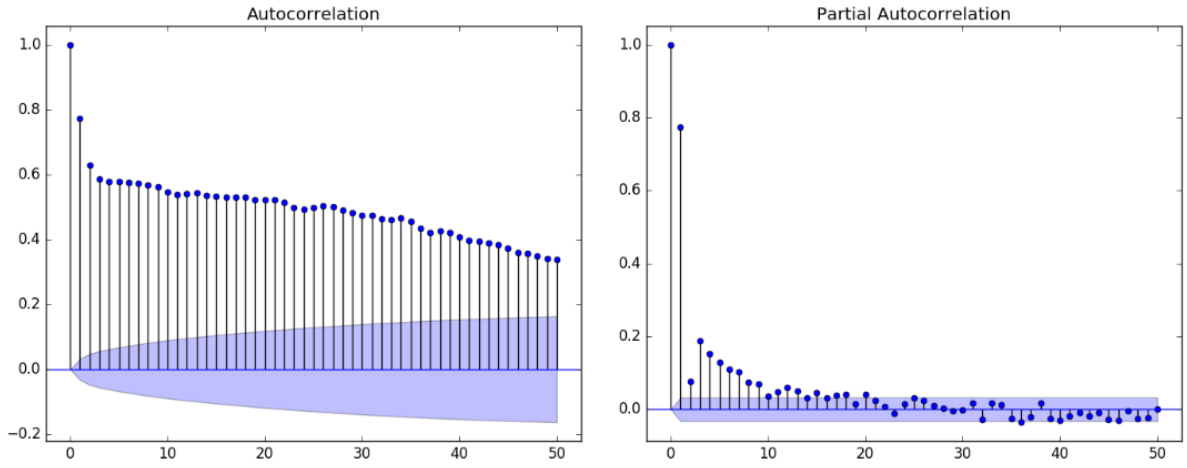
gdje su:

r_k – koeficijent korelacije između vrijednosti y_t i y_{t-k} , a poprima vrijednosti u intervalu [-1,1],

$\overline{y}$ – aritmetička sredina vrijednosti podataka vremenske serije,

T – duljina vremenske serije.

Slika 3.7 prikazuje korelograme autokorelacijskih funkcija te su intervali povjerenja označeni površinom plave boje.

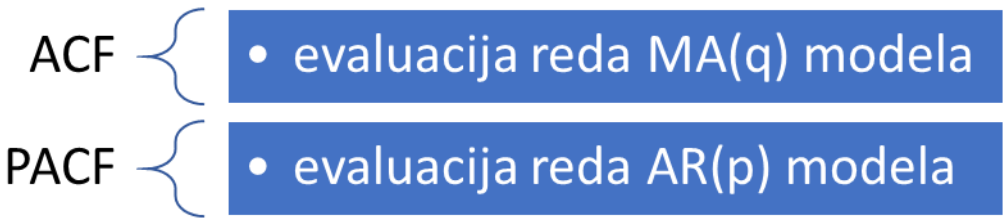


Slika 3.7 Korelogrami – ACF i PACF [7]

Postoje dvije vrste autokorelacijskih funkcija:

- ACF – autokorelacijska funkcija,
- PACF – parcijalna autokorelacijska funkcija.

Autokorelacijska funkcija (ACF) prikazuje korelaciju vremenske serije s obzirom na njene prijašnje vrijednosti (mjerenja). Ukratko, ACF funkcija prikazuje u kojoj su mjeri vremenski odijeljene vrijednosti serije korelirane uzimajući u obzir sve indirektne vrijednosti vremenske serije koje se nalaze između njih. ACF funkcijom evaluira se red MA model (Slika 3.8).
Parcijalna autokorelacijska funkcija (PACF) također prikazuje korelaciju vremenske serije s obzirom na prijašnje vrijednosti (mjerenja), međutim ona uzima u obzir direktnu korelaciju između dvije vrijednosti vremenske serije. Drugim riječima, PACF funkcija poništava efekte svih indirektnih vrijednosti koje se nalaze između vrijednosti (mjerenja) za koje se računa koeficijent autokorelacije. PACF funkcijom evaluira se red AR modela (Slika 3.8).



Slika 3.8 Veza između autokorelacijskih funkcija i parametara *p* i *q*

3.3.4. ARMA, ARIMA i SARIMA model

Podjela modela koje nastaju ARIMA metodom za predviđanje vremenskih serija može se izvršiti s obzirom na komponente koje model sadrži. Komponente koje model može sadržavati, a prije su spomenute su:

- autoregresivna komponenta modela AR (*p*),
- integrirana komponenta I (*d*),
- komponenta pomičnog prosjeka MA (*q*),
- komponenta sezonalnosti.

Podjela modela koji mogu nastati, izuzevši modele koji nastaju kao produkt samostalnih komponenti, je sljedeća:

- ARMA model,
- ARIMA model,
- SARIMA model.

Najčešće korišteni model u praksi je ARIMA. Model se može zapisati kao ARIMA (*p, d, q*), a jednadžba kojom se opisuje je sljedeća [4]:

$$y'_t = C + \sum_{i=1}^{p} \varphi_i \cdot y'_{t-i} + \sum_{i=1}^{q} \theta_i \cdot \varepsilon_{t-i} + \varepsilon_t \qquad (3.13)$$

gdje je y'_t vrijednost diferencirane serije (postupak i svrha diferencijacije opisana je u poglavlju *3.3.3.1*). Ostale oznake jednake su onima definiranim u prijašnjim poglavljima. Konstanta *C* može, a i ne mora biti uključena u model.

Ukoliko analiziranu vremensku seriju nije potrebno diferencirati, odnosno u model nije uključena I (*d*) komponenta, ARIMA model postaje ARMA model. Za ARMA model koristi se ista formula (3.13), uz jednu razliku u zapisu. Kako vrijednosti vremenske serije nisu diferencirane, y'_t prelazi u zapis y_t.

Opis svakog pojedinog parametra ARIMA (*p, d, q*) modela, odnosno njihova značenja prikazani su u sljedećoj tablici (Tablica 3.2).

Tablica 3.2. Popis parametara ARIMA modela

p	Red autoregresivnog modela – AR (*p*).
d	Stupanj diferencijacije modela – I (*d*).
q	Red modela pomičnog prosjeka – MA (*q*).

Ukoliko u vremenskoj seriji postoji i sezonalnost, ARIMA model postaje SARIMA model koji uspješno rješava „problem“ sezonalnosti. SARIMA model ima dodatne parametre, a općeniti zapis je sljedeći:

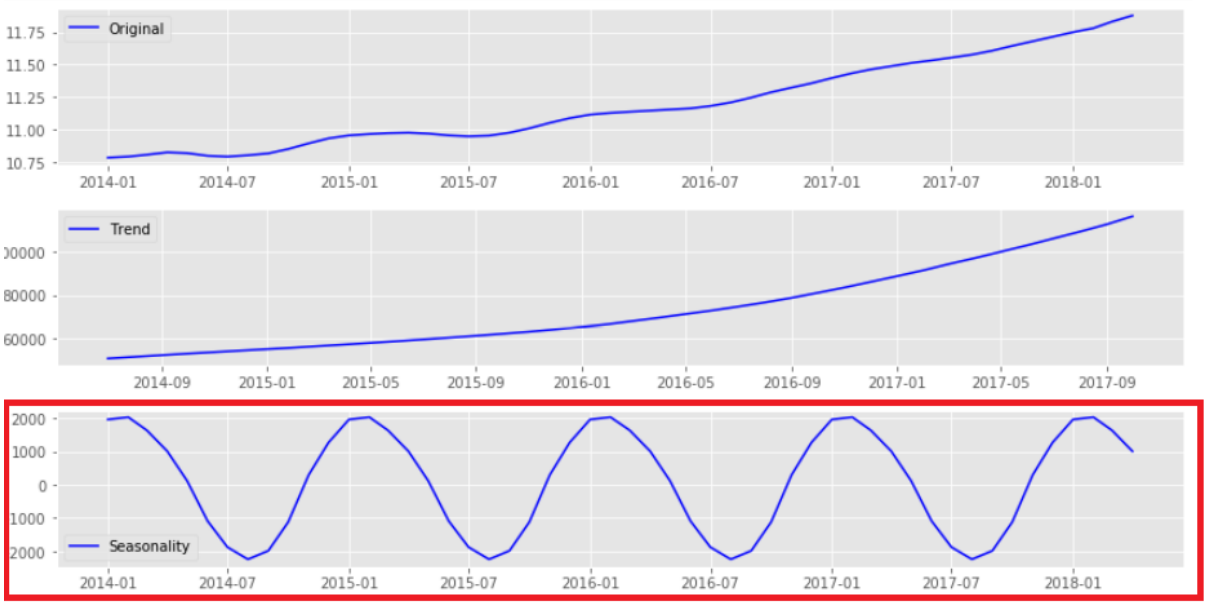
$$\text{ARIMA } (p, d, q)\ (P, D, Q)_m \tag{3.14}$$

gdje su:

(p ,d ,q) – parametri koji se odnose na dio modela koji nema sezonalnost,

(P, D, Q) – parametri koji se odnose na dio modela koji sadrži sezonalnost,

m – parametar koji označava frekvenciju sezonalnosti s obzirom na jednu godinu (npr. m = 12 predstavlja mjesečnu, dok m = 4 predstavlja kvartalnu sezonalnost).



Slika 3.9 Komponenta sezonalnosti u vremenskoj seriji [8]

Slika 3.9 prikazuje detaljnu dekompoziciju vremenske serije, a komponenta sezonalnosti unutar iste posebno je označena crvenim pravokutnikom. Iz slike se može očitati da je frekvencija sezonalnosti jedna godina, odnosno da bi parametar m imao vrijednost 12.

3.3.4.1. Odabir parametara p i q – AIC kriterij

Osim odabira parametara *p* i *q* pomoću već spomenutih autokorelacijskih funkcija u poglavlju *3.3.3.2*, iste je moguće odrediti pomoću AIC kriterija (eng. *Akaike's Information Criterion*). AIC kriteriji pomaže pri evaluaciji *p* i *q* parametara, odnosno daje informaciju o tome koje bi parametre bilo najbolje odabrati za model.

Vrijednost AIC funkcije može se izračunati pomoću sljedeće formule [9]:

$$\mathrm{AIC} = -2 \cdot \log(L) + 2 \cdot (p + q + k + 1) \tag{3.15}$$

gdje su:

L – parametar vjerojatnosti podataka,

k – parametar koji iznosi $k = 1$ uz uvjet ($C \neq 0$), odnosno iznosi $k = 0$ uz uvjet ($C = 0$).

Niža vrijednost AIC funkcije predstavlja bolje prilagođeni model što je bitno za evaluaciju rezultata. Također treba napomenuti da se AIC kriteriji ne bi trebao koristiti za određivanje stupnja diferencijacije (*d*) vremenske serije. Izračun parametra *L*, a samim time i njegova vrijednost, ovisi o podacima iz vremenske serije. Upravo je zbog toga bitno jesu li ti podaci diferencirani ili ne. Vrijednosti AIC kriterija ne mogu se uspoređivati između različitih stupnjeva diferencijacije zbog prije spomenutih razloga.

3.4. Mjere točnosti modela

Nakon izrade modela za predviđanje vremenskih serija, potrebno je znati i evaluirati iste. Drugim riječima, potrebno je ocijeniti točnost izrađenih modela i njihovom usporedbom donijeti odluku o tome koji model će se koristiti za buduća predviđanja. Logika nalaže da će se koristiti onaj model koji ima najveću točnost. Postoje različite mjere kojima se mogu evaluirati kreirani modeli odnosno njihova točnost. Različite mjere daju različite informacije o kreiranim modelima te nas mogu uputiti, odnosno usmjeriti prema ispravnom odabiru, ali isto tako i prema mogućim poboljšanjima. Najčešće korištene statističke mjere točnosti su:

- prosječno apsolutno odstupanje (eng. *mean absolute error,* MAE),
- prosječno postotno apsolutno odstupanje (eng. *mean absolute percentage error,* MAPE),
- korijen srednjeg kvadratnog odstupanja (eng. *root mean square error,* RMSE).

U literaturi se također mogu pronaći nazivi MAD, MAPD i RMSD koji u nazivu umjesto pojma greška (eng. *error*) imaju pojam odstupanje (eng. *deviation*), međutim radi se o istim mjerama. Prosječno apsolutno odstupanje (MAE) predstavlja srednju vrijednost odstupanja svih predviđenih vrijednosti. Ova mjera je zanimljiva jer je dobivena vrijednost izražena u istoj mjernoj jedinici u kojoj je izražena i predviđana varijabla. MAE se računa prema formuli [10]:

$$\mathrm{MAE} = \frac{1}{n} \cdot \sum_{i=1}^{n} |y_t - f_t| \tag{3.16}$$

gdje su:

y_t – stvarna vrijednost podatka vremenske serije u trenutku *t*,

f_t – predviđena vrijednost podatka y_t vremenske serije u trenutku *t*,

n – broj predviđenih vrijednosti vremenske serije.

Prosječno postotno apsolutno odstupanje (MAPE) predstavlja srednju vrijednost odstupanja svih postotnih vrijednosti predviđene varijable. MAPE se računa prema formuli [10]:

$$\mathrm{MAPE} = \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{|y_t - f_t|}{y_t} \tag{3.17}$$

oznake koje se koriste identične su onima onim u MAE mjeri.

Korijen srednjeg kvadratnog odstupanja RMSE najčešće se koristi kod evaluacije modela koji predviđaju vrijednosti vremenske serije. RMSE daje iznos odstupanja koji je izražen u mjernim jedinicama predviđane varijable. RMSE se računa prema formuli [10]:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (y_t - f_t)^2} \tag{3.18}$$

oznake koje se koriste identične su onima onim u MAE mjeri.

4. DEFINIRANJE PROBLEMA PREDVIĐANJA

Definicija problema često predstavlja najteži dio predviđanja. Pažljivo definiranje problema zahtijeva razumijevanje načina na koji će predviđanje biti korišteno i kako se funkcija predviđanja uklapa u širu sliku promatranog problema.

U sustavima proizvodnje i isporuke energije, važno je poznavati potrebe za zahtijevanom količinom energije u nekom budućem vremenskom periodu kako bi se odredili optimalni parametri proizvodnje i dobave. Upotrebom historijskih podataka o utrošku energije na pojedinim objektima te apliciranjem postojećih modela, moguće je predvidjeti zahtjeve za energijom. Temeljem takve predikcije moguće je preciznije odrediti optimalne parametre proizvodnje, a samim time i nadolazeće troškove za dobavu, odnosno proizvodnju potrebne energije.

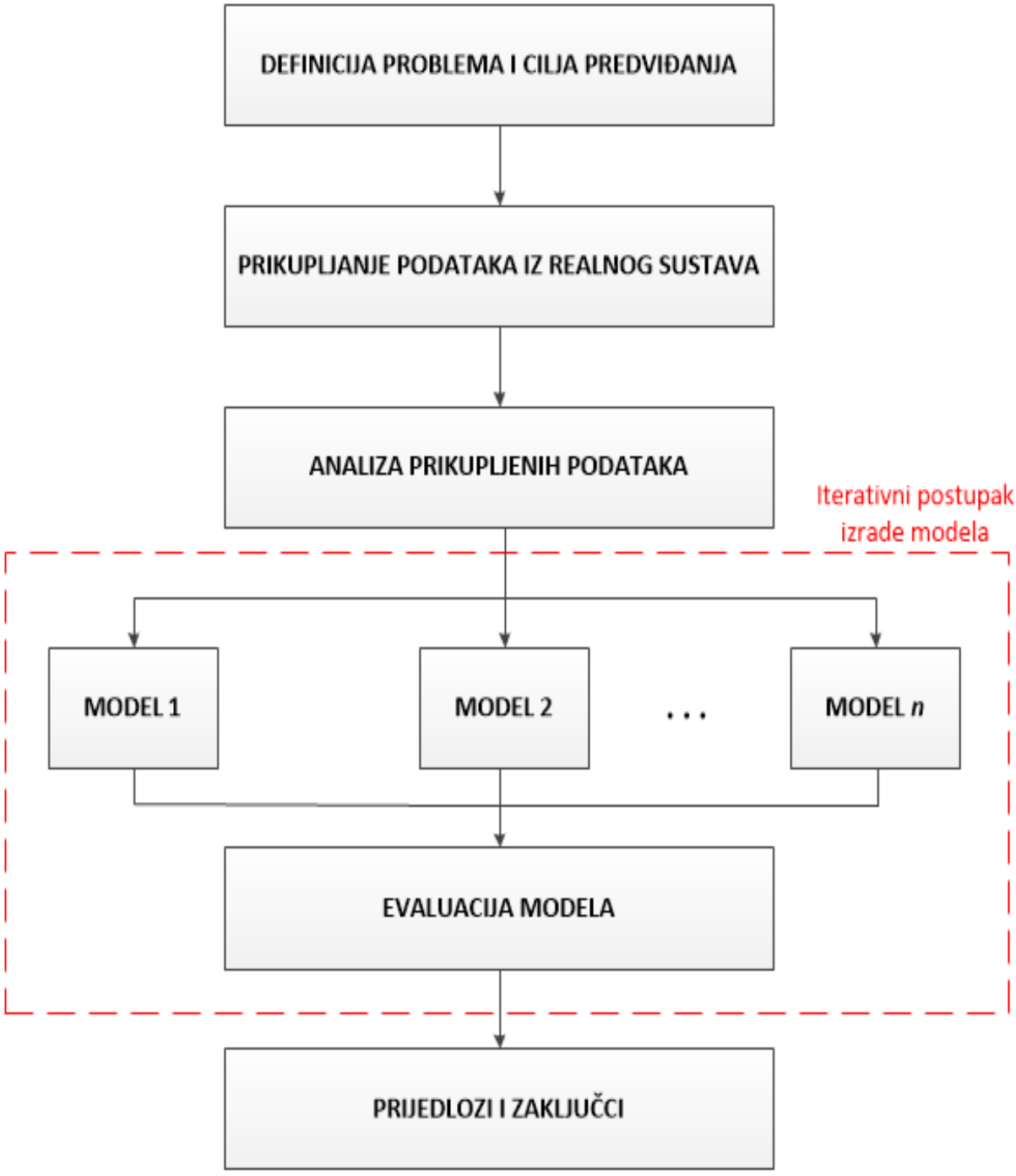
U ovom radu problem je predvidjeti utrošak energije u budućem vremenskom periodu pomoću historijskih podataka. Cilj rada je predložiti model za predviđanje koji će se koristiti za određivanje optimalnih parametara, odnosno optimalnog plana proizvodnje potrebne energije. Definicija i cilj ovog rada prikazani su u sljedećoj tablici (Tablica 4.1).

Tablica 4.1 Definicija problema i cilja predviđanja

Definicija problema predviđanja	Predvidjeti utrošak energije u budućem vremenskom periodu pomoću historijskih podataka.
Cilj predviđanja	Predložiti model za predviđanje koji će se koristiti za određivanje optimalnog plana dobave/proizvodnje iste.

Slika 4.1 prikazuje hodogram aktivnosti koje je potrebno izvršiti kako bi se došlo do postavljenog cilja ovog rada, a to je predviđanje utroška energije pomoću vremenskih nizova i donijeti zaključke o optimalnom planu dobave, odnosno proizvodnje iste. Cijeli proces započinje definicijom problema i cilja predviđanja. Nakon toga slijedi prikupljanje podataka iz realnog sustava opskrbe energijom. Nadalje, potrebno je upoznati se s prikupljenim podacima, odnosno analizirati iste pomoću statističkih metoda. Poslije same analize slijedi izrada odgovarajućih modela za predviđanje, a postupak izrade je iterativan. Evaluacija izrađenih

modela potrebna je kako bi se odredio najtočniji model prema kojem će se vršiti predviđanje i donositi prijedlozi i zaključci o optimalnom planu dobave potrebne energije.



Slika 4.1 Hodogram aktivnosti u procesu predviđanja budućih stanja

5. ANALIZA REALNOG SUSTAVA I PODATAKA

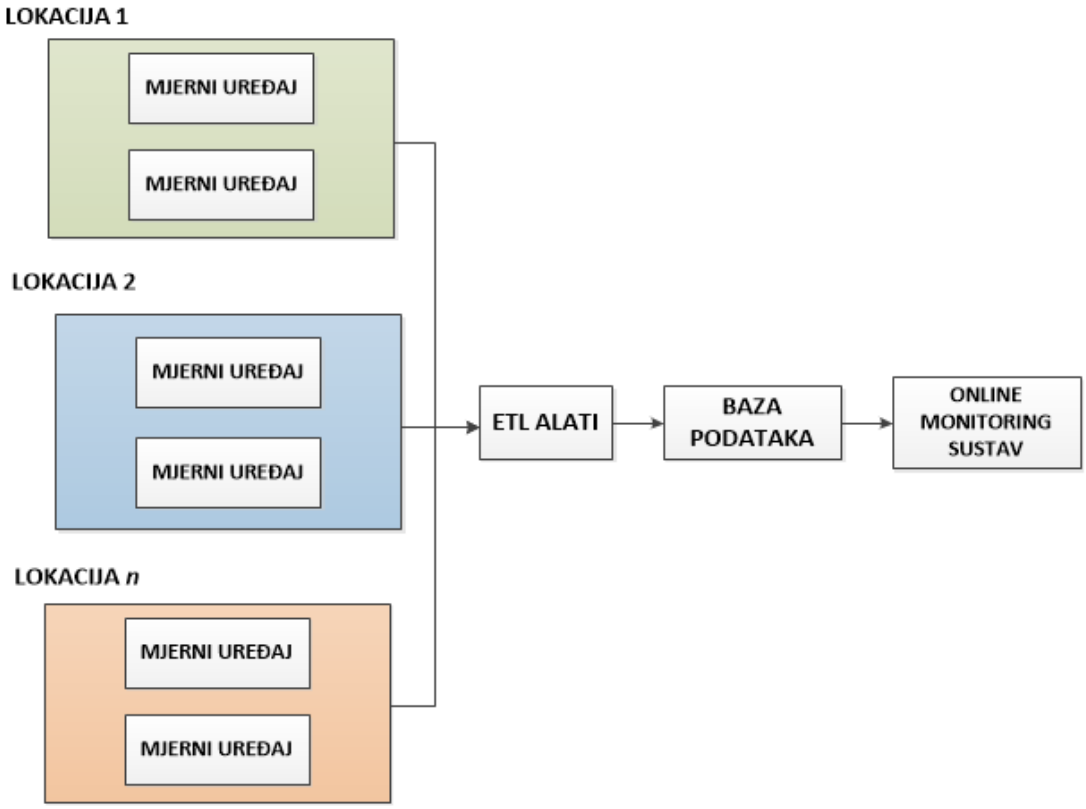
Analiza podataka važan je korak pri odabiru modela predviđanja koji će se koristiti. Dijagrami tj. grafovi podataka vremenske serije trebaju se prikazati i vizualno pregledati kako bi se prepoznale pojave poput trendova, sezonalnosti ili drugih komponenti. Deskriptivna statistika također treba biti uključena u analizu podataka kako bi se uočile zakonitosti i osobine samih podataka, odnosno pretočile u kvantificirani oblik. Cilj analize podataka je stjecanje znanja o podacima koji su dostupni i koji će se koristiti za modeliranje.

U nastavku će biti prikazani sljedeći koraci:

- svojstva i obilježja korištenog realnog sustava,
- prikupljanje podataka iz realnog sustava,
- analiza podatka iz realnog sustava.

5.1. Realni sustav – svojstva i obilježja

Realni sustav koji će se koristiti u ovom radu prikazan je shemom na sljedećoj slici (Slika 5.1):



Slika 5.1 Shema realnog sustava

Prikazani realni sustav sastoji se od većeg broja lokacija na kojima se nalaze brojni mjerni uređaji za praćenje potrošnje energije. Prikupljeni podaci pomoću mjernih uređaja automatski se šalju u bazu podataka koja će biti korištena za analizu i samu izradu modela za predviđanje budućih stanja, odnosno potrošnje energije. Kako bi podaci prikupljeni na mjernim uređajima bili kompatibilni za spremanje u bazu podataka, isti se transformiraju pomoću alata za ekstrakciju, transformaciju i spremanje (eng. *extract, transform, load*, ETL). Tako kreirana baza podataka dostupna je u mrežnom (eng. *online*) sustavu za monitoring svih mjerenja.

5.2. Prikupljanje podataka iz realnog sustava

Prije same analize, potrebno je prikupiti podatke iz realnog sustava. Prikupljanje podataka predstavlja proces prikupljanja relevantnih povijesnih podataka koji potencijalno imaju utjecaj na varijablu ili varijable koje se predviđaju. U ovom slučaju ta varijabla je utrošak energije.

Podaci se prikupljaju daljinski pomoću različitih mjernih uređaja ovisno o tipu energije koja je u fokusu.



Slika 5.2 Mjerni uređaji za praćenje potrošnje energije i drugih parametara

Uređaji za praćenje i prikupljanje podataka o temperaturi i potrošnji energije prikazani su na slici (Slika 5.2). Riječ je o jednotarifnom mjerilu električne energije, kalorimetru (mjeri potrošnju toplinske energije), te senzoru za mjerenje temperature. Podaci koje bilježe ovi

uređaji zapisuju se u bazu podataka kako je objašnjeno u prethodnom poglavlju. Tako stvorena baza podataka predstavljat će različite vremenske serije koje će se koristiti za izradu modela za predviđanje.

5.3. Analiza podataka iz realnog sustava

Analiza podataka iz realnog sustava podijelit će se na nekoliko faza kako je prikazano u sljedećoj tablici (Tablica 5.1).

Tablica 5.1 Faze analize podataka iz realnog sustava

#	Opis faze
1.	Usporedba različitih lokacija (objekata) po njihovim obilježjima.
2.	Analiza vremenske serije odabrane lokacije i energenta.

5.3.1. Usporedba lokacija (građevina)

Svaka lokacija ima svoje specifičnosti odnosno svojstva i obilježja. U nastavku će biti prikazana usporedba četiri lokacije, odnosno četiri objekta.

Energetska bilanca bazira se na informacijama prikupljenim od korisnika energije na definiranim lokacijama. Troškovna bilanca građevine omogućuje prikaz financijskog troška potrošene energije i vode, te koliko u postocima otpada na pojedini energent. Troškovi za energente temeljeni su na tržišnoj cijeni energenta za poslovne korisnike.

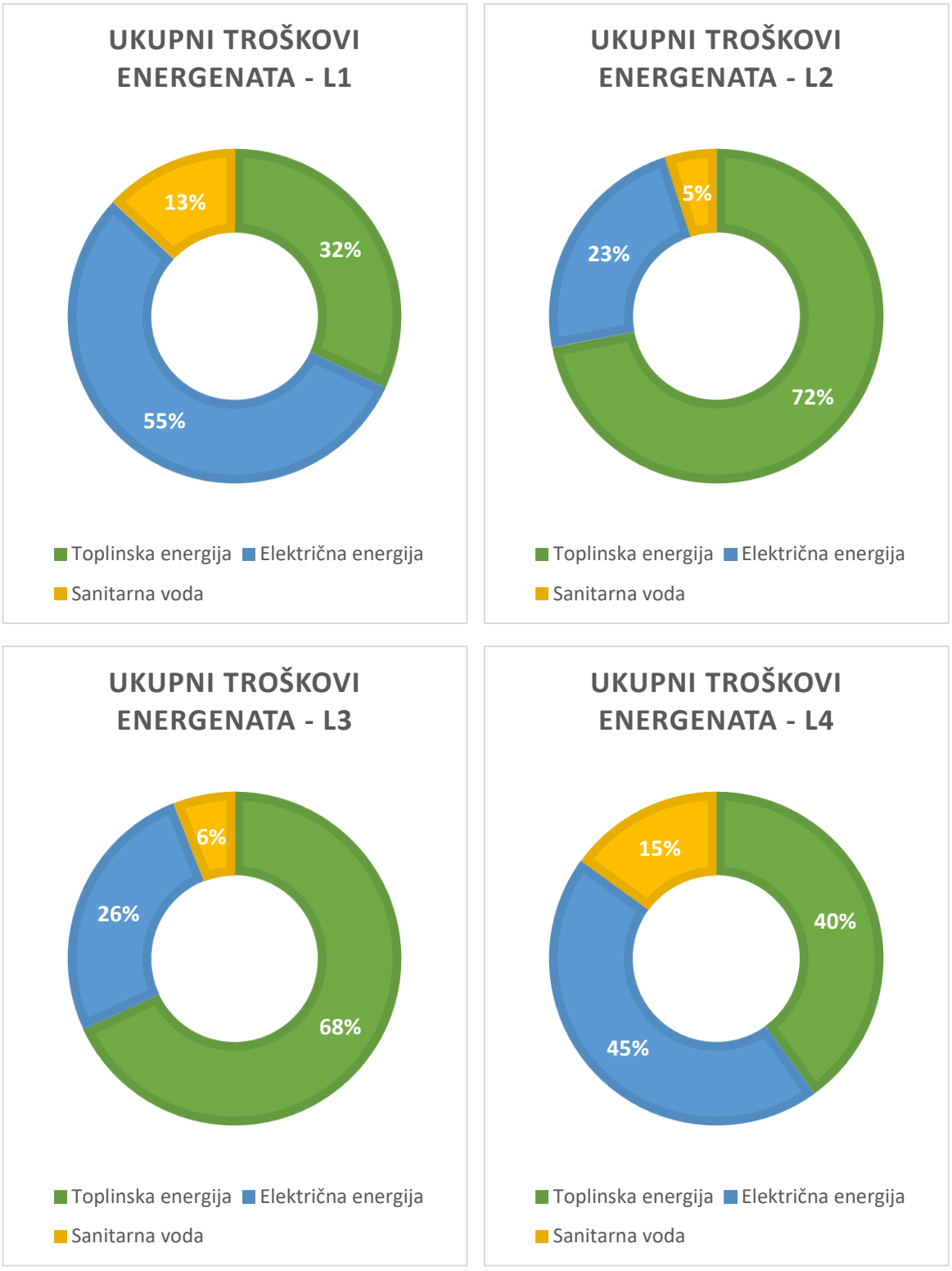
Za pojedine energente treba definirati mjerne jedinice koje će se koristiti u ovom radu. U tablici su prikazani energenti i njihove mjerne jedinice (Tablica 5.2).

Tablica 5.2 Popis energenata i mjernih jedinica

Naziv energenta	Mjerna jedinica
Toplinska energija	kWh
Električna energija	kWh
Sanitarna voda	m^3

5.3.1.1. Usporedba troškova po energentima

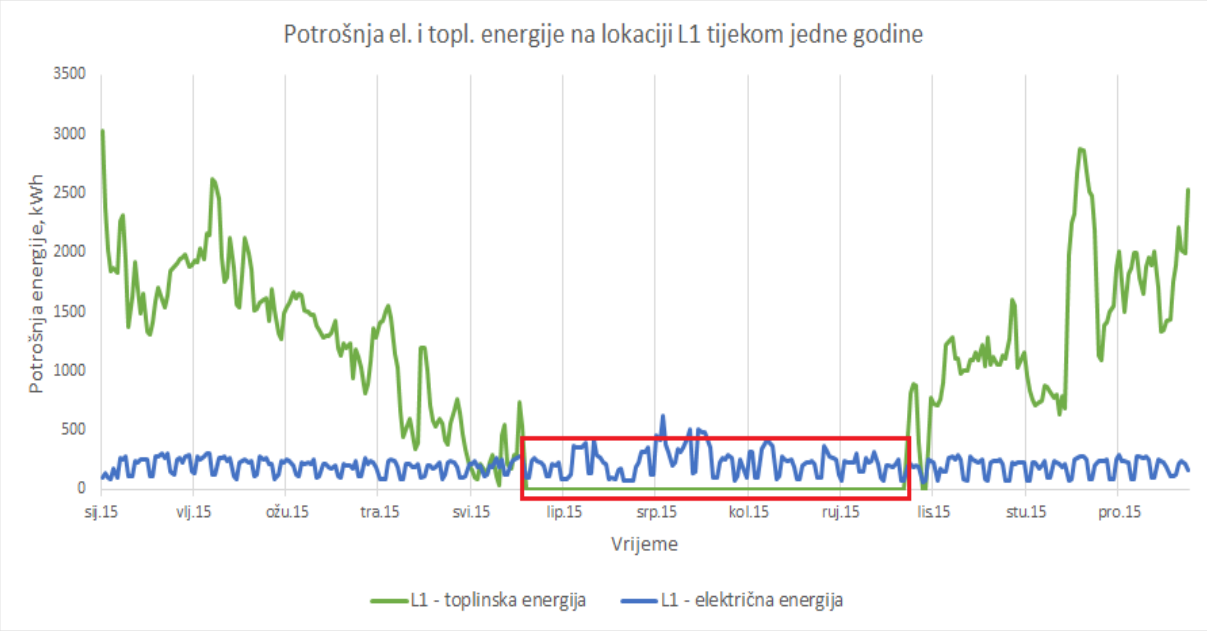
Prva usporedba lokacija (objekata) bit će temeljena na ukupnim modeliranim troškovima pojedinih energenata.



Slika 5.3 Grafovi ukupnog troška energenata po lokacijama

Slika 5.3 prikazuje dijagrame modeliranih ukupnih troškova s obzirom na vrstu energenta i lokaciju. Odabrane su četiri različite lokacije (objekta), a označene su šiframa: L1 – L4.

Iz grafova je vidljivo da se najmanji udio ukupnog troška odnosi na sanitarnu vodu na sve četiri lokacije. Nadalje, bazirano na strukturi bilance troškova, može se zaključiti da lokacija L1 i L4, te lokacija L2 i L3 imaju sličnu strukturu. Odnosno na lokacijama L1 i L4 najveći trošak čini utrošak električne energije, dok na lokacijama L2 i L3 to predstavlja toplinska energija. Zbog slične strukture u potrošnji energenata, može se zaključiti da se na lokacijama L1 i L4 nalaze slični objekti, a isto to vrijedi i za objekte na lokacijama L2 i L3.



Slika 5.4 Potrošnja toplinske i električne energije tijekom jedne godine (L1)

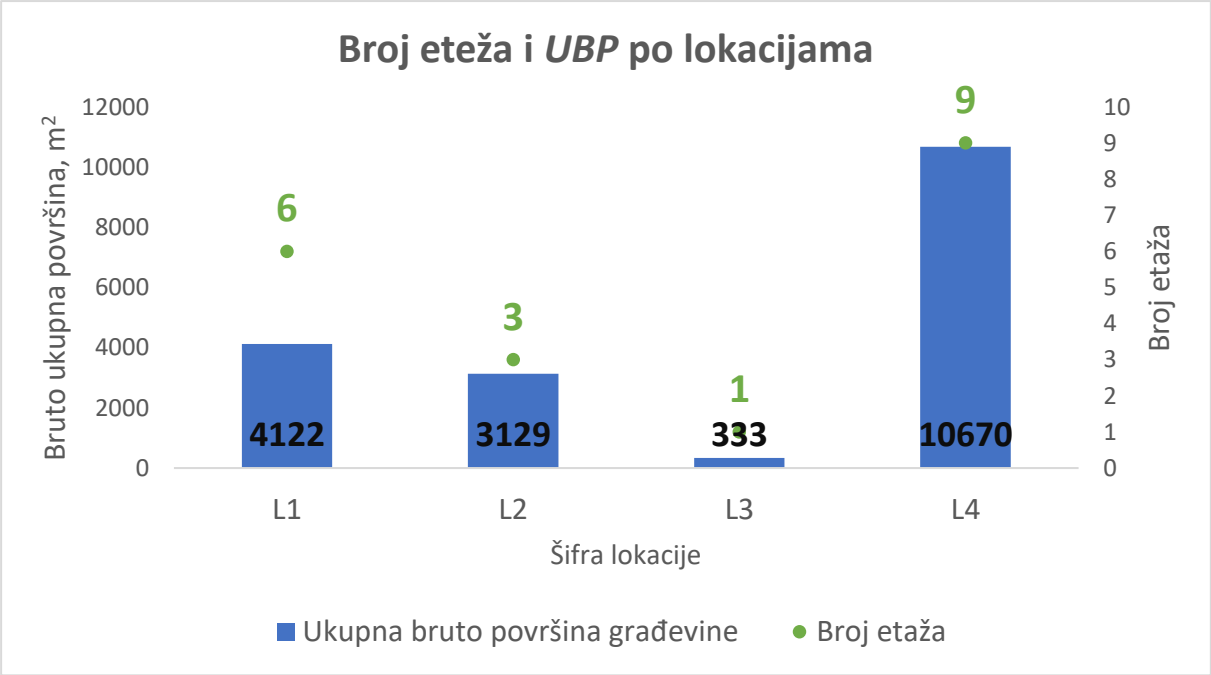
Slika 5.4 prikazuje potrošnju toplinske i električne energije na lokaciji L1 tijekom jedne godine. Uspoređujući količinski u kWh, više je potrošeno toplinske nego električne energije. Razlog razlike u ukupnom trošku koja ide u korist električnoj energiji jest viša cijena električne energije po potrošenom kWh. Međutim, bitno je naglasiti da se električna energija koristi intenzivno kroz cijelu godinu, dok za toplinsku energiju to nije slučaj. Crvenim pravokutnikom označen je period kada nema potrošnje toplinske energije (ljetno doba godine – otprilike četiri mjeseca) (Slika 5.4). Iz toga slijedi bitan zaključak da trećinu ukupnih troškova na potrošene energente odlazi na toplinsku energiju koja se koristi kroz samo tri godišnja kvartala, odnosno kroz osam mjeseci godišnje. Upravo je zbog toga predviđanje potrošnje toplinske energije krucijalno za planiranje godišnjih troškova i plana opskrbe.

5.3.1.2. Usporedba geometrijskih podataka lokacija (objekata)

Druga usporedba lokacija (objekata) bit će temeljena na geometrijsko-građevinskim podacima istih. Prije same usporedbe potrebno je definirati fizikalne veličine koje će se koristiti. Definicije fizikalnih veličina prikazan je u sljedećoj tablici (Tablica 5.3).

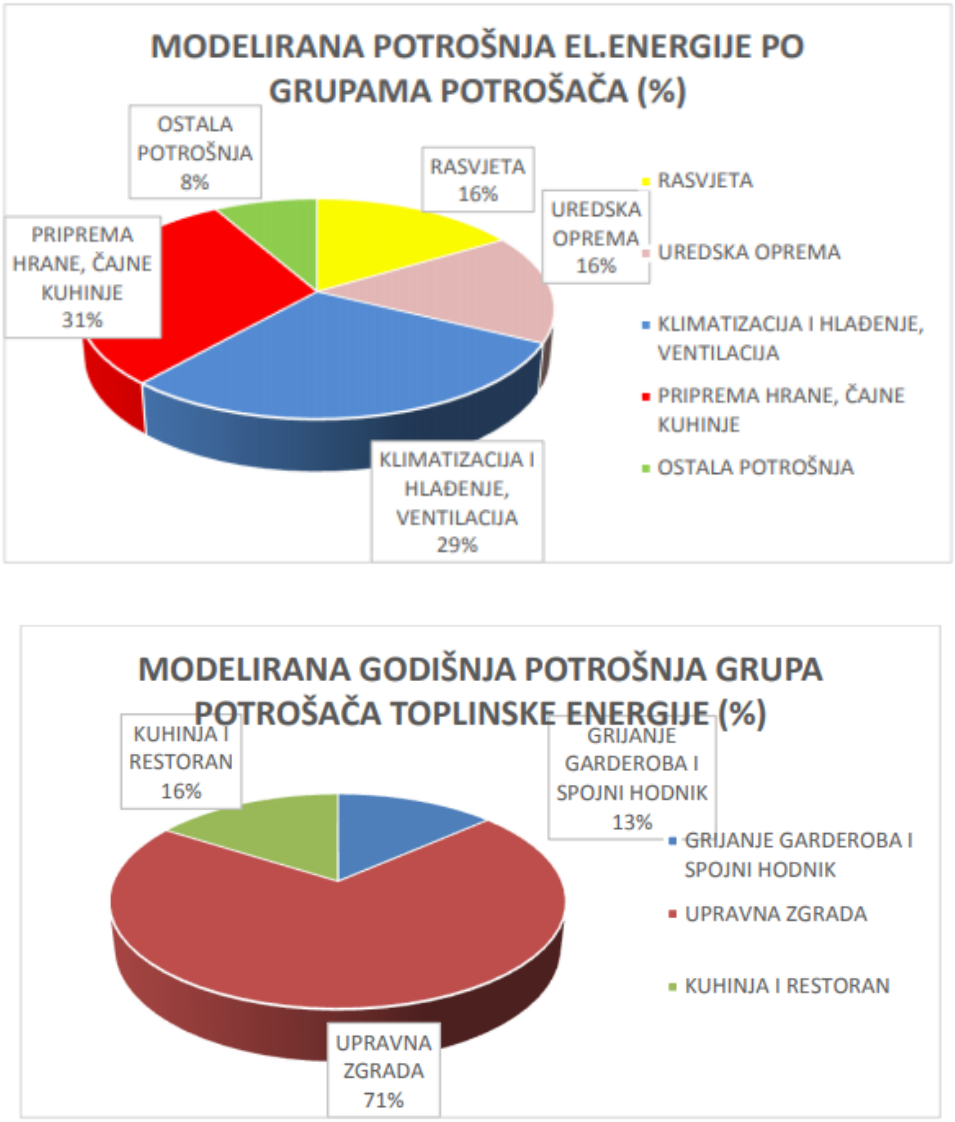
Tablica 5.3 Definicije korištenih fizikalnih veličina [11]

Oznaka	Naziv oznake	Definicija	Mjerna jedinica
UBP	Ukupna bruto površina građevine	Površina grijanog i negrijanog dijela građevine.	m^2
A_k	Ploština korisne površine grijanog dijela građevine	Ukupna ploština neto površine grijanog dijela građevine.	m^2



Slika 5.5 Broj etaža i ukupna bruto površina po lokacijama

Slika 5.5 prikazuje broj etaža i ukupnu bruto površinu objekata na različitim lokacijama. Lokacija L1 i L4 imaju veći broj etaža i ukupnu bruto površinu u odnosu na lokacije L2 i L3. Prema ovim podacima može se zaključiti da su lokacije L1 i L4 zgrade (uredski prostori), dok su lokacije L2 i L3 niže građevine (radionice). Upravo je zbog toga struktura troškova na energente slična između ovih objekata kako je uočeno u prethodnom poglavlju. Uredski prostori generiraju veće troškove električne energije, dok radionice generiraju veće troškove toplinske energije.



Slika 5.6 Struktura potrošača električne energije i toplinske energije na lokaciji L1

Slika 5.6 prikazuje strukturu potrošača električne i toplinske energije za lokaciju L1. Može se zaključiti da je struktura grupe potrošača električne energije puno kompleksnija od strukture grupe potrošača toplinske energije. U nastavku je detaljan opis potrošača za oba energenta.

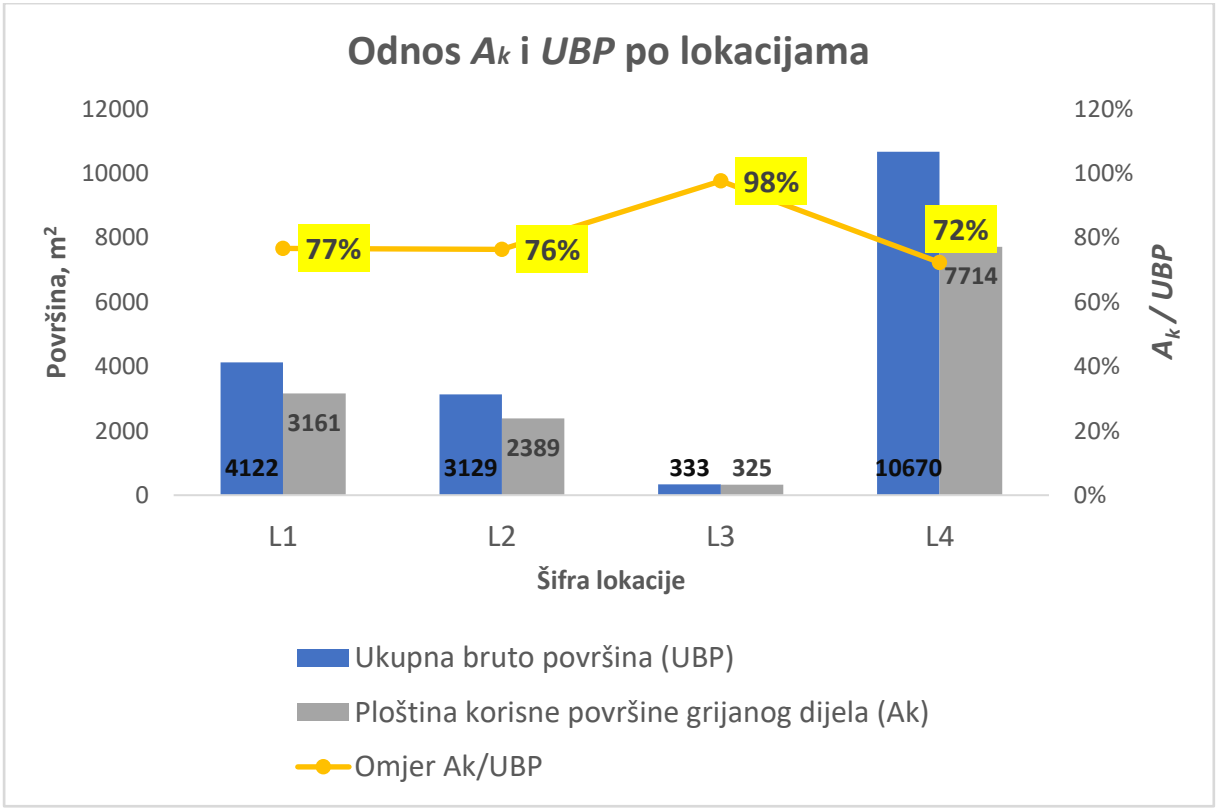
Grupe potrošača toplinske energije na lokaciji L1:

- radijatorsko grijanje upravne zgrade i kuhinje,
- grijanje grijača klima komora kuhinje i restorana,
- radijatorsko grijanje garderobe i spojnog hodnika.

Grupe potrošača električne energije na lokaciji L1:

- rasvjeta,
- uredska oprema,
- klimatizacija, ventilacija,
- oprema za pripremu hrane i napitaka,
- dizala i ostala trošila.

Slična struktura potrošača može se pretpostaviti i za lokaciju L4.



Slika 5.7 Odnos *UBP* i A_k po lokacijama

Slika 5.7 prikazuje odnos ukupne bruto površine objekata (*UBP*) i ploštine korisne površine grijanog dijela objekta (A_k). Može se zaključiti da su objekti na lokacijama L1, L2 i L4 gotovo identično projektirani jer je omjer UBP/A_k vrlo sličan i iznosi oko 75%. Lokacija L3 je iznimka i ima gotovo stopostotnu iskoristivost bruto površine objekta tj. gotovo cijela površina je grijana. Ovako visok postotak ovog omjera uzrokovan je time što objekt na lokaciji L3 ima samo jednu etažu što je vidljivo na dijagramu sa slike (Slika 5.5).

5.4. Analiza vremenske serije – lokacija L1 / toplinska energija

Uzimajući u obzir prethodnu analizu različitih lokacija, kompleksnost prikupljanja svih potrebnih podataka, te samu dostupnost i konzistentnost podataka unutar baze u koju se pohranjuju, za analizu i izradu predikcijskog modela odabrana je sljedeća kombinacija (Tablica 5.4).

Tablica 5.4 Odabrana kombinacija za analizu i predikcijski model

Lokacija:	**Lokacija 1 (L1)**
Energent:	**Toplinska energija**

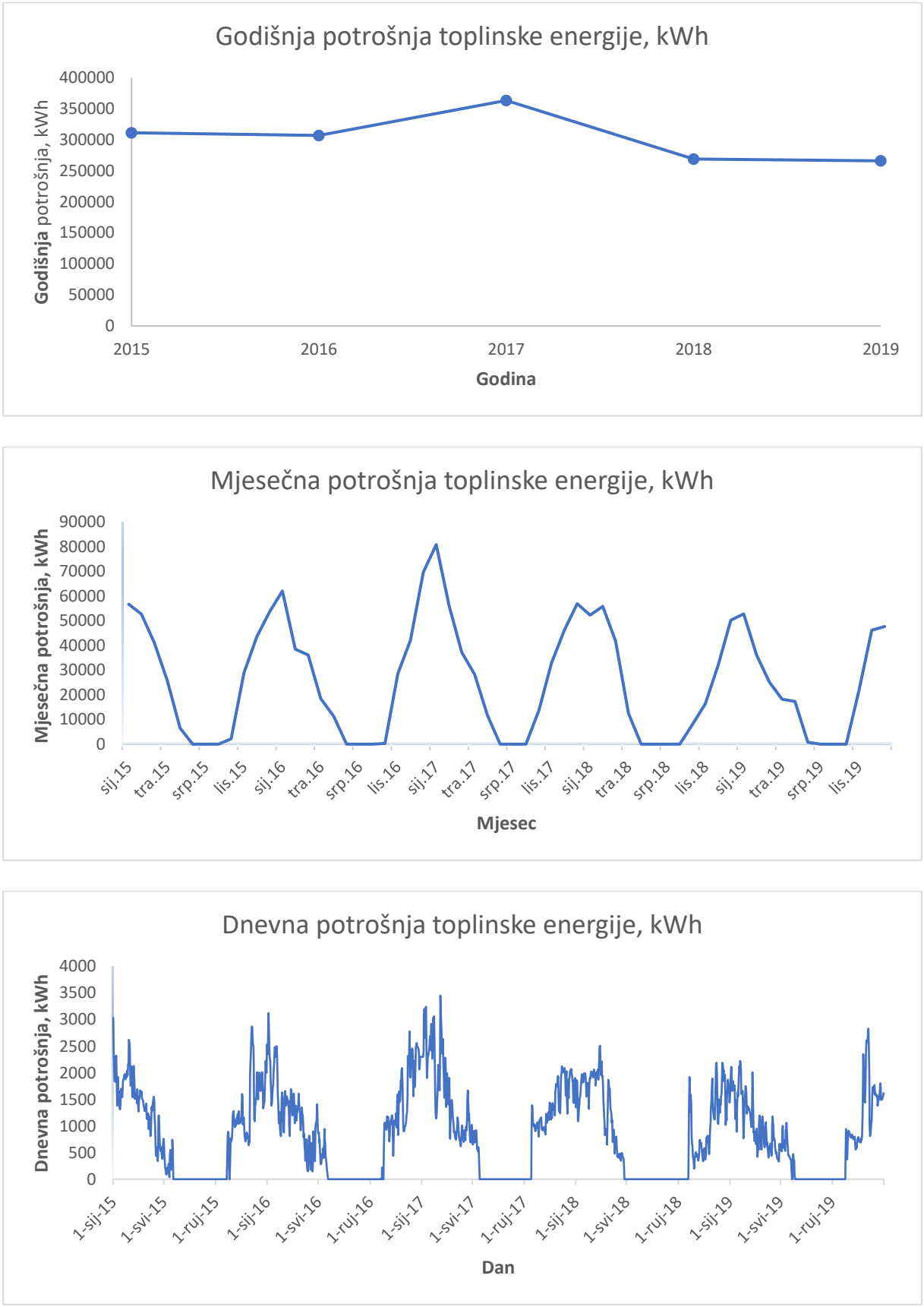
Obrazloženje odabira kombinacije L1 / toplinska energija:

- toplinska energija čini trećinu troškova energenata tijekom godine na lokaciji L1,
- toplinska energija intenzivno se koristi u kraćem periodu tijekom godine (~ tri kvartala), dok se električna energija koristi intenzivno tijekom cijele godine, odnosno drugim riječima uz manje intenzivnu eksploataciju daje vrlo visoke troškove, te bi se zbog toga fokus trebao usmjeriti na predviđanje potrošnje i održivo planiranje dobave iste,
- kompleksnost strukture grupa potrošača je manja kod toplinske energije – lakše praćenje, mjerenje i zapisivanje u bazu podataka – konzistentniji podaci,
- projektirani omjer A_k / *UBP* sličan je ostalim lokacijama,
- kvantitativna i kvalitativna struktura podataka o potrošnji u bazi podataka.

Osnovne osobine vremenske serije potrošnje toplinske energije koja će se koristiti za daljnji dio rada prikazane su u nastavku (Tablica 5.5).

Tablica 5.5 Osnovni podaci odabrane vremenske serije

Energent:	toplinska energija
Mjerna jedinica:	kWh
Raspon godina mjerenja (podataka):	01.01.2015. – 31.12.2019.
Moguće frekvencije mjerenja:	sat, dan, mjesec, godina



Slika 5.8 Usporedba vremenske serije s obzirom na frekvenciju mjerenja

Slika 5.8 prikazuje potrošnju toplinske energije na lokaciji L1. Prikazana tri grafa predstavljaju istu vremensku seriju, ali s različitom frekvencijom mjerenja.

Prvi graf predstavlja godišnju potrošnju toplinske energije. Može se reći da do 2020. nije vidljiv nikakav trend. Također, može se vidjeti da je maksimalna potrošnja toplinske energije bila tijekom 2017. godine.

Drugi graf predstavlja mjesečnu potrošnju toplinske energije. Iz ovog grafa može se vidjeti da je u vremenskoj seriji prisutna sezonalnost. Sezonalnost se javlja kada vremenski niz utječe na vrijednost promatrane varijable. Definira se kao pravilno odstupanje vrijednosti vremenske serije koje se opetovano ponavlja unutar nekog redovitog razdoblja. Iz promatranog grafa može se zaključiti da je taj fiksni, opetovani period 12 mjeseci, odnosno jedna godina. Unutar jedne godine može se primijetiti da sezonalnost ima oblik slova „U“.

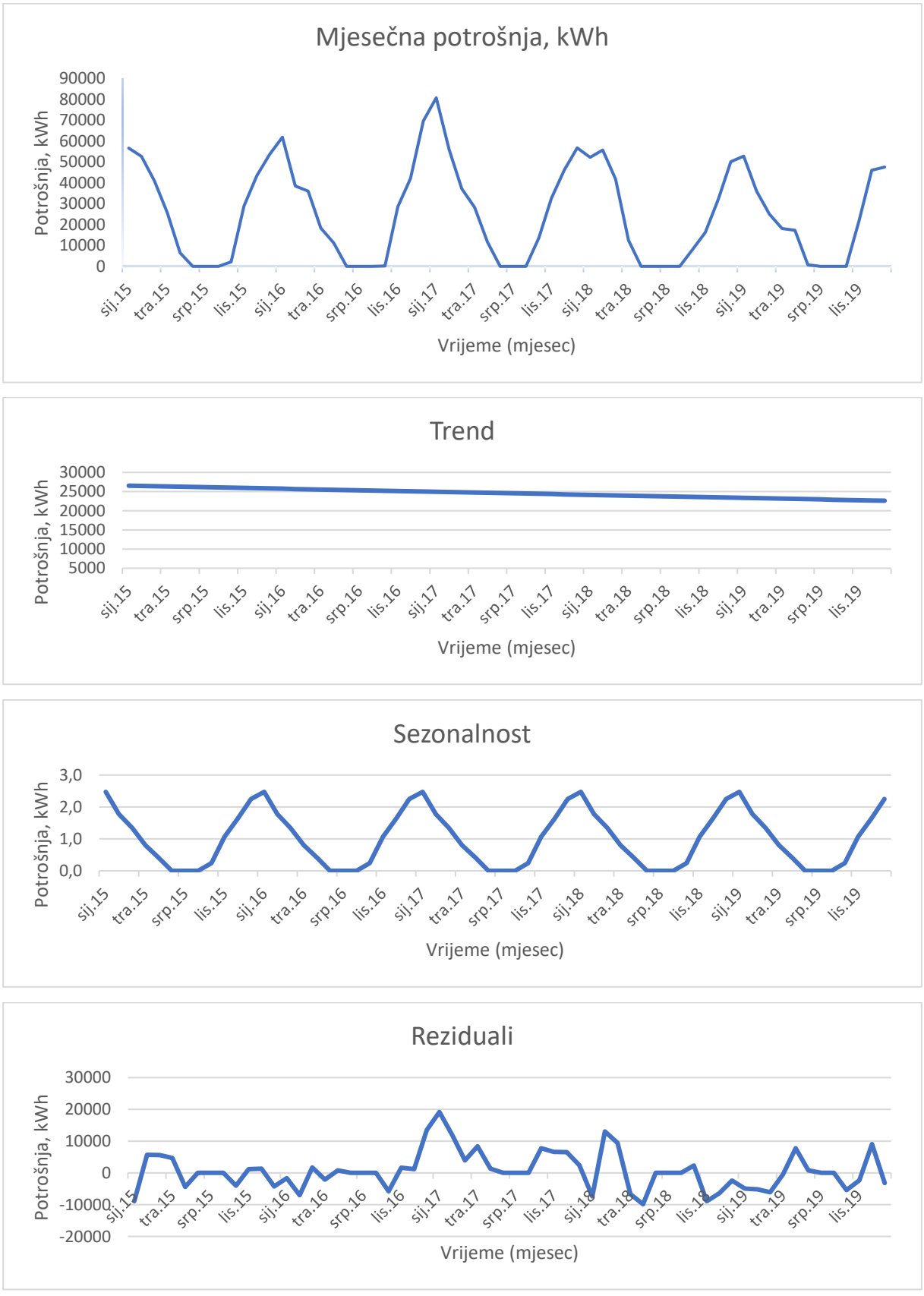
Treći graf predstavlja dnevnu potrošnju toplinske energije, a samim time najdetaljniji je prikaz izabrane vremenske serije. Vidi se da dnevna potrošnja prati oblik drugog grafa, odnosno mjesečne potrošnje, ali s određenim odstupanjima. Dnevna potrošnja koristit će se u daljnjem radu jer sadrži najveći broj mjerenja, odnosno podataka. Upravo zbog toga sadrži i najviše „skrivenog“ znanja.

5.4.1. Dekompozicija vremenske serije

Grafički prikazi omogućuju vizualizaciju mnogih značajki (osobina) podataka, uključujući obrasce, neuobičajene vrijednosti, promjene tijekom vremena i odnose između zabilježenih podataka (mjerenja). Upravo se zbog toga vrši dekompozicija vremenske serije kako bi se uočile i razdvojile prije spomenute komponente:

- trend,
- sezonalnost,
- rezidualne komponente.

Dekompozicija vremenske serije vršit će se na podacima mjesečne potrošnje toplinske energije kako bi se što jednostavnije otkrilo postoji li trend ili ne. Sezonalnost je već uočena na prijašnjim grafovima (Slika 5.8).



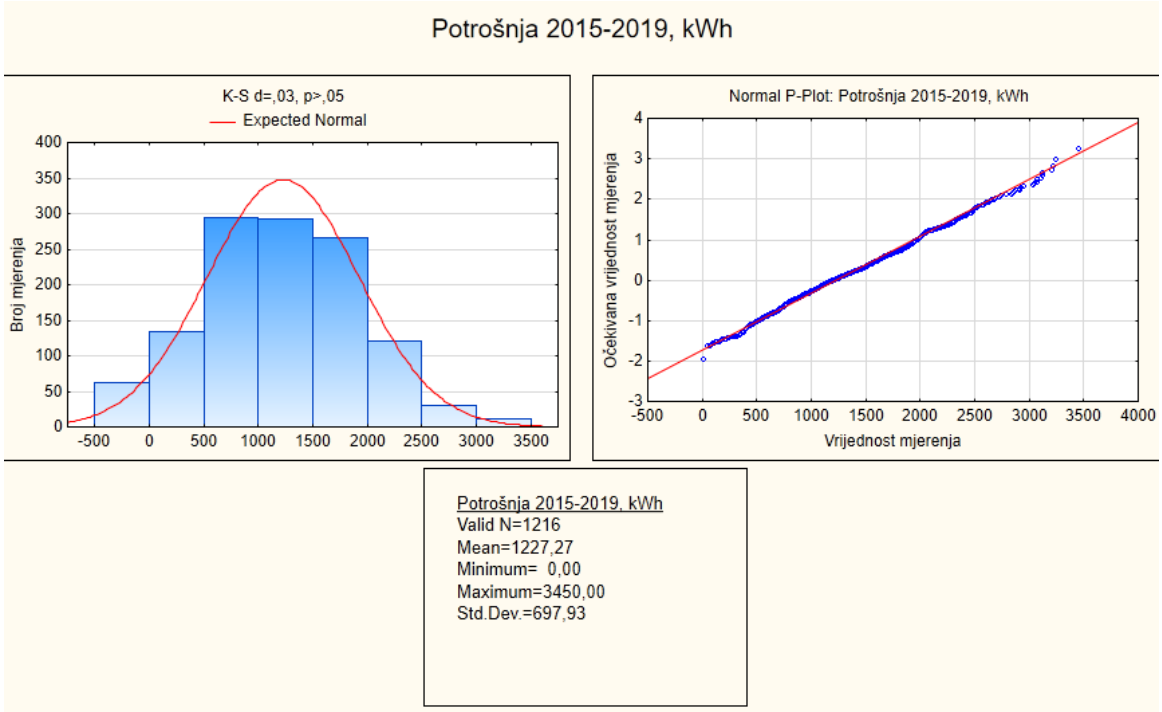
Slika 5.9 Dekompozicija vremenske serije (mjesečna potrošnja)

Slika 5.9 prikazuje multiplikativnu dekompozicija vremenske serije. Iz dekompozicije je vidljivo da postoji blagi trend opadanja, te je potvrđena komponenta sezonalnosti. Također, vidljivo je da su vrijednosti reziduala poprilično visoke što znači da unutar njihovog skupa postoji „znanje“ o ponašanju i karakteristikama vremenske serije. Zbog svega navedenog, pri modeliranju predikcije, morat će se koristiti određene transformacije kako bi samo predviđanje bilo što točnije. Ono što se svakako može iščitati iz ove dekompozicije jest da ova vremenska serija nije stacionarna.

5.4.2. Analiza varijance – ANOVA

Analiza varijance ili ANOVA (eng. *analysis of variance*) metoda je za analizu i provjeru razlika između različitih skupina iste pojave. Kako bi ova metoda bila primjenjiva, potrebno je da podaci zadovoljavaju uvjet normalnosti.

Analiza varijanci vršit će se nad prije definiranim skupom podataka, odnosno skupom podataka s frekvencijom dnevnog mjerenja (Tablica 5.5). Treba naglasiti da su se za test normalnosti isključila mjerenja između perioda lipanj-rujan jer se tada nije koristila toplinska energija. Mjerenja moraju zadovoljavati normalnu distribuciju kako bi se analiza varijanci mogla provesti.



Slika 5.10 Provjera normalnosti podataka

Slika 5.10 jasno prikazuje da se podaci (mjerenja) ponašaju po normalnoj razdiobi. Normalnost je vidljiva iz samog grafičkog prikaza, ali isti zaključak se može donijeti pogledom na *p* vrijednost Kolmogorov – Smirnov testa koja je veća od 0,05 tj. veća je od razine značajnosti $\alpha = 0{,}05$.

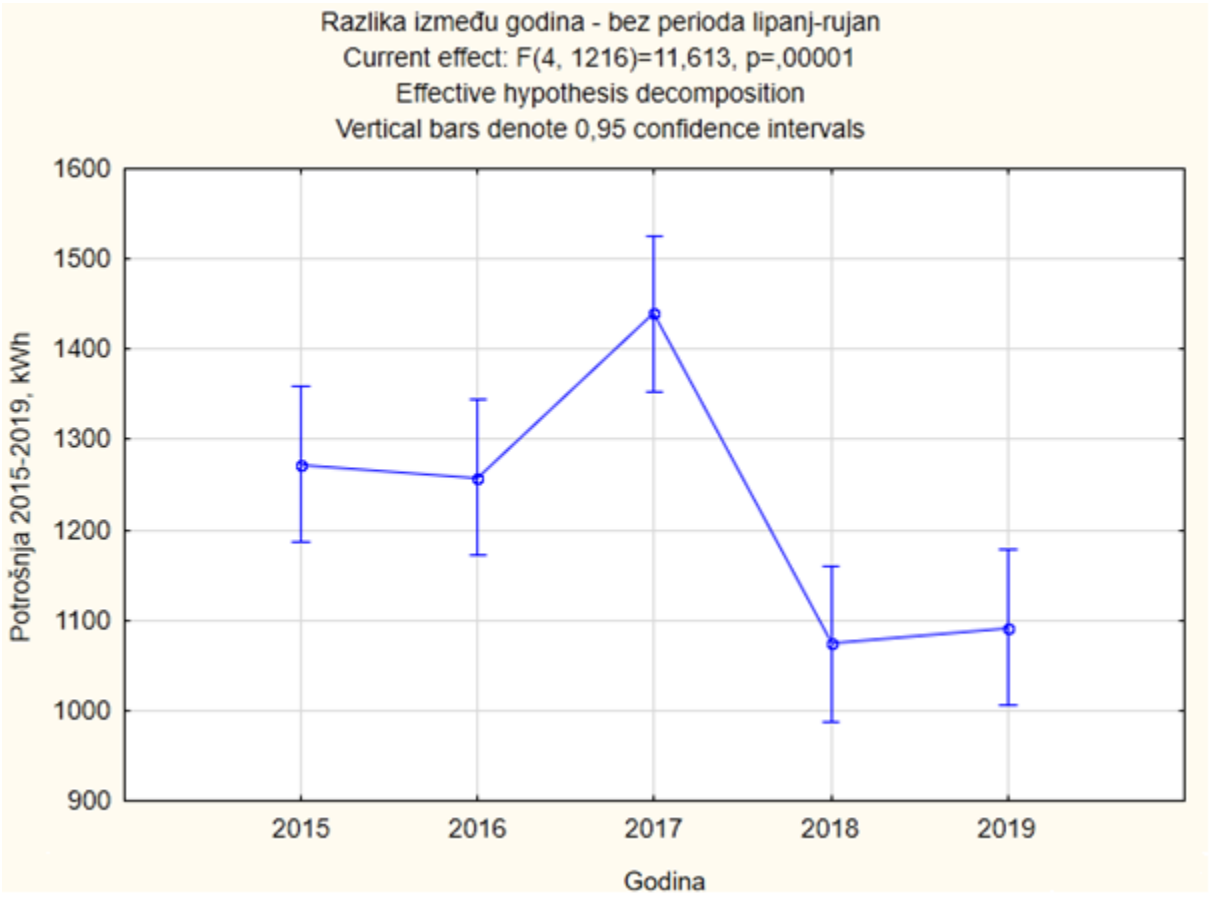
Ono što će se ispitivati analizom varijanci jest statistička razlika (ili sličnost) između različitih godina potrošnje toplinske energije. Zavisna varijabla su godine, a nezavisna ili prediktorska varijabla je potrošnja toplinske energije.

Prije provedbe same analize, postavit će se hipoteze:

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ → nema značajne razlike u potrošnji toplinske energije među godinama.

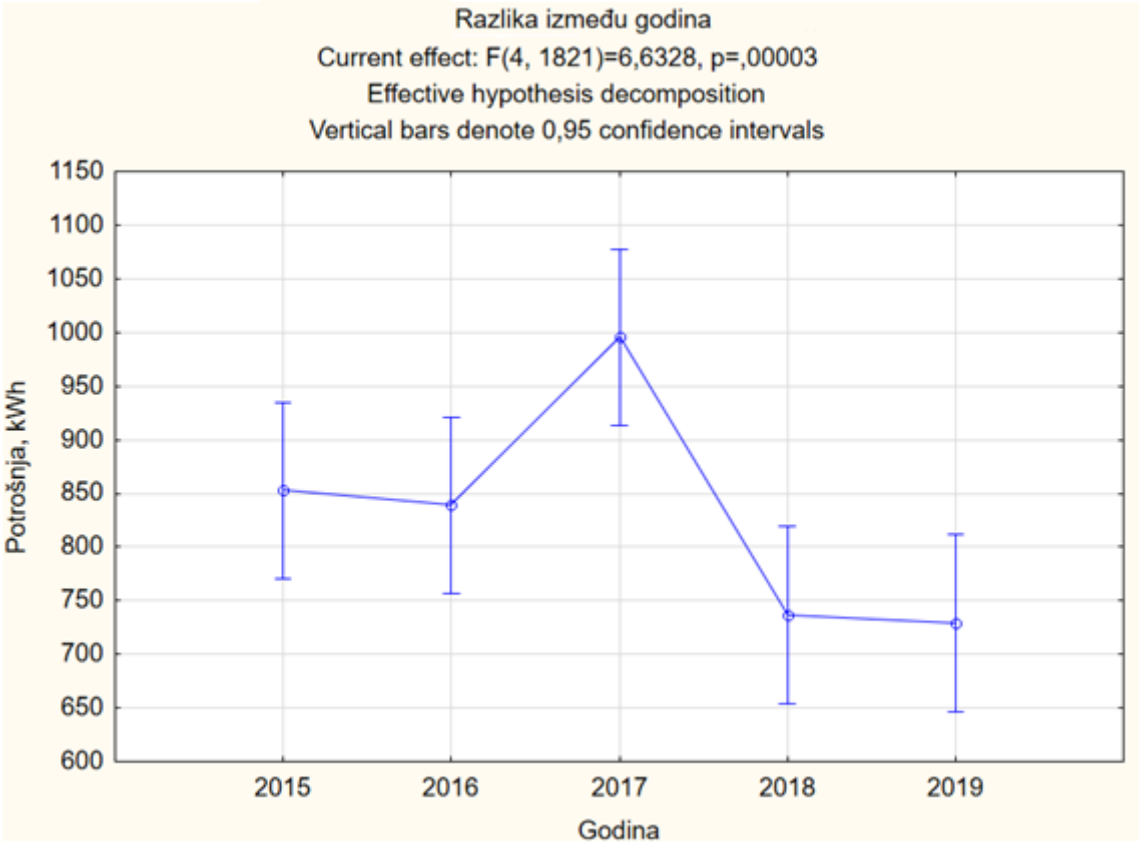
$H_1: \mu_i \neq \mu_{i+k}$ → postoji značajna razlika u potrošnji toplinske energije barem jedne godine.

Nakon provedene analize, dobiveni su sljedeći rezultati.



Slika 5.11 Razlika između godina po potrošnji toplinske energije – bez perioda lipanj-rujan

Slika 5.11 prikazuje dijagram razlika između godina po potrošnji toplinske energije, ali bez perioda lipanj – rujan jer su to mjerenja kada nije bilo potrošnje energije – 0 kWh. Dijagram je rezultat analize varijanci. Zaključak je da postoji značajna statistička razlika između ovih godina potrošnje jer je F-test pokazao vrijednost $p \approx 0$, što je manje od razine značajnosti $\alpha = 0{,}05$. Drugim riječima, postoji barem jedna godina koja je statistički značajno različitija od ostalih. S grafa se može naslutiti da je riječ o 2017. godini kada je potrošnja toplinske energije bila povišena.



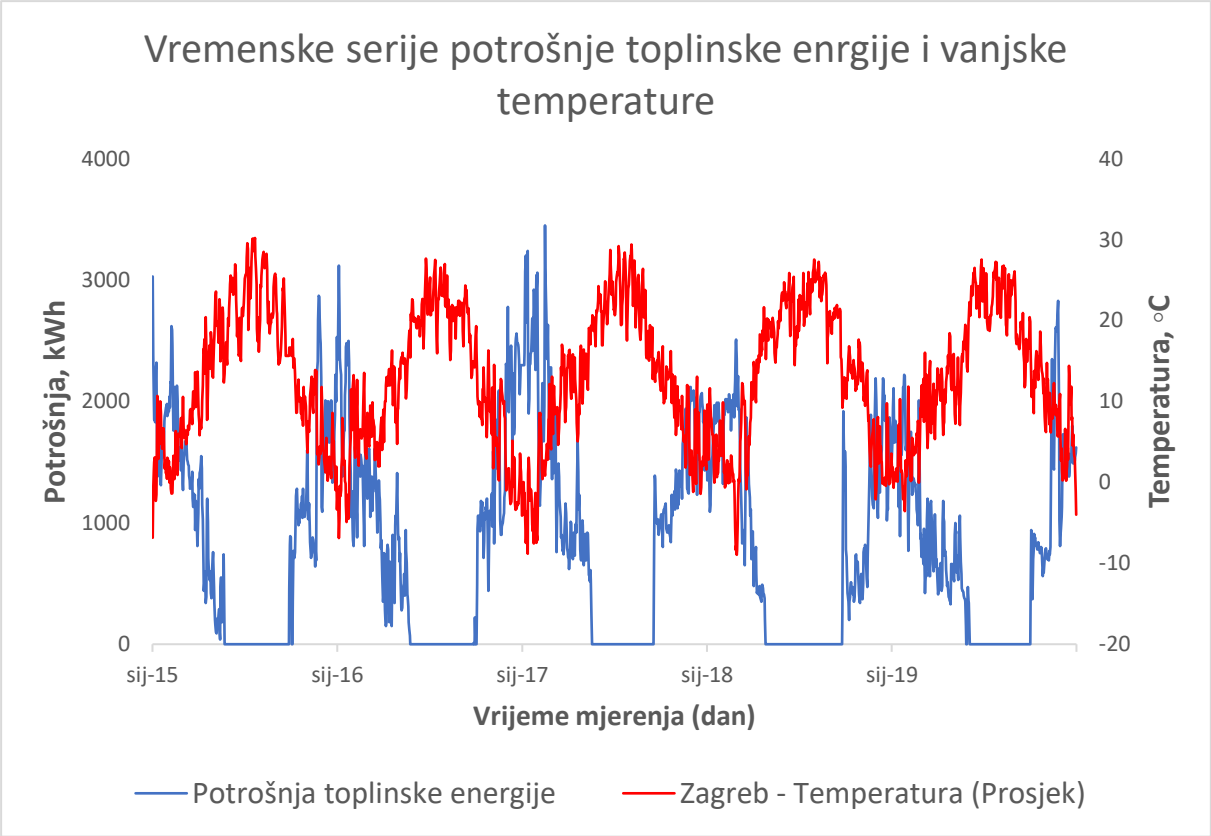
Slika 5.12 Razlika između godina po potrošnji toplinske energije – s periodom lipanj-rujan

Slika 5.12 prikazuje dijagram razlika između godina po potrošnji toplinske energije, ali s uključenim periodom lipanj-rujan. Iz ovog, ali ujedno i prethodnog dijagrama, može se vidjeti kako su godine 2015. i 2016., te 2018. i 2019. vrlo slične. Međutim godina 2017. značajno je različitija od njih te je zbog toga F test još jednom pokazao da postoji statistički značajna razlika među godinama i potrošnjom toplinske energije. Analiza varijanci donijela je dodatne informacije o podacima koji će se koristiti za izradu predikcijskog modela. Statistička razlika između godina i potrošnje unijet će dodatnu kompleksnost pri izradi predikcijskih modela.

5.4.3. Korelacijska analiza

Korelacijska analiza je metoda kojom se utvrđuje postoji li među varijablama funkcionalna ovisnost, odnosno to je mjera povezanosti dviju ili više varijabli. Korelacija se iskazuje pomoću Pearsonovog koeficijenta korelacije r i može poprimiti vrijednost od -1.00 do +1.00. Ukoliko je $r < 0$ definira negativnu korelaciju, odnosno ako je $r > 0$ definira pozitivnu korelaciju. Pozitivna korelacija jest ona u kojoj vrijednosti promatranih varijabli opada ili raste u istom smjeru, odnosno ako vrijednost jedne varijable raste, raste vrijednost i druge varijable (isto vrijedi za opadanje vrijednosti). Negativna korelacija predstavlja suprotan rast vrijednosti promatranih varijabli, odnosno rastom vrijednosti jedne varijable, vrijednost druge opada, te obratno.

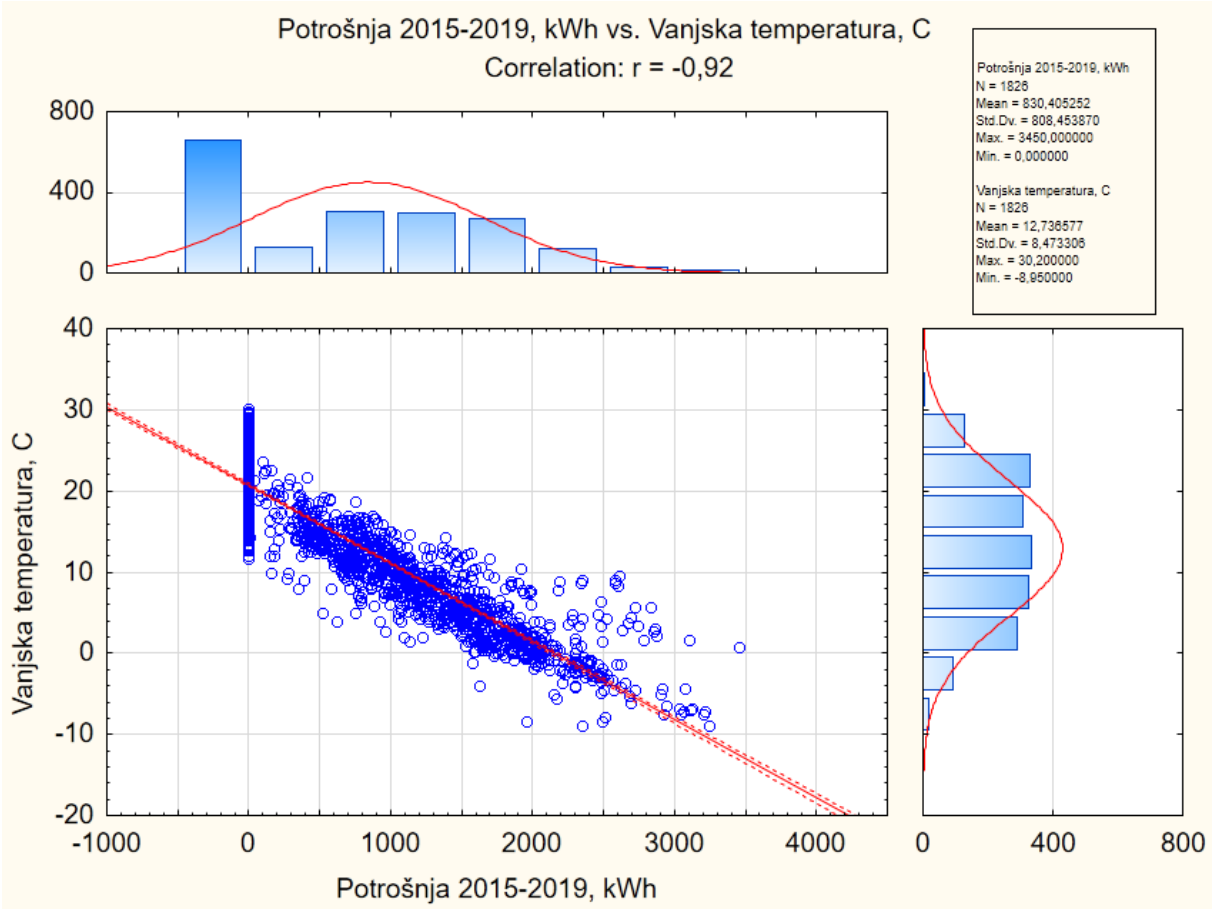
U nastavku će biti prikazana korelacijska analiza između potrošnje toplinske energije i vrijednosti vanjske temperature. Analiza će se vršiti na setu podataka dnevnih mjerenja. Pretpostavka je da će postojati negativna korelacija, odnosno ukoliko je vanjska temperatura niža, potrošnja toplinske energije bit će viša.



Slika 5.13 Vremenske serije potrošnje toplinske energije i vanjske temperature

Slika 5.13 prikazuje vremenske serije potrošnje toplinske energije i vanjske temperature. Vidljivo je kako su obrnuto proporcionalne, odnosno kada je vanjska temperatura na vrhuncu, potrošnja toplinske energije je vrlo mala ili uopće ne postoji. Također, vidljivo je kako obje vremenske serije imaju komponentu sezonalnosti – oblik slova „U“.

Nakon vizualne analize ovih dviju varijabli, potrebno je provesti statistički test korelacije kako bi se ona utvrdila i dobio krajnji oblik iste.



Slika 5.14 Korelacijska analiza – rezultati

Slika 5.14 prikazuje rezultate korelacijske analize. Kao što se i pretpostavilo, između ove dvije varijable postoji negativna korelacija koja prema Pearsonovom koeficijentu ima vrijednost r = – 0,92. Drugim riječima, što je vanjska temperatura niža, to je potreba i potrošnja toplinske energije sve veća.

Informacije o korelaciji bit će od koristi pri modeliranju kompleksnijih modela za predviđanje buduće potrošnje toplinske energije s obzirom na period u godini.

6. PREDIKCIJSKI MODELI

Nakon prikupljanja, strukturiranja i analiziranja podataka na red dolazi izrada predikcijskih modela. Predviđanje vremenskih serija vrši se s ciljem predviđanja, odnosno prognoze budućih vrijednosti vremenske serije. U ovom slučaju, predviđat će se buduća potrošnja toplinske energije. U nastavku će biti prikazano više predikcijskih modela, te će se na kraju usporediti njihove performanse i odabrati najbolji.

6.1. Model eksponencijalnog zaglađivanja

Prema poglavljima *3.2 i 5.4.1,* zbog blagog trenda i sezonalne komponente, odabire se aditivna Holt-Wintersova metoda eksponencijalnog zaglađivanja. Aditivna varijanta je odabrana zbog vrijednosti mjerenja koje iznose 0 kWh (nema potrošnje toplinske energije) jer multiplikativnu varijantu nije moguće izvesti ukoliko postoje takve vrijednosti.

6.1.1. Izrada modela eksponencijalnog zaglađivanja

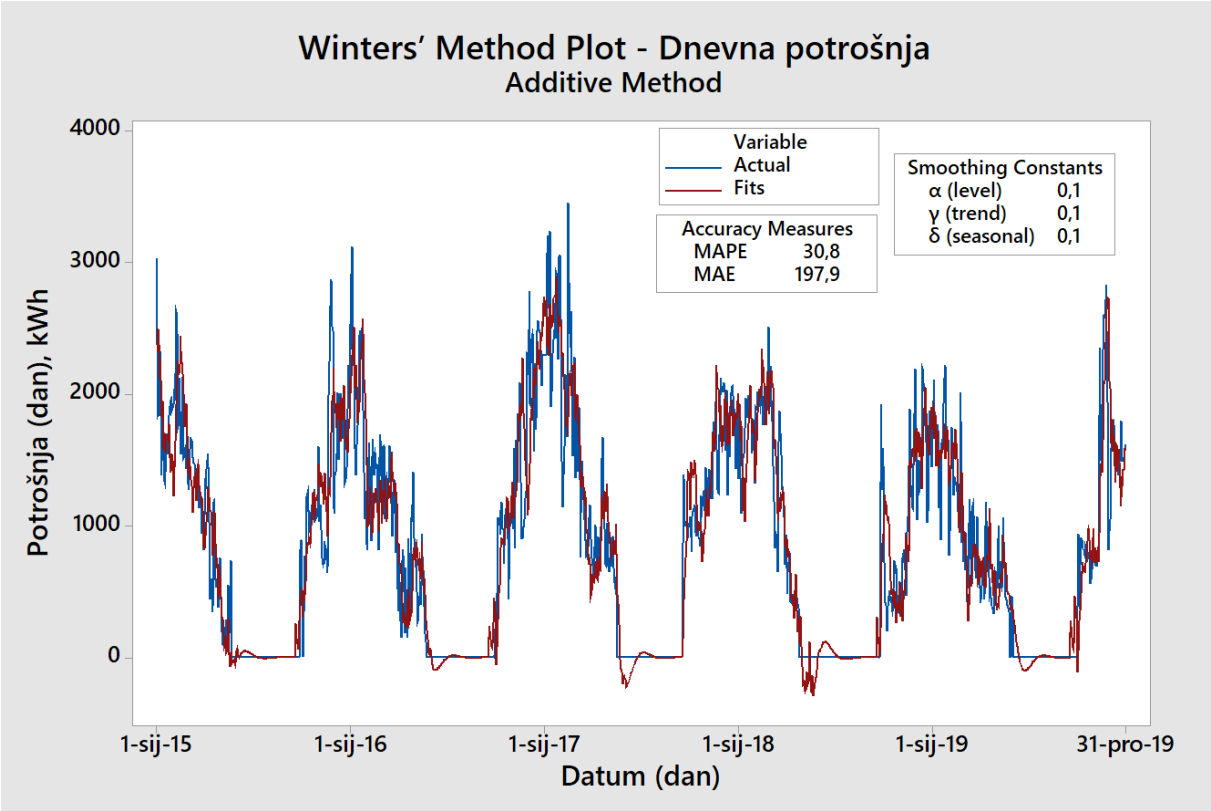
- **Model: $\alpha = \beta = \gamma = 0{,}1$**

Za inicijalni model bit će korišteni početni parametri $\alpha = \beta = \gamma = 0{,}1$. Skup podataka za učenje definiran je sljedećom tablicom:

Tablica 6.1 Skup podataka (mjerenja) za učenje modela

Energent:	toplinska energija
Mjerna jedinica:	kWh
Raspon godina mjerenja (podataka):	01.01.2015. – 31.12.2019.
Ukupan broj mjerenja:	1826
Frekvencije mjerenja:	dan

Prvotni model koristit će se kao referentni model za buduća unapređenja, odnosno optimizaciju parametara α, β i γ.



Slika 6.1 Model: $\alpha = \beta = \gamma = 0{,}1$

Slika 6.1 predstavlja izgled inicijalnog modela. Plava krivulja predstavlja stvarne izmjerene vrijednosti, dok crvena krivulja predstavlja prilagođeni model nastao korištenjem prije spomenutih parametara. Vidljivo je kako model, odnosno crvena krivulja prati kretanje plave krivulje odnosno stvarnih mjerenja.

Mjere točnosti modela koje će se koristiti su MAPE i MAE.

Inicijalni model ima sljedeću točnost:

- MAPE = 30,8%
- MAE = 197,9 kWh

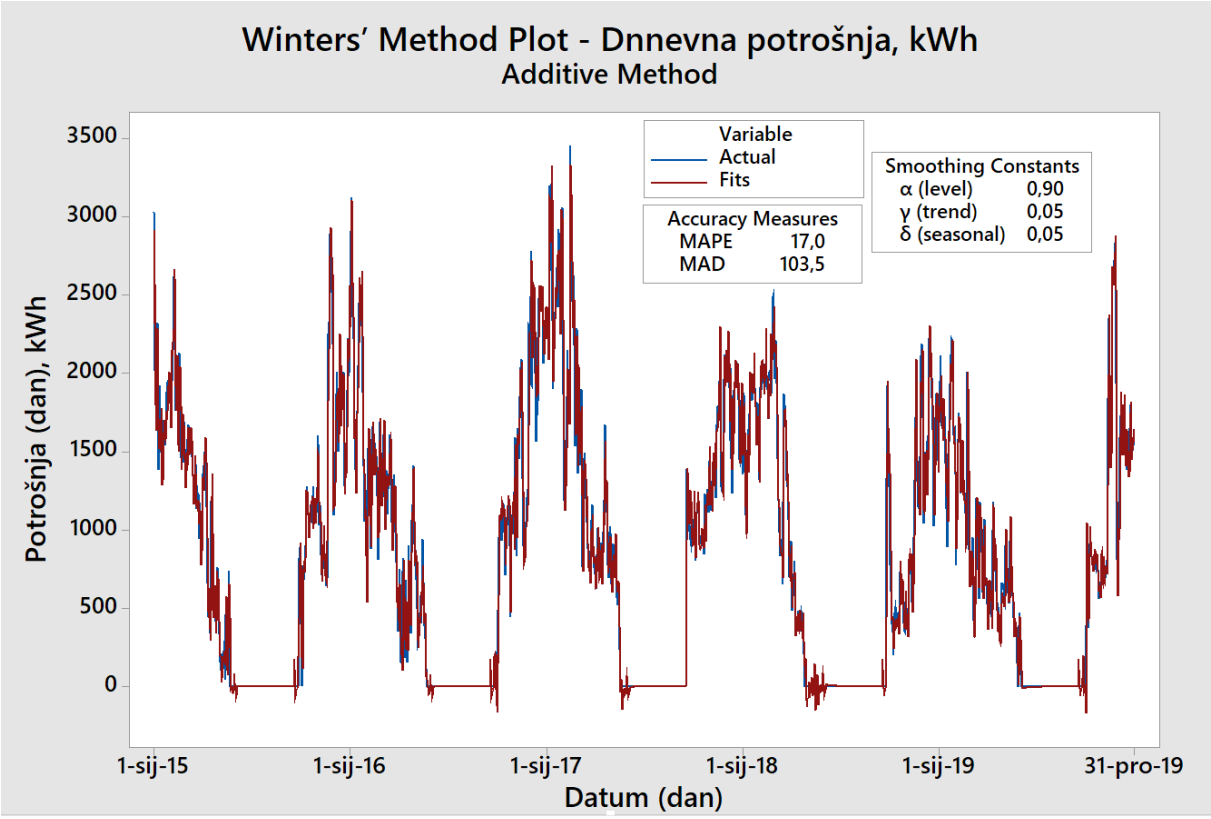
Prema definicijama iz poglavlja *3.4*, prosječno apsolutno odstupanje kreiranog modela od stvarnog iznosi 197,9 kWh, drugim riječima prosječno postotno apsolutno odstupanje kreiranog modela iznosi 30,8%.

Modelu je potrebno optimizirati parametre α, β i γ kako bi se dobio točniji model.

- **Model: α = 0,9 ; β = 0,05 ; γ = 0,05**

Nakon optimizacije parametara pomoću metode pretraživanja mreže (eng. *grid search*), dobile su se sljedeće vrijednosti parametara modela: α = 0,9 ; β = 0,05 ; γ = 0,05.

Ove vrijednosti parametara mogu se interpretirati na sljedeći način. Vrijednost parametra koji predstavlja težinu mjerenja (eng. *weight*) α = 0,9 ukazuje na to da će se pri predviđanju u modelu veća značajnost dati recentnijim mjerenjima unutar povijesnih podatka. Niska vrijednost parametra β = 0,05 koji govori o trendu unutar vremenske serije, ukazuje na to da se pri predviđanju trenda u obzir mora uzeti veća količina povijesnih mjerenja. Ovako niska vrijednost ovog parametra ima smisla jer je trend vrlo blag i vidljiv je tek kada se u obzir uzme veća količina historijskih podataka, odnosno mjerenja. Parametar γ = 0,05 koji govori o sezonalnosti također ima nisku vrijednost. Razlog je isti kao kod parametra β koji daje informaciju o trendu.



Slika 6.2 Model: α = 0,9 ; β = 0,05 ; γ = 0,05

Slika 6.2 prikazuje model s optimiziranim vrijednostima parametara. Plava krivulja predstavlja stvarne izmjerene vrijednosti, dok crvena krivulja predstavlja prilagođeni model nastao korištenjem optimiziranih parametara. Vidljivo je kako model, odnosno crvena krivulja prati kretanje plave krivulje odnosno stvarnih mjerenja.

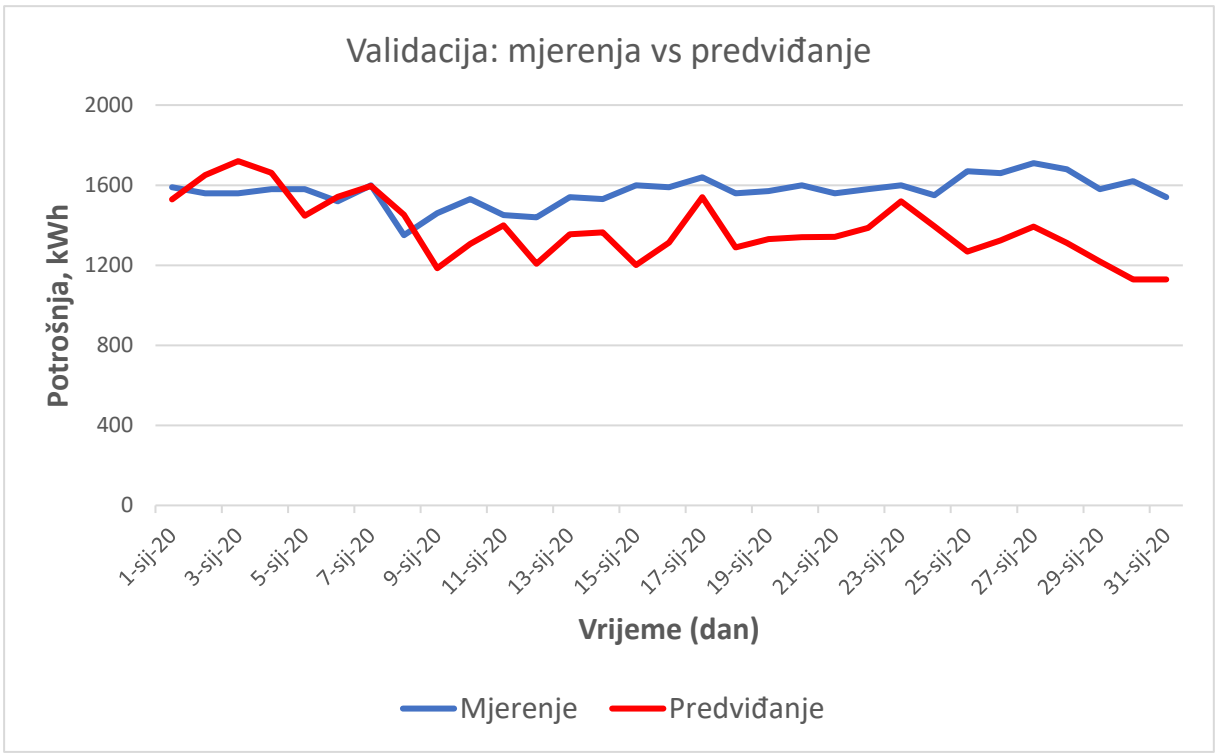
Optimizirani model ima sljedeću točnost:

- MAPE = 17,0 %
- MAE = 103,5 kWh

Optimizacijom parametara modela došlo je do poboljšanja modela. Mjera točnosti prosječnog postotnog apsolutnog odstupanja MAPE poboljšala se za 30,8% u odnosu na inicijalni model, dok se prosječno apsolutno odstupanje smanjilo za 48%.

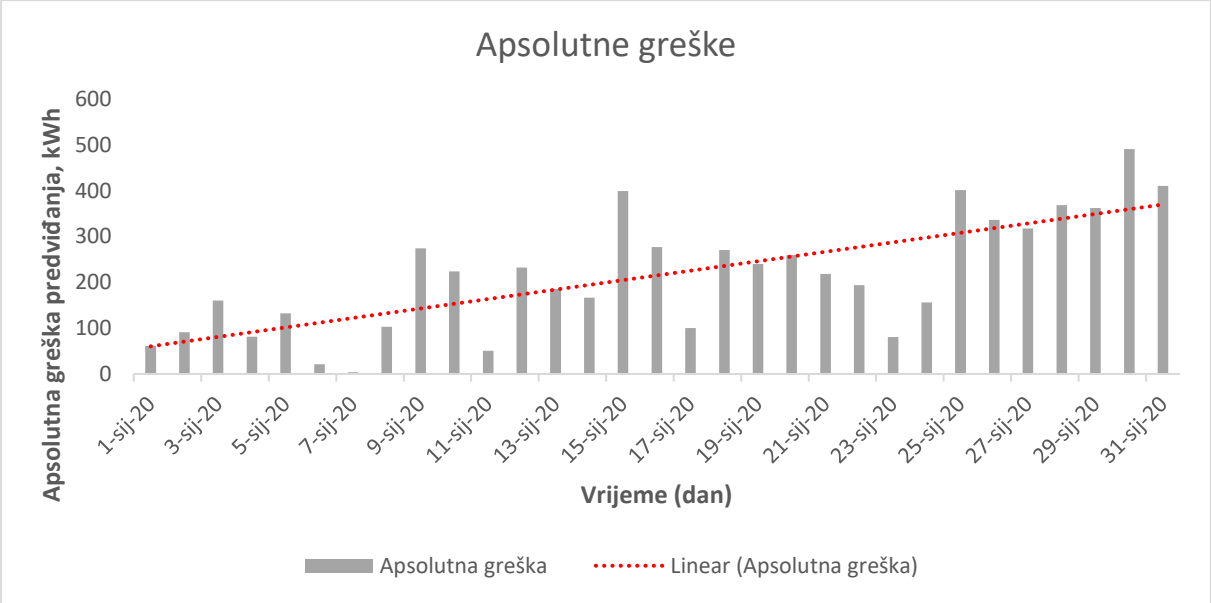
6.1.2. Validacija modela

U nastavku će se optimizirani model (α = 0,9 ; β = 0,05 ; γ = 0,05) validirati na mjerenjima nastalim u periodu od 1.1.2020. do 31.1.2020.



Slika 6.3 Validacija optimiziranog modela

Slika 6.3 prikazuje vrijednosti stvarnih mjerenja u periodu od 1.1.2020. do 31.1.2020. te predviđenih vrijednosti kreiranih optimiziranim modelom. Vidljivo je kako su trendovi i kretanje krivulja u istom smjeru što je pozitivan znak za validaciju modela.



Slika 6.4 Analiza apsolutnih grešaka modela

Slika 6.4 ukazuje na to kako s protekom vremena dolazi do povećavanja apsolutne greške što je bilo za očekivati jer je Holt-Winters metoda pogodna za kratkoročno predviđanje, odnosno za predviđanje manjeg broja koraka.

Mjera točnosti kreiranog predviđanja je sljedeća:

- MAPE = 13,6 %,
- MAE = 115, 1 kWh.

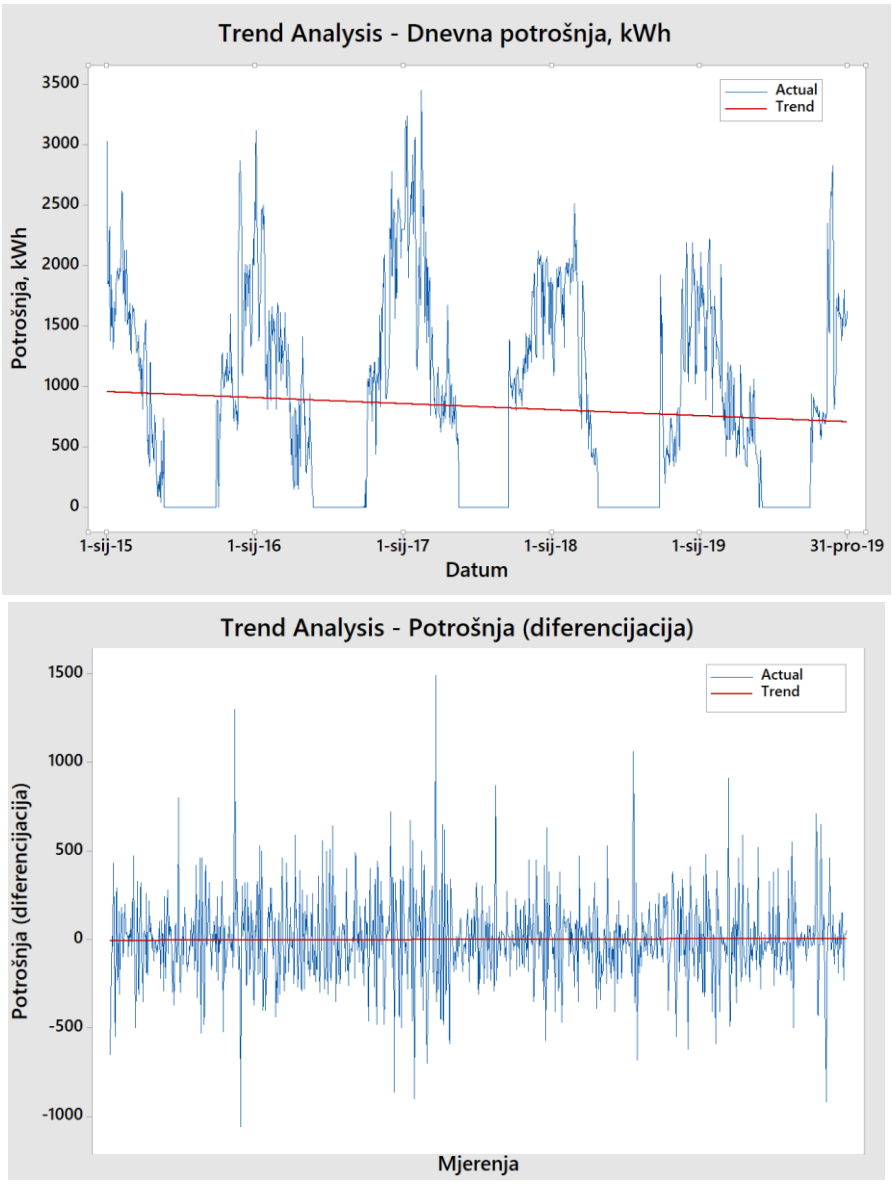
Model je uspješno prošao validaciju jer je prosječno postotno apsolutno odstupanje (MAPE) i prosječno apsolutno odstupanje (MAE) predviđene potrošnje toplinske energije manje od istih mjera u Holt-Winters optimiziranom modelu nastalom iz testnog skupa.

6.2. SARIMA model

Cilj za SARIMA model je odrediti parametre $(p,d,q)(P,D,Q)_m$ koji će najtočnije predviđati buduća stanja, odnosno potrošnju energije.

6.2.1. Izrada SARIMA modela

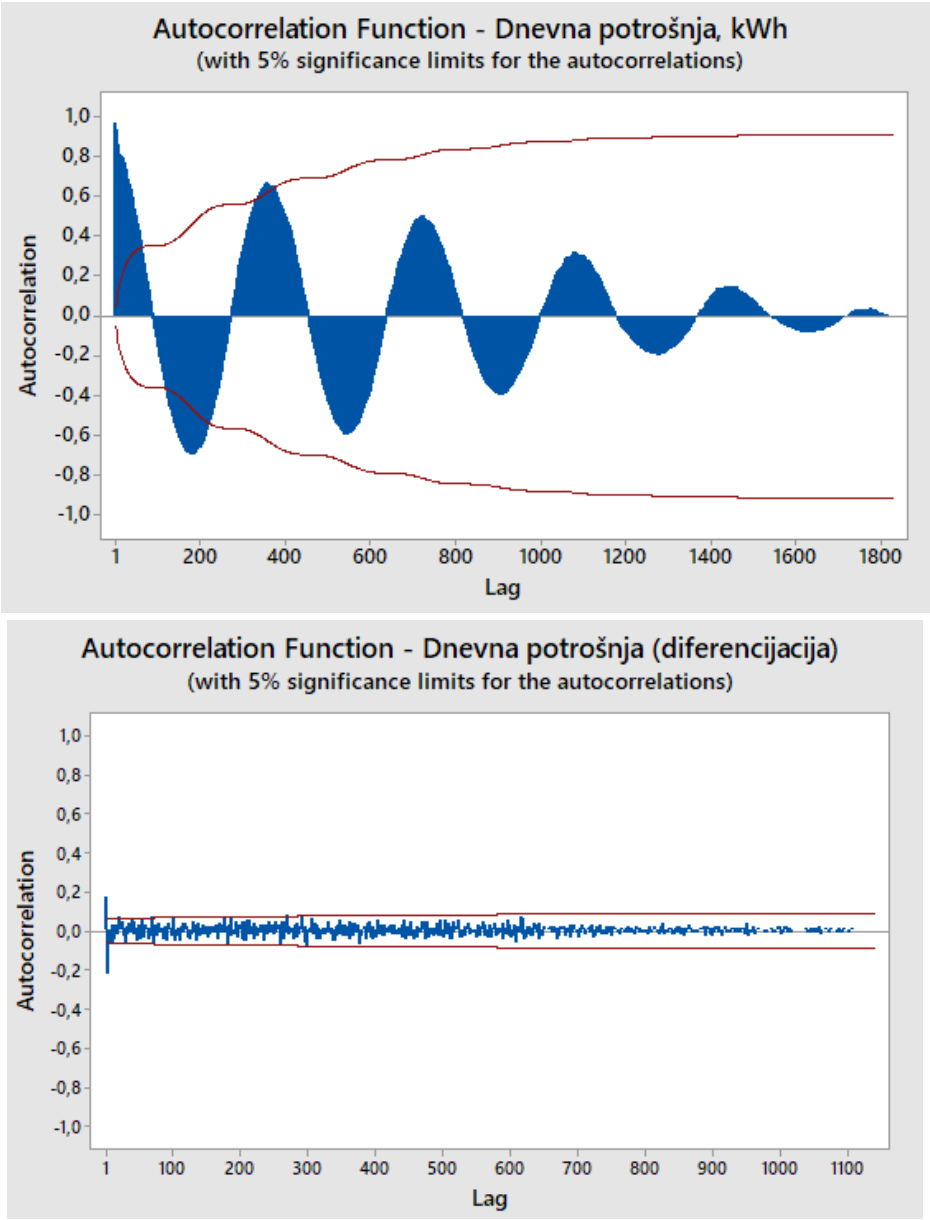
Prema poglavljima *3.3* i *5.4.1* zbog blagog trenda i sezonalne komponente, poznato je da vremenska serija nije stacionarna te će se zbog toga morati izvršiti transformacija podataka kako bi se utjecaji ovih komponenti na vremensku seriju i predviđanje buduće potrošnje toplinske energije maksimalno smanjili.



Slika 6.5 Diferencijacija mjerenja – uklanjanje trenda i sezonalnosti

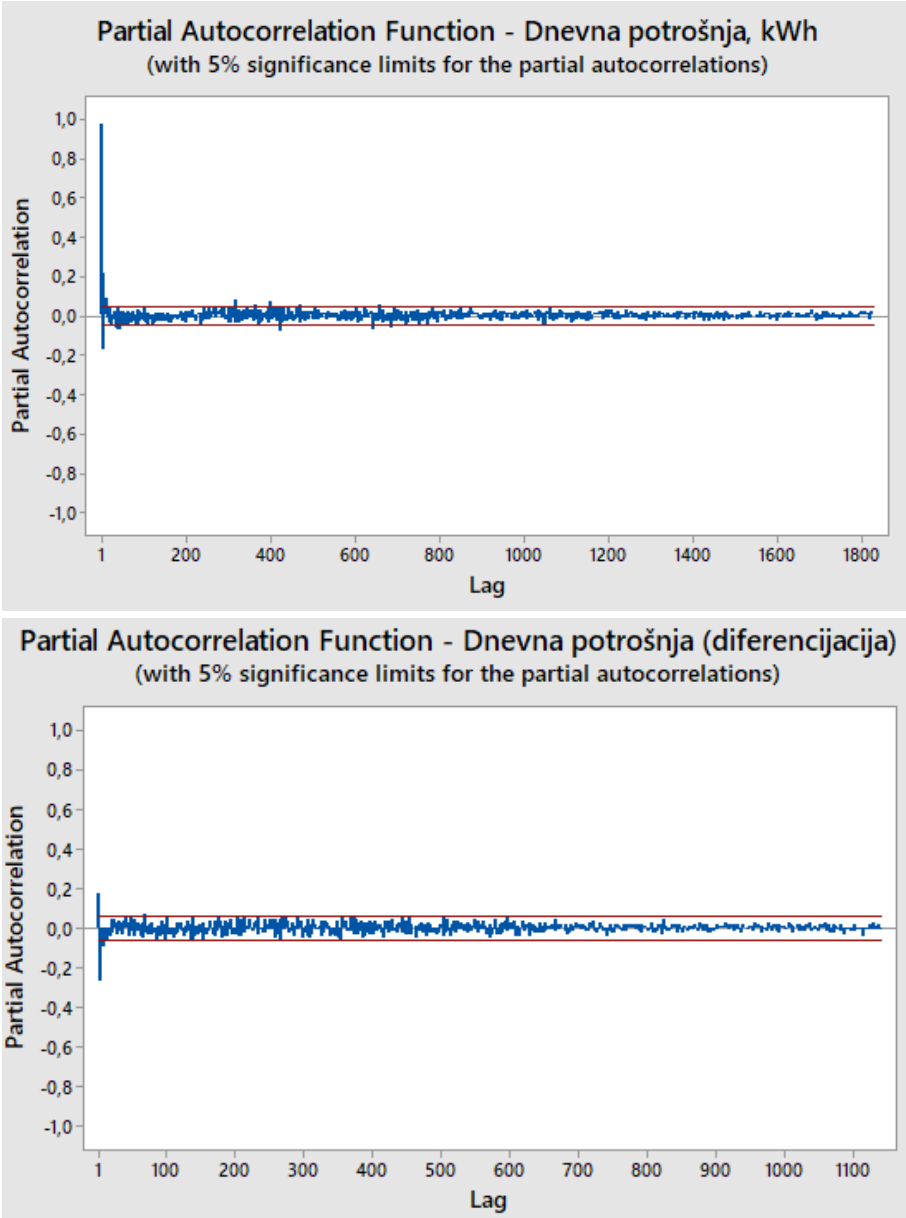
Slika 6.5 prikazuje vremensku seriju prije i nakon postupka diferencijacije. Prvi graf predstavlja vremensku seriju potrošnje toplinske energije (plava krivulja) i njezin trend (crveni pravac). Uz trend vidljiva je i prije spomenuta komponenta sezonalnosti. Drugi graf na istoj slici prikazuje rezultat postupka diferencijacije vremenske serije. Vidljivo je kako su komponente trenda i sezonalnosti uklonjene, a samim time vremenska serija je pogodna za izradu SARIMA modela. Također, diferencijacijom se riješio problem različitih varijanci opisan u poglavlju *5.4.2*.

Uklanjanjem trenda i sezonalnosti jasno je da će postojati diferencijacija u SARIMA modelu, odnosno otkriveni su parametri $d = 1$, i $D = 1$. Sljedeći korak je pronaći preostale parametre modela, a za njih potrebno je provjeriti izgled autokorelacijskih funkcija – ACF i PACF.



Slika 6.6 Usporedba ACF dijagrama

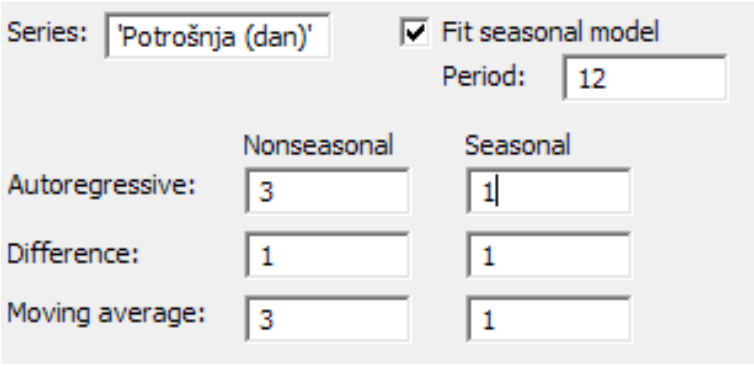
Slika 6.6 prikazuje usporedbu ACF funkcija. Prvi dijagram predstavlja autokorelacijsku funkciju vremenske serije bez transformacije. Vidljivo je kako postoji snažna sezonalna komponenta što se očituje geometrijskim opadanjem vrijednosti autokorelacije (naizmjenična promjena vrijednosti iz pozitivne u negativnu autokorelaciju). Na drugom grafu vidljiva je funkcija autokorelacije nakon transformacije podataka (diferencijacija). Vidljivo je kako je uklonjena sezonalna komponenta, ali isto tako je vidljivo kako postoji nagli skok autokorelacijskih vrijednosti na samom početku vremenske serije. Ovo je pokazatelj da će se u SARIMA model morati uključiti MA(q) model – q parametar će se testirati na vrijednosti od 1 do 3.



Slika 6.7 Usporedba PACF dijagrama

Slika 6.7 prikazuje usporedbu PACF funkcija. Za razliku od ACF funkcija, za PACF funkcije prije i nakon transformacije dijagrami imaju vrlo sličan izgled. PACF funkcija ima značajan skok na početku vremenske serije koji je ublažen transformacijom, te je to pokazatelj da će u SARIMA model morati biti uključena komponenta AR(p) – p parametar će se testirati kao i q parametar na vrijednosti od 1 do 3, te će se odabrati onaj koji rezultira najboljim modelom.

Nakon analize autokorelacijskih funkcija, potrebno je odrediti optimalne parametre modela.



Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	1	0,2255	0,2181	1,03	0,301
AR	2	0,1332	0,1761	0,76	0,450
AR	3	0,1645	0,0732	2,25	0,025
SAR	12	-0,0092	0,0243	-0,38	0,706
MA	1	0,0144	0,2170	0,07	0,947
MA	2	0,3974	0,1290	3,08	0,002
MA	3	0,3356	0,0732	4,59	0,000
SMA	12	0,9880	0,0003	3287,13	0,000

Slika 6.8 Optimizacija parametara modela

Za model je prihvaćen prvi stupanj diferencijacije. Parametri $d = 1$ i $D = 1$ nisu se ispitivali dalje jer su uspješno eliminirali trend i sezonalnost, a njihovim rastom kompleksnost modela bi se povećala, a rezultati bi bili netočniji.

Za interpretaciju dobivenih rezultata potrebno je objasniti kriteriji odabira optimalnih parametara. Ključnu ulogu ima parametar P (Slika 6.8).

Za parametar *P* vrijede sljedeća pravila [12]:

- $P \leq \alpha$ → koeficijent SARIMA modela je značajan i usvaja se,
- $P > \alpha$ → koeficijent SARIMA modela nije značajan i ne usvaja se.

Odabrani koeficijent značajnosti α iznosi $\alpha = 0{,}05$.

Prema rezultatima najmanju vrijednost imaju sljedeći koeficijenti (Slika 6.8):

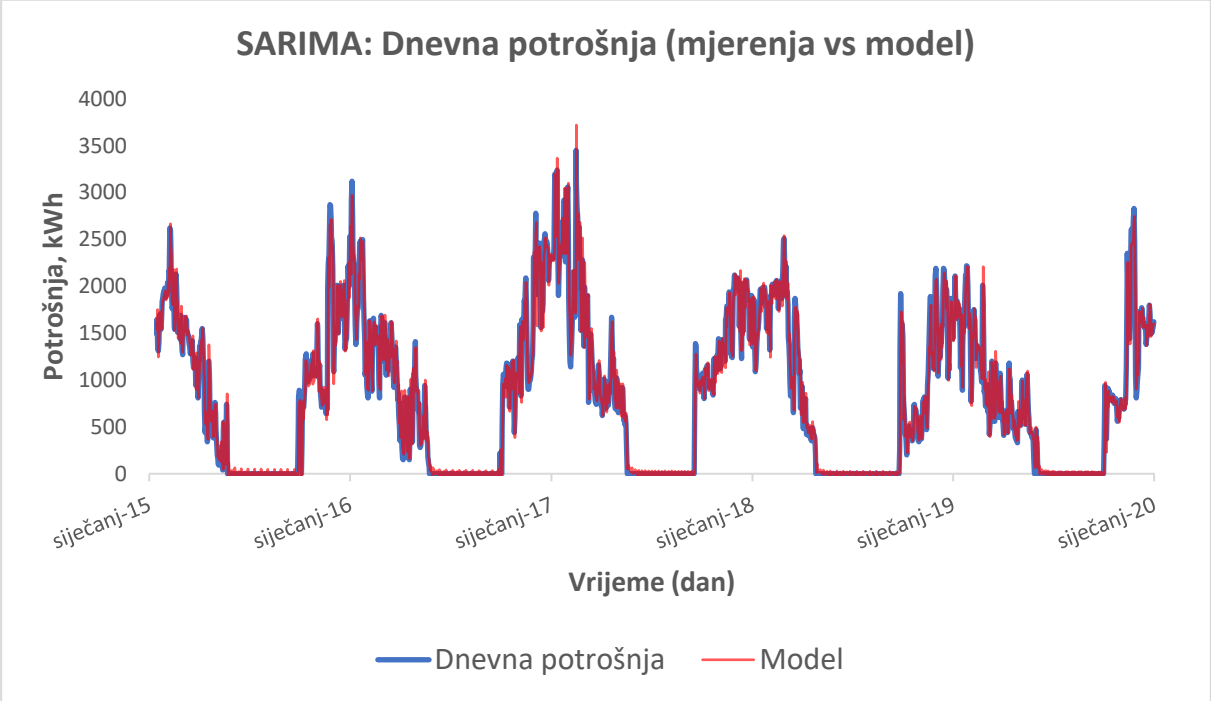
- AR(3) → $p = 3$,
- MA(3) → $q = 3$,
- SMA(1) → $Q = 1$,

te se ovi koeficijenti prihvaćaju za SARIMA model. Koeficijent koji nije prihvaćen je sezonalni autoregresijski model SAR(1) jer je *P* vrijednost puno veća od koeficijenta značajnosti α, te se za SARIMA model usvaja parametar $P = 0$. Sezonalna komponenta je definirana jednom godinom, odnosno s 12 mjeseci.

Nakon optimizacije parametara, odabrani SARIMA model je:

$$\mathbf{SARIMA = (3, 1, 3)(0, 1, 1)_{12}}$$

Objašnjenje parametara prikazan je u poglavlju *3.3.4.*



Slika 6.9 SARIMA model (3, 1, 3)(0, 1, 1)$_{12}$

Slika 6.9 prikazuje model s optimiziranim vrijednostima parametara. Plava krivulja predstavlja stvarne izmjerene vrijednosti, dok crvena krivulja predstavlja prilagođeni model nastao korištenjem optimiziranih parametara. Vidljivo je kako model, odnosno crvena krivulja prati kretanje plave krivulje odnosno stvarnih mjerenja što je pozitivan znak da je model adekvatan.

SARIMA model s optimiziranim parametrima ima sljedeću točnost:

- MAPE = 10,0 %
- MAE = 98,7 kWh

Prema definicijama iz *3.4*, prosječno apsolutno odstupanje kreiranog modela iznosi 98,7 kWh, odnosno prosječno postotno apsolutno odstupanje kreiranog modela iznosi 10 %.

6.2.2. Validacija modela

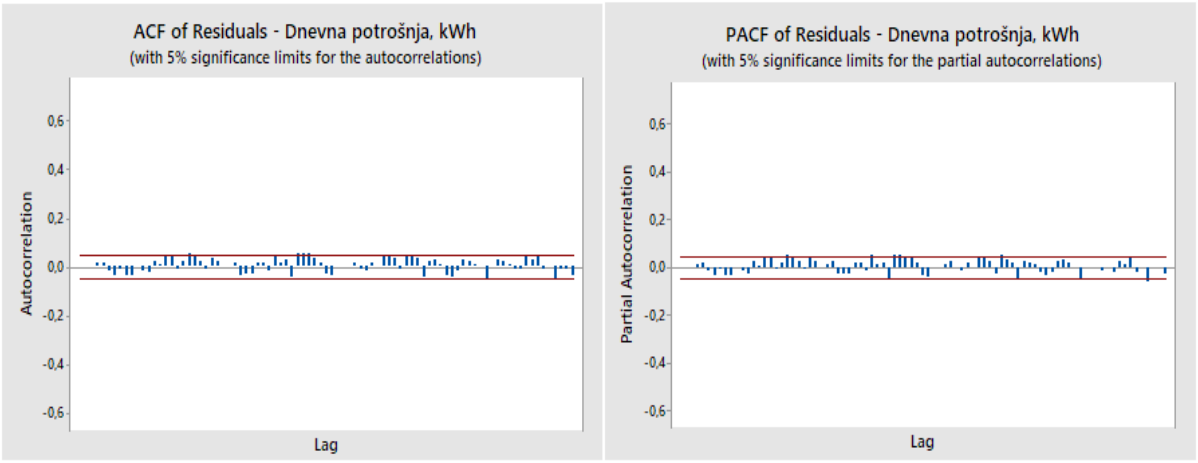
Validacija modela će se vršiti u dva koraka:

1. provjera rezidualnih vrijednosti,
2. validacija modela na testnom skupu podataka (1.1.2020. – 31.1.2020.).

6.2.2.1. Provjera rezidualnih vrijednosti modela

Kako bi se model validirao, potrebno je da rezidualne vrijednosti prate normalnu distribuciju i da je autokorelacija istih unutar kontrolnih granica.

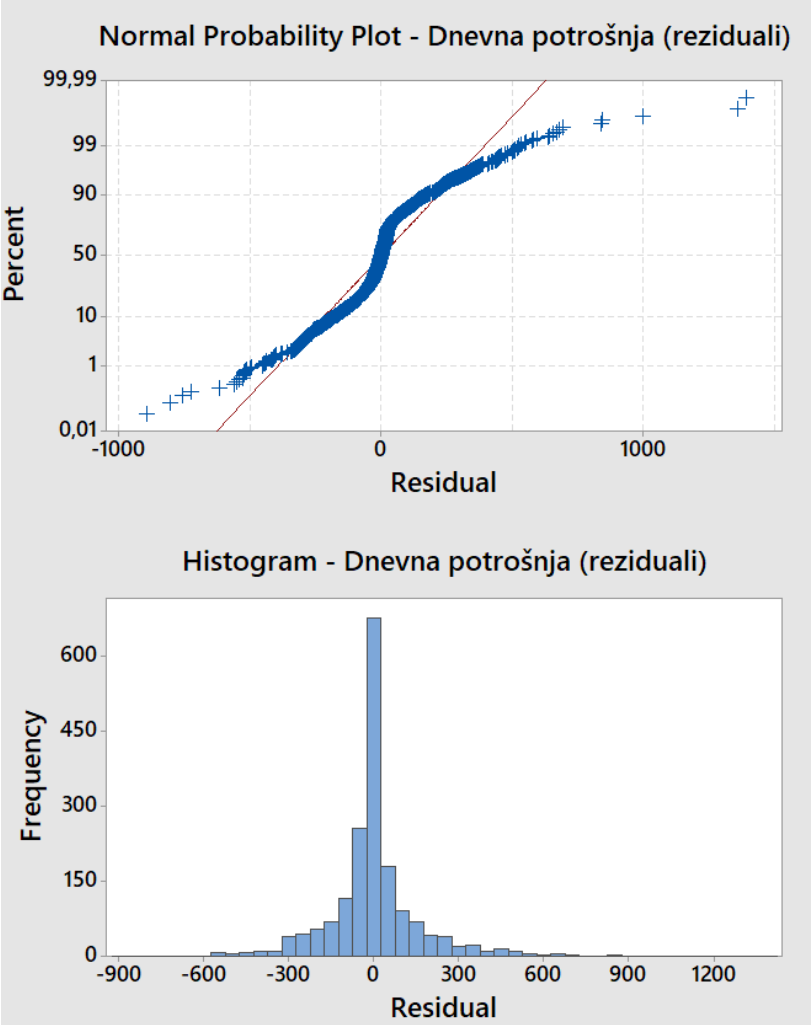
U nastavku će se provjeriti izgled autokorelacijskih funkcija.



Slika 6.10 Provjera autokorelacijskih funkcija rezidualnih vrijednosti

Slika 6.10 prikazuje funkcije autokorelacije ACF i PACF za rezidualne vrijednosti, te je vidljivo kako su vrijednosti unutar kontrolnih granica ili na samoj granici. Uvjet nekoreliranih rezidualnih vrijednosti je ispunjen.

Sljedeći korak je provjera normalnosti rezidualnih vrijednosti.

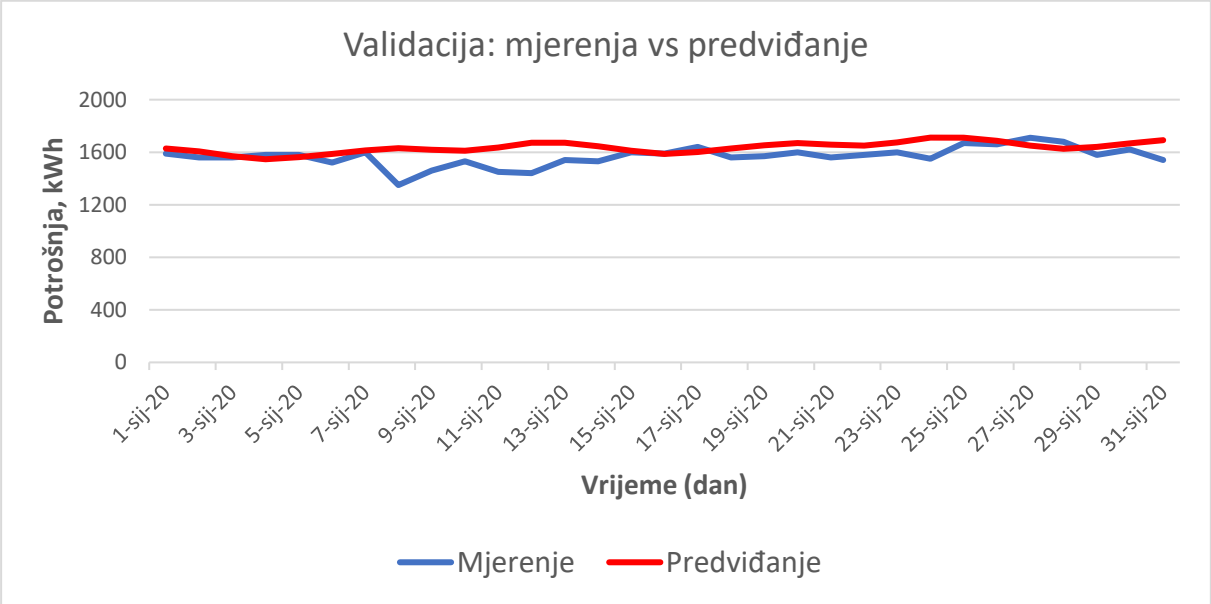


Slika 6.11 Provjera normalnosti rezidualnih vrijednosti

Slika 6.11 predstavlja histogram rezidualnih vrijednosti i dijagram normalnosti. Na histogramu je vidljivo kako rezidualne vrijednosti imaju zvonoliki izgled distribucije što ukazuje na normalnost podataka. Na dijagramu normalnosti vidljivo je kako rezidualne vrijednosti ne prate idealno krivulju normalnosti, ali u dovoljnoj mjeri da zadovolje uvjet normalnosti. Ova krivulja ukazuje na to da unutar rezidualnih vrijednosti postoji još „znanja“ odnosno da rezidualne vrijednosti nisu savršeno normalno (nasumično) distribuirane. Međutim korištenim programskim paketom ovo je najbolji model koji je moguće izraditi.

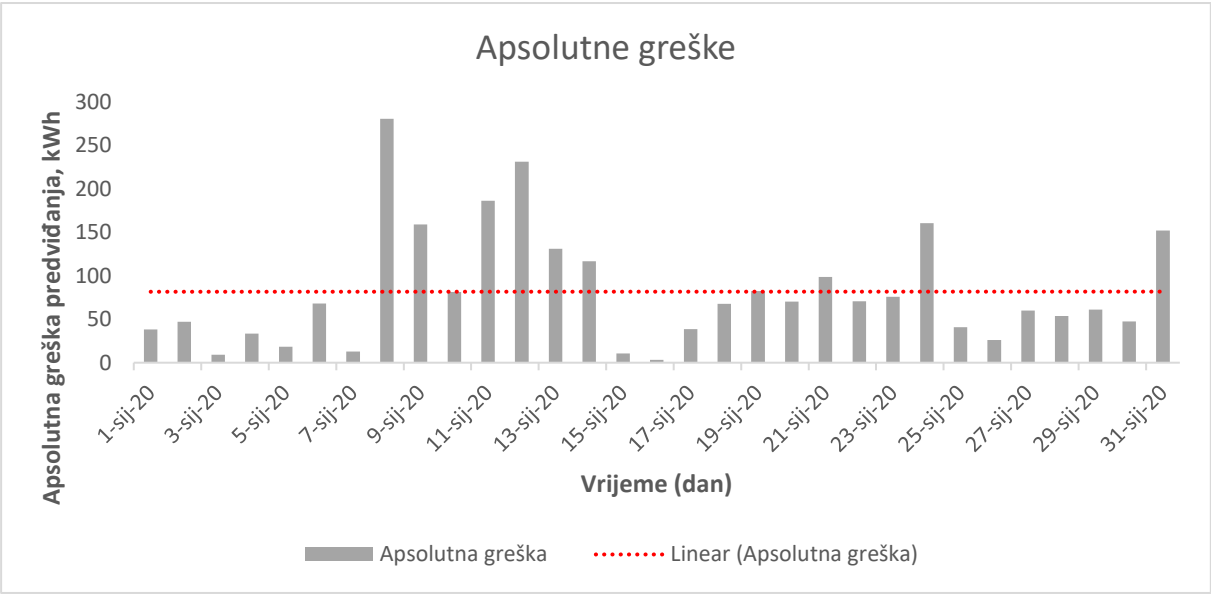
6.2.2.2. Valdiacija modela – testni skup

U nastavku će se kreirani SARIMA model $(3, 1, 3)(0, 1, 1)_{12}$ validirati na mjerenjima nastalim u periodu od 1.1.2020. do 31.1.2020.



Slika 6.12 Validacija SARIMA modela

Slika 6.12 prikazuje vrijednosti stvarnih mjerenja u periodu od 1.1.2020. do 31.1.2020. te predviđenih vrijednosti kreiranih SARIMA modelom. Vidljivo je kako su trendovi i kretanje krivulja u istom smjeru što je pozitivan znak za validaciju modela.



Slika 6.13 Analiza apsolutnih grešaka modela

Slika 6.13 ukazuje na to kako s protekom vremena dolazi do stabiliziranja apsolutne greške. To je vidljivo kroz linearni trend koji je praktički nepostojan (crvena linija na dijagramu). Iz toga se može zaključiti da je srednje odstupanje apsolutne greške ujednačeno za kreirani SARIMA model.

Mjera točnosti kreiranog predviđanja je sljedeća:

- MAPE = 5,0 %,
- MAE = 81,6 kWh.

Model je uspješno prošao validaciju jer su prosječno postotno apsolutno odstupanje (MAPE) i prosječno apsolutno odstupanje (MAE) predviđene potrošnje toplinske energije manji od istih mjera u SARIMA modelu nastalom iz testnog skupa.

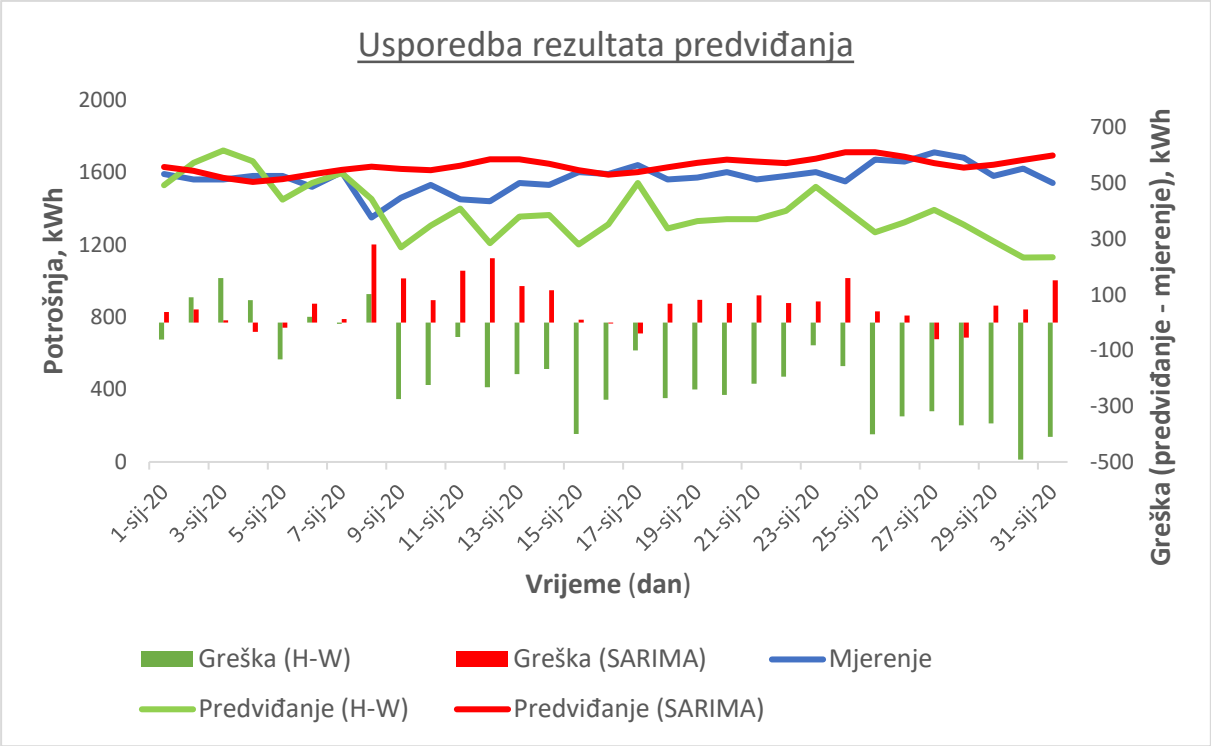
6.2.3. Analiza rezultata i prijedlog modela

Kreirani modeli i njihovi rezultati prikazani su u sljedećoj tablici (Tablica 6.2).

Tablica 6.2 Rezultati predviđanja potrošnje toplinske energije

MODEL			
Holt-Winters model ($\alpha = 0,9$; $\beta = 0,05$; $\gamma = 0,05$)		SARIMA $(3, 1, 3)(0, 1, 1)_{12}$	
MAE	MAPE	MAE	MAPE
115,1 kWh	13,6 %	81,6 kWh	5,0 %

Iz mjera točnosti može se zaključiti da je SARIMA model puno bolji u predviđanju potrošnje toplinske energije. Imajući u vidu kompleksnost modela, bilo je za očekivati ovakve rezultate. Prosječno postotno apsolutno odstupanje (MAPE) SARIMA modela je za 8,6% preciznije od Holt – Winters modela, dok je srednje apsolutno odstupanje predviđanja (MAE) preciznije za 33,5 kWh.



Slika 6.14 Usporedba rezultata kreiranih modela

Slika 6.14 prikazuje usporedbu rezultata kreiranih modela. Krivulje predstavljaju stvarne vrijednosti mjerenja i predviđenih vrijednosti potrošnje toplinske energije, dok stupci predstavljaju stvarnu grešku (greška je razlika između predviđene vrijednosti i mjerenja). Vizualno je vidljivo kako krivulja predviđanja SARIMA modela puno bolje opisuje stvarna mjerenja za razliku do Holt – Winters modela. Zanimljivo je primijetiti kako u većini slučajeva greška SARIMA modela ima pozitivan iznos dok je kod Holt – Winters modela obrnuto tj. uglavnom su greške negativnog predznaka. Iz ovoga se može zaključiti da SARIMA model daje „optimističnije“ rezultate, a Holt – Winters model daje „pesimističnije“ rezultate. Pri provjeri budućih vrijednosti potrošnje toplinske energije, svakako bi se u obzir trebala uzeti oba modela jer je vidljivo kako je stvarna vrijednost potrošnje negdje između predviđanja ova dva modela. Drugim riječima, ova dva modela mogla bi dati interval moguće buduće potrošnje.

Također, vidljivo je kako su greške puno manje pri početku procesa predviđanja te protjecanjem vremena one rastu. Razlika je što se kod SARIMA modela veličina greške vremenom stabilizira dok kod Holt – Winters modela greška linearno raste s protekom vremena. Zbog ovog se za predviđanje nekoliko unaprijednih koraka mogu koristiti obje metode, dok se za duže vremenske periode preporuča korištenje SARIMA modela.

7. ZAKLJUČAK

U sustavima proizvodnje i isporuke energije bitno je poznavati potrebe za zahtijevanom količinom energije u nadolazećem vremenskom periodu kako bi se olakšalo planiranje i kako bi se odredili optimalni parametri proizvodnje i dobave iste.

Kroz ovaj rad prikazane su statističke metode predviđanja vremenskih nizova. Konkretno, opisano je predviđanje potrebne toplinske energije za jedan poslovni objekt pomoću dvije metode, a to su SARIMA i Holt -Wintersova metoda eksponencijalnog zaglađivanja.

Kriterij po kojem su se kreirani modeli uspoređivali su mjere točnosti MAPE i MAE. SARIMA model je ostvario bolji MAPE rezultat za 8,6 %, odnosno 33,5 kWh bolji MAE rezultat od Holt – Wintersovog modela. Ipak, u vidu treba imati da je SARIMA model puno kompleksniji za modeliranje od Holt – Wintersovog. Nadalje, uz analizu grešaka modela, zaključeno je kako bi se za buduća predviđanja trebali u obzir uzeti rezultati i jednog i drugog modela te usporedbom i analizom doći do konačne odluke o predviđanju. Također, uočeno je kako oba modela imaju vrlo visoku točnost pri predviđanju nekoliko prvih koraka. Iz rezultata kreiranih modela, vidljivo je kako modeli nisu savršeni te postoji mogućnost za unapređenjem istih.

Svakako bi bilo zanimljivo usporediti performanse kreiranih statističkih modela s nekim drugim rješenjima kao što su prognoze temeljene na neuronskim mrežama (strojnom učenju) koje se također koriste za predviđanje vremenskih nizova.

Kreirani modeli i predikcije ukazuju na to kako je predviđanje vremenskih serija, ali i svih drugih pojava, i dalje kompleksan problem za koji ne postoji univerzalno rješenje. Za kreiranje kvalitetnog predikcijskog modela potrebna je kvalitetna strategija te spremnost na iscrpno strukturiranje podataka i optimiziranje parametara modela.

LITERATURA

[1] Kassambara A. *R Graphics Essentials for Great Data Visualization.* Marseille: STDHA; 2017.

[2] Daitan. *A Visual Guide to Time Series Decomposition Analysis*. Dostupno na: https://medium.com/better-programming/a-visual-guide-to-time-series-decomposition-analysis-a1472bb9c930 (pristupljeno: ožujak, 2020.)

[3] Montgomery DC, Jennings CL, Kulahci M. *Introduction to Time Series Analysis and Forecasting*. Hoboken: J. Wiley & Sons Inc.; 2015.

[4] Hyndman RJ, Athanasopoulos G. *Forecasting principles and practice*. 2nd ed. Melbourne: OTexts; 2018.

[5] Zhang B, Pu Y, Li J. *Forecasting Hotel Accommodation Demand Based on LSTM Model Incorporating Internet Search Index*. Sustainability. 2019 Aug; 11(17): 4708. doi: 10.3390/su11174708.

[6] Shumway RH, Stoffer DS. *Time Series Analysis and its Applications*. 4th ed. New York: Springer. 2016.

[7] Brownlee J. *Introduction to Time Series Forecasting with Python*. E-book: Machine Learning Mastery. 2017.

[8] Markos O. *Time series forecasting – SARIMA vs Auto ARIMA models.* Dostupno na: https://medium.com/analytics-vidhya/time-series-forecasting-sarima-vs-auto-arima-models-f95e76d71d8f (pristupljeno: ožujak, 2020.)

[9] Burnham, KP, Anderson DR. *Understanding AIC and BIC in Model Selection*. Sociological Methods & Research. 2004 Nov; 33(2): 261–304. doi:10.1177/0049124104268644.

[10] Hyndman RJ, Koehler AB. *Another look at measures of forecast accuracy*. International Journal of Forecasting. 2006 Mar; 22(4): 679–688. doi: 10.1016/j.ijforecast.2006.03.001.

[11] *Energetski certifikat nestambene zgrade*. Direktiva 2010/31/EU. Zagreb; 2016.

[12] Minitab. *Model statistics for ARIMA*. Dostupno na: https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/time-series/how-to/arima/interpret-the-results/all-statistics-and-graphs/model-statistics/ (pristupljeno: travanj, 2020.)

Napomena:

Za statističku obradu podataka korišten je programski paket *STATISTICA 13.3 (StatSoft)* i *Minitab 17* (edukacijska verzija).