



UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI
FAKULTET
DEPARTMAN ZA MATEMATIKU I
INFORMATIKU



Jelena Samardžija

Optimizacioni metodi za dugoročne predikcije koncentracije polena glavnih alergena u Vojvodini

MASTER RAD

Mentor

dr Nataša Krklec Jerinkić

Novi Sad, 2019.

Sadržaj

1. Uvod	4
2. Polenska alergija kroz istoriju	6
2.1. O polenu.....	7
2.2. O alergiji	7
3. Metodologija istraživanja	9
3.1. Prostor verovatnoća, slučajne promenljive i stohastički procesi	9
3.2. Osnovni pojmovi vremenskih serija.....	10
3.3. STARI metod i kriterijumi.....	16
3.4. Problem najmanjih kvadrata	18
4. Istraživanje	22
4.1. Podaci, geografski položaj i klima	22
4.2. Koncentracija polena u vazduhu tokom godine	24
4.3. Metode za popunjavanje nedostajućih podataka.....	27
4.4. Statističke karakteristike podataka	27
4.5. NOVI metodi.....	32
5. Rezultati	35
5.1. Novi Sad	35
5.2. Vrbas.....	40
5.3. Sombor.....	45
5.4. Zrenjanin.....	50
Zaključak	54
Literatura	55
Biografija	56

Veliku zahvalnost dugujem svojoj mentorki dr Nataši Krklec Jerinkić koja me je svojim predavanjima zainteresovala za predmet Vremenske serije, probudila želju za dublje istraživanje obrađenih tema, a samim tim i izučavanje i praktičnu primenu istih u cilju izrade ovog master rada. Dugujem joj zahvalnost i na nesebičnoj pomoći koju je pružala u svakom trenutku i sugestijama koje su me usmeravale prilikom pisanja rada. Zahvaljujem se i ostalim članovima komisije, kao i svim profesorima koji su mi predavali tokom studija i koji su me uveli u različite, ali opet povezane matematičke svetove.

Veliku zahvalnost na pomoći oko izrade rada dugujem i dr Branku Šikopariji sa Departmana za biologiju Prirodno matematičkog fakulteta u Novom Sadu. U radu su korišteni meteo podaci koji su izmereni na meteorološkoj stanici Republičko Hidrometeorološkog zavoda Srbije Rimski Šančevi, koji su javno dostupni i preuzeti sa sajta <http://www.tutiempo.net>. Podatke o koncentracijama polena prikupila je Laboratorija za palinologiju PMF Novi Sad u okviru redovnog monitoringa aeroalergena koje finansira Gradska uprava za zaštitu životne sredine (Grad Novi Sad) i Pokrajinski sekretarijat za urbanizam i zaštitu životne sredine (AP Vojvodine). Za potrebe realizacije master rada podaci su dostupni zahvaljujući RealForAll projektu (sufinansiraju IPA Cross-border Cooperation programme Croatia – Serbia 2014-2020 i Pokrajinski sekretarijat za finansije, AP Vojvodina).

Hvala prijateljima koji su činili, a i dalje čine, svaki dan vrednim sećanja i koji su postali moja izabrana porodica.

Na kraju, najveću zahvalnost dugujem svojim roditeljima, bratu i Maji. Bez njih ništa ne bi bilo kao što jeste i ne bi vredelo koliko vredi uz njih.

Novi Sad, 2019.

Jelena Samardžija

1. Uvod

Sa problemom alergije na polen susreće se sve veći broj ljudi. Njeni simptomi utiču veoma nepovoljno na kvalitet života na taj način što obavljanje svakodnevnih aktivnosti postaje otežano i neprijatno. Sama alergija ne predstavlja opasnost po zdravlje, ali ukoliko joj se ne pristupi na pravilan način može se razviti u alergijski rinitis koji može da dovede do niza ozbiljnijih zdravstvenih problema.

Alergija se ne može izlečiti, ali se njeni simptomi mogu kontrolisati i svesti na minimum. U lečenju se koriste preventivne mere i terapijski postupci. Izbegavanje boravka i fizičkih aktivnosti na otvorenom prostoru za vreme velike koncentracije polena u vazduhu jedna je od glavnih mera prevencije. Takođe, uzimanje određene terapije koja bi pripremila organizam na povećanu koncentraciju polena olakšava i smanjuje simptome alergije. Time dolazimo do teme ovog rada, a to je formiranje dugoročne prognoze koncentracije polena u vazduhu za glavne alergene tj. ambroziju, brezu i travu na teritoriji Novog Sada, Vrbasa, Sombora i Zrenjanina. Predviđanje varijabilnosti dnevnih koncentracija polena za navedene alergene bilo bi od ogromnog značaja za veliki broj pojedinaca osetljivih na polen. Informacija o očekivanoj količini polena omogućava alergičnim osobama da pripreme organizam na odgovarajući način kao i da usklade svoje obaveze na taj način da izbegnu ili svedu na minimum boravak na otvorenom za vreme povećane koncentracije polena u vazduhu.

Za navedene gradove beležena je koncentracija polena na svaka dva sata, svaki dan u toku godine. Za Novi Sad imamo najveću bazu sa istorijskim podacima i zbog toga je nad tim podacima sprovedeno glavno istraživanje. Kao relevantni metod za formiranje dugoročne prognoze koncentracije polena u vazduhu korišten je metod iz rada „Aerobiologija“ [3]. Pored njega u ovom radu korištena su još dva metoda koja predstavljaju modifikaciju navedenog. Svi metodi koriste prosečne dnevne vrednosti o koncentraciji polena izmerene na teritoriji određenog grada za izračunavanje prosečnih dnevnih vrednosti u budućem periodu. Pomoću odgovarajućih kriterijuma vršeno je poređenje dobijenih predikcija u cilju odabira najboljeg metoda. Dakle, glavni deo istraživanja i formiranje novih metoda sprovedeno je na podacima za Novi Sad, a zatim su dobijeni metodi primenjeni na Vrbasa, Sombor i Zrenjanin.

Problemi koji se često javljaju u rutinskom praćenju koncentracije polena su tzv. „rupe“ u podacima, odnosno nedostajući podaci. Njihov uzrok je najčešće gubitak električne energije, zaustavljanje sata na uređaju, itd. Popunjavanje ovih podataka je neophodno da bi se izbegle greške u predikcijama i obično se vrši nekom od interpolacionih metoda. Od svih metoda linerana interpolacija se pokazala kao najbolja, pa je ona korištena za interpolaciju i u ovom istraživanju kako za Novi Sad tako i za ostale gradove.

Predviđanje koncentracije polena, kao i ostalih pojava koje se dešavaju u prirodi, nije jednostavno iz razloga što zavisi od velikog broja drugih prirodnih faktora. Emisija polena zavisi od cvetanja biljke, a cvetanje biljke od vremenskih uslova. Ukoliko su uslovi povoljniji biljka može cvetati ranije nego što je očekivano i obrnuto, ukoliko su nepovoljni kasnije.

Zatim kada dodje do emisije polena njegovo širenje i rasprostranjenost zavisti od broja vetrovitih dana i brzine vetra. Ali ukoliko je određenog dana padala kiša, nakon nje količina polena u vazduhu je smanjena. Dakle, možemo predvideti dobru vrednost ali iznenadni letnji pljusak može da smanji koncentraciju polena i takvu grešku u predikciji ne možemo pripisati modelu. Ukoliko izuzmemo ovakve iznenadne i kratkotrajne skokove i padove, čak i približni podaci o količini očekivanog polena u vazduhu mogu u velikoj meri da pomognu alergičnim osobama da olakšaju simptome alergije.

Celokupna analiza, popunjavanje nedostajućih podataka, formiranje predikcija kao i poređenje modela pomoću odabranih kriterijuma sprovedeno je u programu Matlab.

2. Polenska alergija kroz istoriju

Alergija na polen je problem sa kojim se danas susreće sve veći broj ljudi. Međutim, postoje dokazi da preosteljivost na prisustvo cvetnog praha u vazduhu nije problem samo modernog doba. Zabeleženo je da je pre oko 2500 godina persijski vođa Hinijski prilikom iskrcavanja na grčku obalu dobio napad kašlja i kijanja. Uzročnik je najverovatnije bilo kasno cvetanje neke mediteranske biljke. U drugom veku naše ere, poznati grčki lekar i filozof Galen alergiju je opisao na sledeći način: „Postoje ljudi koji u prisustvu nekog cveća dobijaju napade kašlja i kijanja.”

Tek 1819. godine nastao je prvi naučni opis ove alergije. Načinio ga je engleski lekar Džon Bostok koji je u izveštaju “Kraljevskom medicinskom društvu” naveo da postoji oboljenje očiju i nosa koje se javlja tokom leta. U čitavoj Engleskoj uspeo je da pronađe svega 28 pacijenata sa istim simptomima. Svi pacijenti su imali kontakt sa svežim senom, pa je Bostok smatrao da je sveže seno glavni uzrok ovih tegoba, a bolest je nazvao senska groznica. Ovaj termin se zadržao i do danas na engleskom govornom području. Izveštaj dr Bostoka predstavlja početak praćenja simptoma polenskih alergija i smatra se možda najznačajnijim korakom u borbi čovečanstva sa alergijama.

Prošlo je više od pola veka, kada je drugi engleski lekar, Čarls Harison Blekli, 1873. godine objavio da je uzrok ove bolesti polen trava, a ne seno. Kada je otkrio da kod pojedinih ljudi zrnca polena trave postavljena na sluzokožu nosa izazivaju kijavicu napisao je: „Važnije je prvo identifikovati šta vam leti iznad glave, koji poleni, pa tek onda uraditi kožne probe za alergiju“. Takođe, njegov doprinos je značajan i u tome što je uočio i opisao osnovne razlike između polenske kijavice i prehlade što predstavlja prvi korak u borbi protiv alergije.

Prvi učitelj i tvorac alergološke škole u Srbiji bio je čuveni profesor medicine i akademik Vladimir Spužić. Njegov nemerljivi doprinos razvoju alergologije ogledao se kroz stručni, naučni i nastavnički rad. Prof. dr Vladimir Spužić organizovao je rad prve alergološke ambulante u Srbiji, marta 1946. godine, i rukovodio njenim radom do penzionisanja.

Situacija koja danas vlada u svetu je zabrinjavajuća, uprkos sve većem broju istraživanja i generalno sve većim bavljenjem problemom alergije na polen. U poslednjih 25 godina prošlog veka, broj Evropljana koji pate od alergije povećao se za sedam puta. Od 2 odsto, koliko ih je bilo 1968. godine, u 2000. godini registrovano ih je čak 30. Prema podacima iz 2004. godine, od alergijskog rinitisa u različitim delovima sveta patilo je od 10 do 25 odsto stanovništva, pa se slobodno može reći da bolest ima epidemijske razmere. Procenjuje se da danas u svetu svaka treća osoba ima alergijske reakcije na polen.

2.1. O polenu

Polen je fini, cvetni prah koji se sastoji od nekoliko hiljada polenovih zrna biljaka. Biološka funkcija polena je oplođenje. Pored osnovne funkcije, polen je jedan od najznačajnijih bioloških izvora alergena. Alergena svojstva polena potiču od hemijskih jedinjenja koja se nalaze na površini i unutar samog polenovog zrna.

Polenovo zrno je bogato različitim hemijskim jedinjenjima među kojima su proteini najodgovorniji za izazivanje alergijske reakcije. Naprimera, kod ambrozije je utvrđeno 6 alergeni proteina. Kad polenovo zrno dođe u kontakt sa vlažnom sluznicom, oslobađa biološke aktivne supstance, koje dovode do senzibilizacije i reakcije pacijenta na taj polen, tj. pokreće se delovanje imunog sistema.

U prirodi polen se može prenositi na različite načine: vetrom, preko životinja i vodom. Rasprostiranje polena pomoću vetra je najveći problem jer ne postoji način da se spreči njegovo širenje i prisustvo u vazduhu. U toku vegetacionog perioda na području Srbije postoje tri maksimuma koncentracije polena u vazduhu:

- ranoprolećni (drveće i žbunje)
- letnji (trave)
- letnje-jesenji (korov)

Od temperature, vlažnosti vazduha, padavina i ostalih parametara spoljašnje sredine zavisi kada, koliko dugo i koja količina polena će se naći u vazduhu. Viša temperatura i niža vlažnost vazduha, kao i vetrovito vreme pogoduju rasipanju i širenju polena dok kišni periodi onemogućavaju širenje i smanjuju koncentraciju polena u vazduhu. Ipak, dugotrajna visoka vlažnost vazduha takodje dovodi do bubrenja i prskanja polenovih zrna i oslobađanja alergnog materijala u vazduh, te može da izazove alergijske reakcije kod osoba koje su alergične na polen.

2.2. O alergiji

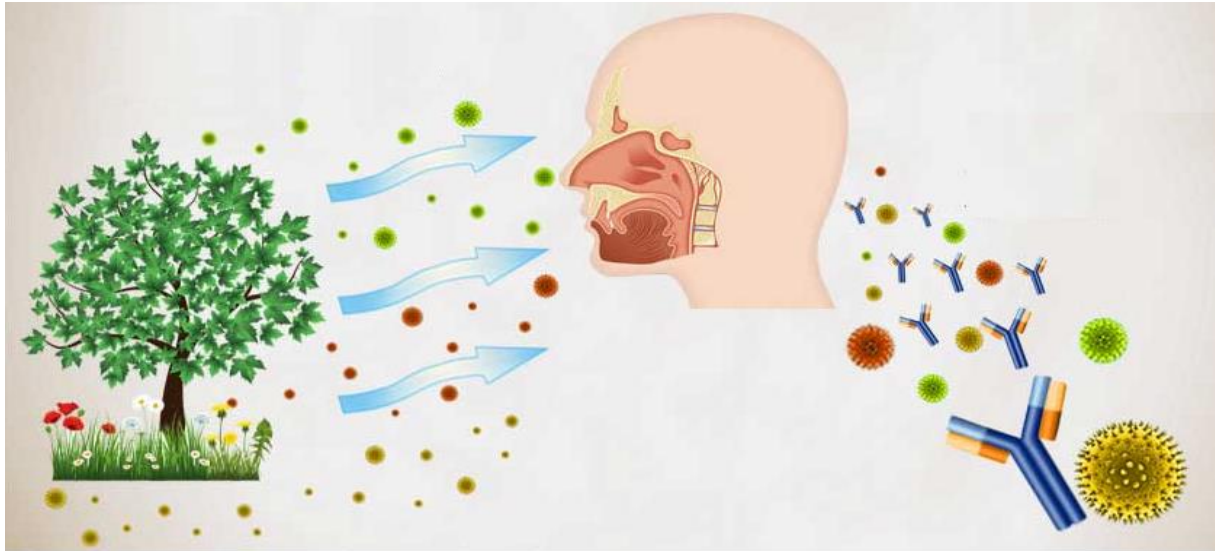
Alergije su poslednjih decenija u velikom porastu. Smatra se da svaki treci stanovnik sveta pati od nekog oblika alergije. Preko 40 posto svetske populacije je genetski predodređeno za razvoj alergije i to je nesumnjivo važan faktor. Ali pored njega veliki uticaj imaju i visoki higijenski standardi i izmenjene životne navike kao što su: boravak u zatvorenim prostorima, smanjena fizička aktivnost, gojaznost, preterana upotreba antibiotika, pušenje, zagađenje životne sredine itd.

Pojam alergija odnosi se na preosetljivost organizma na neke supstance, dok je alergijska reakcija odgovor imunog sistema na, u organizam, unete alergene supstance. Nosioci supstanci (alergena) koji izazivaju alergijsku reakciju mogu biti hrana, spore, delovi tela životinja, otrov insekata, polen itd.

Alergija na polen nastaje tako sto polen ulazi u oči, nos, pluća, povećavajući svaki put osetljivost našeg organizma. Naš organizam tada stvara specifična antitela na polen koja se potom vezuju za mastocite (vrste belih krvnih zrnaca) koji se nalaze u tkivu. Kada polen ponovo dospe u telo, on se vezuje za već nastala antitela, i prouzrokuje oslobađanje histamina

i drugih supstanci iz mastocita. Tada dolazi do alergijske reakcije organizma. Najčešće i najprepoznatljivije reakcije na polen odnosno simptomi alergije su kihanje, zapešćen nos i sinusi, curenje nosa, svrab u predelu nosa, grla i očiju, suzne oči, konjuktivitis, rinitis, itd.

Simptomi alergije, bilo da su u umerenom ili težem obliku, u velikoj meri otežavaju obavljanje svakodnevnih čovekovih aktivnosti. Osim što nepovoljno utiče na kvalitet života, alergija po pravilu nije opasna po zdravlje. Postaje opasna tek ako se iz nje razvije alergijska astma.



Slika 1.1 Ilustracija alergijske reakcije organizma na polen (<http://wp11.americanobserver.net/wp-content/uploads/2012/04/pollen-graphic.jpg>)

Određenim merama prevencije, uzimanjem lekova ili drugim oblicima terapije simptomi se mogu smanjiti ili čak u potpunosti eliminisati. Međutim lek za potpuno izlečenje alergije ne postoji u klasičnoj medicini. Veoma važna preventivna mera i pomoć osobama alergičnim na polen je redovno informisanje o stanju i prognozi polena u vazduhu. Na taj način mogu da se pripreme i odreaguju na odgovarajući način bilo da je u pitanju početak sezone pojave polena u vazduhu ili povećana koncentracija polena u vazduhu u toku sezone.

3. Metodologija istraživanja

3.1. Prostor verovatnoća, slučajne promenljive i stohastički procesi

Rad sa vremenskim serijama zahteva pre svega dobro poznavanje teorije verovatnoće, statistike i stohastičke analize. Zbog toga ćemo navesti osnovne pojmove i definicije iz ovih oblasti koje će nam biti potrebne prvo da bismo uopšte mogli dati definiciju šta je to vremenska serija, a potom i za dalji rad sa njima.

Definicija 2.1 Prostor verovatnoća, određen slučajnim eksperimentom, je trojka (Ω, \mathcal{F}, P) , gde je Ω skup skup svih (logički) mogućih ishoda slučajnog eksperimenta, \mathcal{F} σ -algebra (polje), a P funkcija koja svakom skupu $A \in \mathcal{F}$ dodeljuje broj $P(A)$, koji se naziva verovatnoća da se desi događaj A .

Definicija 2.2 Preslikavanje $X: \Omega \mapsto \mathbb{R}^n$, se zove n -dimenzionalna slučajna promenljiva nad prostorom verovatnoća (Ω, \mathcal{F}, P) , ako važi da je X \mathcal{F} -merljivo, tj. $X^{-1}(S) \in \mathcal{F}$, za svaki Borelov skup S .

Definicija 2.3 Funkcija $F_X(x) = F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n)$, za $(x_1, \dots, x_n) \in \mathbb{R}^n$, zove se zajednička funkcija raspodele slučajnih promenljivih (X_1, \dots, X_n) .

Definicija 2.4 Slučajna promenljiva $X=(X_1, \dots, X_n)$ je apsolutno neprekidna ako postoji integrabilna funkcija $f_X(x) = f_{(X_1, \dots, X_n)}(x_1, \dots, x_n)$, $x_i \in \mathbb{R}$, $i = 1, \dots, n$, takva da $P(X \in S) = \int \dots \int_S f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) dx_1, \dots, dx_n$, za $S \in \mathcal{B}^n$. Specijalno, za $S = \{(u_1, \dots, u_n) \in \mathbb{R}^n, u_i < x_i\}$, dobićemo $F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{(X_1, \dots, X_n)}(u_1, \dots, u_n) du_n, \dots, du_1$.

Definicija 2.5 Neka je X apsolutno neprekidna slučajna promenljiva. Očekivanje slučajne promenljive X dato je sa

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

i ono postoji ako i samo ako i samo ako $\int_{-\infty}^{+\infty} |x| f_X(x) < +\infty$.

Pojam slučajne promenljive je nezavisan o vremenu. Ali, ukoliko posmatramo neki proces koji se odvija u vremenu i čiji ishod je neizvestan, potrebno je da uključimo i vremensku komponentu, odnosno da uopštimo proces slučajne promenljive. Na taj način dolazimo do pojma stohastičkog procesa.

Definicija 2.8 Stohastički (slučajni) proces $\{X(t), t \in \tilde{T}\} := \{X_t, t \in \tilde{T}\}$ je familija slučajnih promenljivih definisanih na istom prostoru verovatnoća (Ω, \mathcal{F}, P) .

Skup \tilde{T} se naziva parametarski (indeksni) skup i to je najčešće skup koji predstavlja vremenski interval. Ako je konačan, dobićemo konačno mnogo slučajnih promenljivih. Ako je prebrojiv, onda govorimo o nizu slučajnih promenljivih. U slučaju da je neprebrojiv, onda govorimo o „pravom“ stohastičkom procesu.

Predikcije koncentracije polena ambrozije, breze i trave za određene gradove koje razmatramo u ovom radu rađene su na osnovu podataka o merenjima iz prethodnih godina, pri čemu znamo početak i kraj merenja. Dakle, u ovom slučaju skup \tilde{T} je konačan skup.

3.2. Osnovni pojmovi vremenskih serija

Analiza vremenskih serija se poslednjih decenija izdvaja kao jedna od statističkih disciplina koja beleži najdinamičniji razvoj. Njenom razvoju doprinela je unutrašnja dinamika same discipline, ali i intrakcija sa ostalim disciplinama. Razni ekonomski modeli, kao i modeli u prirodi evoluirali su od statičkih do dinamičkih, takođe u modele je uvedena neizvesnost, odnosno evolucija je tekla od determinističkih do stohastičkih modela.

Sa vremenskim serijama susrećemo se u različitim oblastima ljudskog života. Zajedničko za sve njih je da se mogu nazvati uređenim nizom opservacija. Uređivanje se najčešće, ali ne i nužno, vrši u odnosu na vreme u jednakim vremenskim intervalima. Tako vremensku seriju možemo zapisati kao skup merenja x_t , od kojih je svako merenje dobijeno za poseban trenutak t . Upravo ovakav, vremenski poredak, daje opservacijama bitnu osobinu, a to je međusobna zavisnost. Pomoću te osobine vršimo analizu vremenskih serija, pokušavamo objasniti zašto i kako su se desile određene pojave, a krajnji cilj analize je formiranja modela. Model zatim koristimo da na osnovu prošlih, prognoziramo buduće opservacije.

Definicija 2.9 *Vremenski niz ili vremenska serija je skup hronološki uređenih vrednosti promenljive, koja predstavlja pojavu ili stohasticki proces u vremenu.*

Vremenska serija se može posmatrati kao jedna realizacija stohastičkog procesa. Radi razumevanja, odnos vremenske serije i stohastičkog procesa može da se posmatra kao odnos uzorka i populacije u klasičnoj statističkoj analizi. Kao što na osnovu uzorka, koji je izvučen iz određene populacije, izvodimo zaključke o karakteristikama populacije tako na osnovu vremenske serije možemo da izvedemo zaključke o osnovnim karakteristikama stohastičkog procesa.

Rad sa vremenskim serijama kao neprekidnim slučajnim promenljivama zahteva navođenje još nekih pojmova iz teorije verovatnoće.

Momenat reda k apsolutno neprekidne slučajne promenljive X se definiše kao

$$m'_k = E(X^k) = \int_{-\infty}^{\infty} x^k f_X(x) dx,$$

gdje je E oznaka za očekivanje, a f_X funkcija gustine za X .

Momenat reda jedan naziva se očekivana vrednost ili sredina slučajne promenljive X . Ona meri centralnu poziciju raspodele. Označavaćemo je sa μ_x .

Centralni momenat reda k neprekidne slučajne promenljive X definiše se kao

$$m_k = E[(X - \mu_x)^k] = \int_{-\infty}^{\infty} (x - \mu_x)^k f_x dx,$$

pod uslovom da dati integral postoji.

Centralni momenat reda dva naziva se varijansa ili disperzija promenljive X . Ona meri „rasipanje” vrednosti slučajne promenljive oko sredine. Označavaćemo je sa σ_x^2 .

Standardna devijacija neprekidne slučajne promenljive X je pozitivan kvadratni koren iz disperzije. Označavamo je sa σ_x .

Normalizovani momenat reda tri neprekidne slučajne promenljive X definiše se kao

$$S(x) = E \left[\frac{(X - \mu_x)^3}{\sigma_x^3} \right].$$

Naziva se koeficijent asimetrije i meri simetričnost raspodele slučajne promenljive, u odnosu na srednju vrednost.

Normalizovani momenat reda četiri neprekidne slučajne promenljive X , definiše se kao

$$K(x) = E \left[\frac{(X - \mu_x)^4}{\sigma_x^4} \right].$$

Naziva se koeficijent ekscesa i meri „debljinu repova” raspodele.

Kovarijansa slučajnih promenljivih X i Y je data sa

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))].$$

Može se pokazati da važi $Cov(X, Y) = E(XY) - E(X)E(Y)$.

Koeficijent korelacije između slučajnih promenljivih X i Y dat je sa

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{D(X)D(Y)}}$$

Posmatrajmo vremesku seriju $x = \{x_1, \dots, x_T\}$ kao slučajni uzorak, koji ima T vrednosti, uzet iz populacije X . Tada je:

- **Uzoračka sredina**

$$\hat{\mu}_x = \frac{1}{T} \sum_{t=1}^T x_t$$

- **Medijana uzorka**

$$\widehat{me}_x = \begin{cases} \frac{x_n + x_{n+1}}{2}, T = 2n \\ x_{n+1}, T = 2n + 1 \end{cases},$$

pod uslovom da su vrednosti u uzorku poređane u neopadajućem redosledu.

- **Najveća vrednost u uzorku**

$$x_{max} = \max\{x_1, \dots, x_T\}$$

- **Najmanja vrednost u uzorku**

$$x_{min} = \min\{x_1, \dots, x_T\}$$

- **Centrirana disperzija uzorka**

$$\hat{\sigma}_x^2 = \frac{1}{T-1} \sum_{t=1}^T (x_t - \hat{\mu}_x)^2$$

- **Standardna devijacija uzorka**

$$\hat{\sigma}_x = \sqrt{\hat{\sigma}_x^2}$$

- **Koeficijent asimetrije uzorka**

$$\hat{S}(x) = \frac{1}{(T-1)\hat{\sigma}_x^3} \sum_{t=1}^T (x_t - \hat{\mu}_x)^3$$

- **Koeficijent ekscesa (spljoštenosti) uzorka**

$$\hat{K}(x) = \frac{1}{(T-1)\hat{\sigma}_x^4} \sum_{t=1}^T (x_t - \hat{\mu}_x)^4$$

- **Uzorački koeficijent linearne korelacije za dva uzorka $\{x_1, \dots, x_T\}$ i $\{y_1, \dots, y_T\}$**

$$\hat{\rho}_{XY} = \frac{\sum_{t=1}^T (x_t - \hat{\mu}_x)(y_t - \hat{\mu}_y)}{\sqrt{\sum_{t=1}^T (x_t - \hat{\mu}_x)^2 \sum_{t=1}^T (y_t - \hat{\mu}_y)^2}}$$

gde se izraz u brojiocu naziva uzoračka kovarijansa uzorka pomožena sa T .

Bitne osobine $\hat{S}(x)$ i $\hat{K}(x) - 3$ su da imaju asimptotski normalu raspodelu, očekivanje nula i disperziju $6/T$ i $24/T$, respektivno. Ove asimptotske osobine se koriste za testiranje normalnosti vremenskih serija. Prilikom testiranja vrednosti koeficijenta asimetrije uzimamo da je nulta hipoteza $H_0(S(x) = 0)$, dok je alternativna $H_1(S(x) \neq 0)$. Vrednost test statistike se dobija na sledeći način

$$t = \frac{\hat{S}(x)}{\sqrt{6/T}}$$

Nulta hipoteza se odbacuje na nivou poverenja α , ako je $|t| > z_{\alpha/2}$, gde je $z_{\alpha/2}$ kvantil standardne normalne raspodele. U suštini, izračunamo verovatnoću p za test statistiku t i odbacujemo nultu hipotezu ako je p manje od α .

Za testiranje koeficijenta ekscesa koristimo nultu hipotezu $H_0(K(x) - 3 = 0)$, protiv alternativne $H_1(K(x) - 3 \neq 0)$. Vrednost test statistike se dobija kao

$$t = \frac{\hat{K}(x) - 3}{\sqrt{24/T}}$$

Nulta hipoteza se prihvata ili odbacuje na osnovu gore navedenih pravila.

1987. godine Jarque i Bera su pojednostavili testiranje normalnosti vremenske serije tako što su spojili ova dva testa u jedan i dobili test statistiku

$$JB = \frac{\hat{S}(x)^2}{6/T} + \frac{(\hat{K}(x) - 3)^2}{24/T},$$

koja ima asimptotsku χ^2 raspodelu sa dva stepena slobode. Nulta hipoteza o postojanju normalnosti se odbacuje ukoliko je verovatnoća p test statistike JB manja nego nivo poverenja α .

Na osnovu osobina koje vremenske serije poseduju možemo ih podeliti na stacionarne i nestacionarne. Uopšteno gledano, stacionarnost je svojstvo vremenske serije čije se kretanje tokom vremena odvija po ustaljenom obrazcu, u smislu nepromenljivosti svojstava. Stacionarnost može takođe da se podeli na striktnu i slabu. Kako uslovi striktnosti često nisu ispunjeni, kada govorimo o stacionarnosti uglavnom se pretpostavlja slaba.

Definicija 2.10 *Vremenska serija $\{x_t, t = 1, \dots, T\}$ je striktno stacionarna ako $(x_{t_1}, \dots, x_{t_k})$ i $(x_{t_1+t}, \dots, x_{t_k+t})$ imaju istu raspodelu za svako t , gde su $t_1, \dots, t_k \in \mathbb{N}$.*

Definicija 2.11 *Vremenska serija $\{x_t, t = 1, \dots, T\}$ je slabo stacionarna ako su $E(x_t)$ i $cov(x_t, x_{t-l})$ nezavisne od t za svako $l \in \mathbb{Z}$.*

Ukoliko posmatramo x_t i formiramo model na osnovu x_{t-1} , tada ako vremenska serija ispunjava uslove slabe stacionarnosti, za predviđanje x_{t+1} možemo koristiti x_t . Dakle, slaba stacionarnost nam garantuje istu kovarijansu, odnosno opravdava predikcije što je neophodno da bi model imao smisla.

Implicitna pretpostavka u slaboj stacionarnosti je da su $E(x_t)$ i $E(x_t^2)$ konačni. Dakle, striktna stacionarnost ne implicira slabu u opštem slučaju, samo ako se pretpostavi da su $E(x_t)$ i $E(x_t^2)$ konačni. To nam govori i sledeća teorema.

Teorema 2.1 Ako je vremenska serija $\{x_t, t = 1, \dots, T\}$ striktno stacionarna i $E(x_t)$ i $E(x_t^2)$ konačni onda je i slabo stacionarna. Obratno ne važi.

Definicija 2.12 $\gamma_l = \text{cov}(x_t, x_{t-l})$ nazivamo autokovarijansom kašnjenja l vremenske serije $\{x_t\}$.

Lema 2.1 Za slabo stacionarnu vremensku seriju $\{x_t, t = 1, \dots, T\}$ važi $\gamma_0 = D(x_t)$ i $\gamma_{-l} = \gamma_l$.

Prilikom pravljenja modela potrebno je odrediti šta sve trebamo uvrstiti u njega. Tu do izražaja dolazi autokorelacija niza.

Definicija 2.13 Autokorelacija kašnjenja l vremenske serije $\{x_t, t = 1, \dots, T\}$ je

$$\rho_{x_t, x_{t-l}} = \frac{\text{cov}(x_t, x_{t-l})}{\sqrt{D(x_t) \cdot D(x_{t-l})}}$$

Ukoliko imamo slabu stacionarnost vremenske serije tada autokorelacija zavisi samo od l , odnosno

$$\rho_l = \frac{\text{cov}(x_t, x_{t-l})}{\sqrt{D(x_t)D(x_{t-l})}} = \frac{\gamma_l}{\sqrt{\gamma_0 \cdot \gamma_0}} = \frac{\gamma_l}{\gamma_0},$$

i važi $\rho_0 = 1$ i $\rho_l = \rho_{-l}$.

Definicija 2.14 Slabo stacionarna vremenska serija $\{x_t, t = 1, \dots, T\}$ je serijski nekorelisana ako i samo ako je $\rho_l = 0$, za svako $l > 0$.

Definicija 2.14 Uzoračka autokorelacija kašnjenja l vremenske serije $\{x_t, t = 1, \dots, T\}$ je

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^T (x_t - \bar{x})(x_{t-l} - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2},$$

gde je $\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t$ i $l \in [0, T - 1]$.

Testiranje autokorelacije za određenu vremensku seriju može se vršiti na dva načina. Prvi je pojedinačno testiranje za svako $l \in [0, T - 1]$. Testiramo nultu hipotezu $H_0(\rho_l = 0)$, tj. da x_t i x_{t-l} nisu korelisane, protiv alternativne $H_1(\rho_l \neq 0)$. Prilikom testiranja koristimo sledeću test statistiku (t ratio)

$$t_r = \frac{\hat{\rho}_l}{\sqrt{\frac{1 + 2 \cdot \sum_{i=0}^{l-1} \hat{\rho}_i^2}{T}}}$$

Nulta hipoteza se odbacuje na nivou poverenja α , ako je $|t_r| > z_{\alpha/2}$, gde je $z_{\alpha/2}$ kvantil standardne normalne raspodele. U suštini, izračunamo verovatnoću p za test statistiku t_r i odbacujemo nultu hipotezu ako je p manje od α .

Drugi način testiranja nam omogućava da proverimo da li postoji autokorelacija između više članova niza odjednom. Dakle, testiramo nultu hipotezu $H_0(\rho_1 = \dots = \rho_m = 0)$ protiv alternativne $H_1(\rho_i \neq 0)$ za neko $i \in \{1, \dots, m\}$. Koristimo sledeću test statistiku (Ljung-Boks)

$$Q(m) = T \cdot (T + 2) \sum_{l=1}^m \frac{\hat{\rho}_l^2}{T - l},$$

koja ima asimptotsku χ^2 raspodelu sa m stepeni slobode. Nulta hipoteza se odbacuje ukoliko je verovatnoća p Ljung-Boks test statistike manja nego nivo poverenja α . U praksi, izbor m može uticati na performance Ljung-Boks statistike. Često se uzima da je $m = \ln(T)$.

Definicija 2.15 *Vremenska serija $\{a_t, t = 1, \dots, T\}$ se naziva beli šum ako su a_t nekorelisane slučajne promenljive koje imaju jednaku raspodelu, gde su $E(a_t) = 0$ i $Var(a_t) = \sigma^2$.*

Prilikom pravljenja modela jedan od najbitnijih zadataka jeste pronalaženje odgovarajuće veze između određenih ulaznih i izlaznih informacija. Ukoliko model pravimo na osnovu stacionarnih vremenskih serija, jedna od najčešćih pretpostavki jeste da je veza između ulaznih i izlaznih parametara linearna. Beli šum je potpuno slučajan proces, koji na izvestan način korespondira slučajnoj grešci linearnog modela.

Definicija 2.16 *Vremenska serija $\{x_t, t = 1, \dots, T\}$ je linearna ako može da se predstavi u obliku*

$$x_t = \mu + \sum_{i=0}^{\infty} \psi_i \cdot a_{t-i},$$

gde je $\mu = \text{const.}$, ψ_i težinski koeficijenti, $\psi_0 = 1$, $\{a_t, t = 1, \dots, T\}$ nekorelisane slučajne promenljive koje imaju jednaku raspodelu sa očekivanjem 0 i konačnom disperzijom.

U ovako definisanoj vremenskoj seriji figuriše beskonačan broj nepoznatih težinskih koeficijenata, pa je cilj da se njihov broj ograniči na konačno mnogo.

Za formiranje kratkoročnih predikcija najčešće se koriste linearni modeli autoregresivnog tipa. Takođe, postoje metodi pomoću kojih možemo dobiti dugoročne predikcije. Neki od tih metoda su korišteni u ovom radu, a više o njima nalazi se u nastavku rada.

3.3. STARI metod i kriterijumi

Za dugoročnu prognozu koncentracije polena u vazduhu kao relevantni metod korišten je metod iz rada [3]. U ovom radu nazvaćemo ga STARI metod jer ćemo u nastavku rada navesti i druge metode koje ćemo nazvati NOVI metodi i koji sadrže određene modifikacije u cilju poboljšanja rezultata.

U svim metodima za predikciju koji su korišteni u radu, radi jednostavnosti, svaka godina se posmatra kao prestupna tj. dodaje se još jedan dan u februaru onim godinama koje to nisu tako da svaka ima 366 dana. Koncentracija polena je tada nula tako da dodavanje ne utiče na predikcije. Pre navođenja formule za dobijanje predikcija navešćemo oznake koje se koriste:

- r_i^k - koncentracija polena u danu i godine k
- r^k - vektor koncentracija u godini k
- stvarne koncentracije (PRAVE) - r
- predikcije - \hat{r}
- skup svih godina za koje imamo podatke - G
- testiranje - $G_k = G \setminus \{k\}$

Predikcija koncentracije polena u danu i godine k dobija se na sledeći način

$$\hat{r}_i^k = \frac{1}{|G_k|} \sum_{j \in G_k} m(a_i^j),$$

gde je sa m označena medijana vektora, a $a_i^j = (r_{i-b}^j, \dots, r_i^j, \dots, r_{i+b}^j)^T$ odnosno, to je vektor stvarnih koncentracija za godinu $j \in G_k$, a b broj dana pre i posle dana r_i koje zajedno sa njim stavljamo u vektor a_i^j . Pored medijane vektora u radu [3] korištena je i srednja vrednost vektora a_i^j u navedenom modelu. Međutim, u kasnijem poređenju dobijenih predikcija medijana se pokazala kao bolji izbor.

Broj dana b se određuje testiranjem prilikom čega b može da uzima vrednosti od 1 do 30. Dakle, jednu od godina iz skupa G , godinu k , izostavimo i za nju na osnovu preostalih podataka pravimo predikcije. Zatim testiramo koliko dobro smo predvideli vrednosti za godinu k . Postupak ponavljamo pri čemu svaki put uzimamo različite vrednosti za b . Testiranje se vrši pomoću sledeća tri kriterijuma:

- NRMSE (normalizovana srednje kvadratna greška)

$$NRMSE_k = \sqrt{\frac{\|\hat{r}^k - r^k\|^2}{366 \cdot (\bar{r}^k)^2}},$$

gde je $\bar{r}^k = \frac{1}{366} \sum_{i=1}^{366} r_i^k$.

Dobijamo vrednost koja nam pokazuje koliku grešku smo napravili u predikcijama ukoliko ih uporedimo sa stvarnim vrednostima.

- PROC (procenat podbacivanja)

$$PROC_k = \frac{1}{366} \sum_{i=1}^{366} I_i^k,$$

$$I_i^k = \begin{cases} 1, & \hat{r}_i^k < r_i^k \\ 0, & \hat{r}_i^k \geq r_i^k \end{cases}$$

Kada je u pitanju predviđanje koncentracije polena, većom greškom se smatra ukoliko predvidimo manju vrednost od stvarne nego veću. Ukoliko predvidimo manju vrednost, ljudi koji su alergični neće biti spremni da reaguju na adekvatan način na povećanu koncentraciju polena u vazduhu što im svakako može nasmetati više nego dodatni oprez i priprema. Dakle, procenat podbacivanja nam pokazuje za koliko dana u određenoj godini smo predvideli vrednost manju od stvarne vrednosti i samo se takvi dani „broje“.

- Spirmanovo ρ

$$\rho_k = \frac{Cov(p^k, \hat{p}^k)}{\sqrt{D(p^k) \cdot D(\hat{p}^k)}} \in (0,1),$$

pri čemu su p^k i \hat{p}^k sortirani vektori r^k i \hat{r}^k , respektivno. Sortiranje se vrši u neopadajućem poredku. Vrednost koju dobijemo nam govori u kakvoj su monotonij vezi predikcije i stvarne vrednosti odnosno, da li važi da je $\hat{p}_i^k < \hat{p}_j^k$ ($\hat{p}_i^k > \hat{p}_j^k$ ili $\hat{p}_i^k = \hat{p}_j^k$) ukoliko je $p_i^k < p_j^k$ ($p_i^k > p_j^k$ ili $p_i^k = p_j^k$), za svako $i < j$, tj. da li rast predviđenih vrednosti prati rast stvarnih vrednosti ili je njihovo kretanje obrnuto.

Nakon testiranja i upoređivanja dobijenih vrednosti dobijamo vrednost b za koju metod daje najtačnije predikcije. Za ovo istraživanje najbolje predikcije dobijamo ukoliko odaberemo da je $b = 11$.

Ove kriterijume ćemo koristiti i prilikom upoređivanja različitih metoda za predviđanje koncentracije polena. Postupak je analogan. Godinu k izostavimo iz skupa G , dobijamo skup G_k i na osnovu vrednosti iz tog skupa želimo da predvidimo koncentraciju polena za godinu k . Predikcije izračunamo pomoću metoda. Postupak se ponavlja onoliko puta koliko godina imamo u skupu G , odnosno svaka godina se po jednom bira kao test godina i za svaku računamo vrednosti $NRMSE$, $PROC$ i Spirmanovo ρ . Metode poredimo na osnovu prosečnih vrednosti. Metod je bolji ukoliko ima manju srednje kvadratnu grešku, manji procenat podbacivanja i veće Spirmanovo ρ . Takodje, vrednosti ovih kriterijuma ćemo računati i samo za predikcije koje se odnose na period sezone polena određene biljke. Dobijene vrednosti ćemo označavati sa $NRMSE_S$, $PROC_S$ i Spirmanovo ρ_S .

3.4. Problem najmanjih kvadrata

U NOVIM metodima koji predstavljaju modifikacije relevantnog metoda i koje ćemo navesti u sledećem poglavlju javlja se problem najmanjih kvadrata. Da bi izračunali predikcije pomoću tih metoda neophodno je prvo rešiti ovaj problem. Metod koja je korišten za rešavanje ovog problema je Levenberg-Marquardt metod [5].

Problem najmanjih kvadrata predstavlja jednu od značajnijih specijalnih klasa problema bez ograničenja i njegov zapis dat je u sledećoj formi

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \sum_{j=1}^m (r_j(x))^2, \quad (3.4.1)$$

gde se obično pretpostavlja da su $r_j : \mathbb{R}^n \rightarrow \mathbb{R}, j = 1, \dots, m$ glatke funkcije i često se nazivaju rezidualima. To su zapravo greške koje najčešće nestaju kada želimo da fitujemo podatke, odnosno da napravimo model koji aproksimira poznate vrednosti. U tom slučaju javlja se problem prilikom određivanja nepoznatih parametra u modelu na osnovu poznatih eksperimentalnih podataka. Na primer, pretpostavimo da postoji m merenja koje ćemo označiti sa $\{t_j, j = 1, \dots, m\}$ i da stvarne vrednosti promenljive y zavise od t . Ako se razmatra linearna regresija tada se funkcija modela može predstaviti u sledećem obliku $\Phi(x; t) = x^T t$, gde x predstavlja nepoznati vektor relevantnih koeficijenata, a grešku definišemo kao

$$r_j(x) = \Phi(x; t_j) - y_j.$$

Naš cilj je tada minimiziranje kumulativnog kvadrata prinosa predstavljenog gore navedenom formulom da bismo dobili model koji najbolje fituje poznate podatke.

Označimo sa r sledeću funkciju koja slika $\mathbb{R}^n \rightarrow \mathbb{R}^m$

$$r(x) = (r_1(x), \dots, r_m(x))^T.$$

Dalje, označimo Jakobijan ove funkcije sa J , tj.

$$J(x) := \nabla r(x).$$

Sada, problem (3.4.1) možemo predstaviti i sledećim ekvivalentnim zapisom

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} \|r(x)\|^2,$$

pri čemu su izvodi posmatrane funkcije sledeći

$$\nabla f(x) = J^T(x)r(x) = \sum_{j=1}^m r_j(x) \nabla r_j(x), \quad (3.4.2)$$

$$\nabla^2 f(x) = \sum_{j=1}^m \nabla r_j(x) \nabla^T r_j(x) + \sum_{j=1}^m r_j(x) \nabla^2 r_j(x) := G(x) + H(x). \quad (3.4.3)$$

Možemo da primetimo da je $G(x) = J^T(x)J(x)$.

Radi boljeg razumevanja prvo ćemo razmatrati problem linearnih najmanjih kvadrata gde je svaki rezidual linearna funkcija, a zatim ćemo objasniti problem nelinearnih najmanjih kvadrata. Primer za problem linearnih najmanjih kvadrata je upravo problem linearne regresije pomenut na početku. U tom slučaju, r se može predstaviti kao

$$r(x) = Ax + b, \quad (3.4.4)$$

gde $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Takođe, primetimo da je $J(x) = A$ i $\nabla f(x) = A^T(Ax + b)$. Sledi da je $\nabla^2 r_j(x) = 0$, $j = 1, \dots, m$ što implicira dalje da je $H(x) = 0$ i

$$\nabla^2 f(x) = A^T A.$$

Zaključujemo da je problem linearnih najmanjih kvadrata konveksan, što pojednostavljuje rešavanje problema jer stacionarna tačka funkcije f ne može biti maksimum. Zbog toga je kandidat za rešenje problema x^* tačka koja zadovoljava $\nabla f(x^*) = 0$. Problem linearnih najmanjih kvadrata se tada svodi na problem rešavanja sistema linearnih jednačina

$$A^T A x^* = -A^T b. \quad (3.4.5)$$

Sistem (3.4.5) se naziva sistem normalnih jednačina i obično se rešava numerički.

Sada ćemo razmatrati problem nelinearnih najmanjih kvadrata, tj. opšti slučaj za (3.4.1). Metod koji je korišten u radu za rešavanje ovog problema je Levenberg – Marquardt metod. Međutim, kako on predstavlja modifikaciju Gaus – Njutnovog metoda potrebno je prvo njega opisati.

Gaus – Njutnov pravac pretraživanja d^k je definisan sa

$$J^T(x^k)J(x^k)d^k = -J^T(x^k)r(x^k). \quad (3.4.6)$$

Primetimo da je ovo Kvazi – Njutnov metod ukoliko Hesijan aproksimiramo samo sa prvom sumom u (3.4.3), tj.

$$\nabla^2 f(x^k) \approx G(x^k).$$

Na ovaj način možemo da izbegnemo računanje drugog izvoda $\nabla^2 r_j(x)$ što može doneti značajne uštede u procesu optimizacije i to predstavlja jednu od prednosti ovog metoda. Takođe, $G(x^k)$ često dominira nad $H(x^k)$ što ukazuje na to da je navedena aproksimacija Hesijana dobra.

Još jedna prednost Gaus – Njutnovog pravca je njegovo opadajuće svojstvo. Zaista,

$$\begin{aligned}
(d^k)^T \nabla f(x^k) &= (d^k)^T J^T(x^k) r(x^k) \\
&= (d^k)^T (-J^T(x^k) J(x^k) d^k) \\
&= -\|J(x^k) d^k\|^2 \leq 0.
\end{aligned}$$

Prema (3.4.6), $J(x^k)d^k = 0$ implicira da je $J^T(x^k)r(x^k) = 0$, tj. $\nabla f(x^k) = 0$ pa zaključujemo da Gaus – Njutnov pravac ima opadajući pravac pod pretpostavkom da x^k nije stacionarna tačka. Konačno, primetimo da je d^k rešenje problema linearnih najmanjih kvadrata

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} \|J(x^k)d + r(x^k)\|^2,$$

pa se problem linearnih najmanjih kvadrata može posmatrati kao podproblem u slučaju problema nelinearnih najmanjih kvadrata.

Razmotrimo sada konvergenciju Gaus – Njutnovog metoda inkorporiranog u Armijo linijsko pretraživanje. Rezultat konvergencije predstavljen nakon Algoritma 3.1 posledica je Teoreme 3.1. koju ćemo prvu da navedemo. Navedene teoreme date su bez dokaza.

Teorema 3.1 *Pretpostavimo da $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^1(\mathbb{R}^n)$ i da je f ograničena odozdo. Takođe, pretpostavimo da je i niz pravaca pretraživanja $\{d^k\}_{k \in \mathbb{N}}$ ograničen. Tada, ili se Armijo linijsko pretraživanje zaustavlja nakon konačnog broja iteracija \bar{k} u stacionarnoj tački $x^{\bar{k}}$ ili je svaka tačka nagomilavanja niza $\{x^k\}_{k \in \mathbb{N}}$ stacionarna tačka funkcije f .*

Algoritam 3.1

Korak 0 Ulazni parametri: $x^0 \in \mathbb{R}^n, \beta, \eta \in (0, 1)$.

Korak 1 Inicijalizacija: $k = 0, x^k = x^0$.

Korak 2 Kriterijum za zaustavljanje: Ako je $\nabla f(x^k) = 0$ STOP. Inace, prelazimo na Korak 3.

Korak 3 Pravac pretraživanja: Izračunamo Gaus – Njutnov pravac d^k tako što rešimo (3.4.6).

Korak 4 Dužina koraka: Potrebno je pronaći najmanji nenegativni ceo broj j takav da $\alpha_k = \beta^j$ zadovoljava Armijov uslov

$$f(x^k + \alpha_k d^k) \leq f(x^k) + \eta \alpha_k \nabla^T f(x^k) d^k.$$

Korak 5 Ažuriranje: Postavimo $x^{k+1} = x^k + \alpha_k d^k$, $k = k + 1$ i idemo na Korak 2.

Teorema 3.2 *Pretpostavimo da $r \in C^1(\mathbb{R}^n)$ i da je nivo skup $\mathcal{L}(x^0) = \{x \in \mathbb{R}^n \mid f(x) \leq f(x^0)\}$ ograničen. Pretpostavimo da je $G(x)$ uniformno pozitivno definitna na otvorenom skupu koji sadrži $\mathcal{L}(x^0)$. Tada, ili se Algoritam 3.1 zaustavlja nakon konačnog broja iteracija \bar{k} u stacionarnoj tački $x^{\bar{k}}$ ili je svaka tačka nagomilavanja niza $\{x^k\}_{k \in \mathbb{N}}$ stacionarna tačka funkcije f .*

Jedna od ključnih pretpostavki koja čini Gaus – Njutnov metod konvergentnim je ta da je matrica sistema (3.4.6) (tj. Kvazi – Njutnova matrica) uniformno pozitivno definitna. Ta pretpostavka nas konačno dovodi do Levenberg – Marquardt metoda. Cilj ovog metoda je upravo da osigura da je Kvazi – Njutnova matrica pozitivno definitna. Da bi to postigli koristimo tzv. regularizaciju. Preciznije, imajući na umu da je matrica $G(x) = J^T(x^k)J(x^k) \succcurlyeq 0$, dodavanje konstantne pozitivno definitne matrice će obezbediti da su karakteristični koreni nove matrice udaljeni od nule. Ukoliko označimo sa $\rho > 0$ parametar regularizacije tada, koristeći osobine simetrične kvadratne matrice, dobijamo da najmanji karakteristični koren λ_{min} regularizovanog Hesijana zadovoljava

$$\lambda_{min}(J^T(x^k)J(x^k) + \rho I) \geq \lambda_{min}(J^T(x^k)J(x^k)) + \lambda_{min}(\rho I) \geq \rho.$$

Odgovarajući pravac pretraživanja Levenberg – Marquardt metoda se dobija kao rešenje sledećeg sistema

$$(J^T(x^k)J(x^k) + \rho I)d^k = -J^T(x^k)r(x^k).$$

Primetimo da je Levenberg – Marquardtov pravac jedinstveno određen dok Gaus – Njutnov pravac nije jer $G(x)$ može biti singularna. Analiza konvergencije Levenberg – Marquardt metoda može biti sprovedena kao u slučaju Gaus – Njutnovog metoda.

4. Istraživanje

4.1. Podaci, geografski položaj i klima

U atmosferi je prisutan polen velikog broja biljaka, ali po svojoj alergijskoj moći i zastupljenosti izdvaja se polen dvanaest tipova biljaka: *Ambrosia*, *Alnus*, *Artemisia*, *Betula*, *Amaranthaceae*, *Corylus*, *Cupressaceae/Taxaceae*, *Olea*, *Platanus*, *Poaceae*, *Quercus* and *Parietaria/Urtica*. Većina alergijskih reakcija u Evropi uzrokovana je polenom breze (*Betula*) i trave (*Poaceae*). Ambrozija (*Ambrosia*) je drugi najvažniji uzročnik sezonske astme i rinitisa u Severnoj Americi, tj. oblasti iz koje izvorno potiče. U Evropu je doneta ne nekom od prekookeanskih brodova krajem 19. veka i od tada se polako i nezadrživo širi njenom teritorijom. U prethodnoj deceniji štetni uticaj ambrozije na zdravlje stanovika Evrope značajno je porastao, a trenutno je ima više u Evropi nego u Americi. Procenjuje se da bi broj Evropljana koji su alergični na polen ambrozije mogao biti više nego dvostruk do 2060. godine.

Ambrozija, breza i trava se od navedenih dvanaest biljaka izdvajaju kao najvažnije za proučavanje, praćenje i predviđanje njihovih koncentracija polena u vazduhu. U ovom radu korišteni su podaci o koncentraciji polena u vazduhu za ove tri biljke merene na nivou sledeća 4 grada u Srbiji: Novi Sad (45°46' SGŠ i 19°20' IGD), Sombor (45°78' SGŠ i 19°12' IGD), Zrenjanin (45°22' SGŠ i 20°23' IGD) i Vrbas (45°34' SGŠ i 19°38' IGD). Koncentracija je beležena na svaka dva sata. Kako bi se osigurala regionalna reprezentativnost podataka uređaji za merenje su bili postavljeni iznad lokalnih izvora polena.

Glavni deo istraživanja sproveden je na osnovu podataka o koncentraciji polena koji se odnose na teritoriju Novog Sada i koji su sakupljeni od 2000. do 2017. godine. Ovaj skup podataka je najveći i najreprezentativniji pa samim tim i najpogoniji za ovaj vid istraživanja. Za Sombor podaci su sakupljeni od 2008. do 2017. godine, Zrenjanin takodje od 2008. do 2017. ali samo za ambroziju i travu i za Vrbas od 2009. do 2017. godine.

Po svom geografskom položaju, svi gradovi pripadaju Panonskoj niziji koja je okružena planinskim masivima. Na istoku se nalaze Karpati, na zapadu Alpi, na jugozapadu Dinarske planine i relativno je otvorena prema severu i severozapadu. Problem ovog geografskog položaja najviše je vezan za koncentraciju polena ambrozije u vazduhu jer je upravo u Panonskoj niziji ta koncentracija najviša u Evropi. Kako je područje ravničarsko, bez geografskih barijera ili visokih planina koje bi mogle da spreče migraciju i širenje biljnih vrsta pogodno je za useljavanje invazivnih i alergeničkih vrsta biljaka. Iz istih geografskih razloga kretanje čestica polena u vazduhu putem vazdušnih strujanja je takodje olakšano. Suprotno, lokalni izvori polena breze su ograničeni na drveće zasađeno kao ukrasno bilje na ulicama i u parkovima. Zbog toga aerobiološka situacija posebno zavisi od atmosferskog transporta polena breze iz udaljenih izvora polena. Kada je u pitanju polen trave, u okolini posmatranih gradova ne postoje veliki travnjaci i najveći dio izvora polena trave je ograničen na polja žitarica i javne i privatne zelene površine kojima se u velikoj meri upravlja.



Slika 4.1 Deo Panonske nizije u Srbiji



Slika 4.2 Panonska nizija sa označenim gradovima u kojima su vršena merenja (<http://www.zrenjanin.rs/userfiles/image/SrbijaKarta1.jpg>)

Posmatrani gradovi se nalaze u oblasti umereno kontinentalne klime. Velika godišnja kolebanja temperature vazduha su njena osnovna karakteristika. Pored toga karakterišu je i jeseni toplije od proleća, temperaturni prelaz od zime ka letu izraženiji od prelaza od leta ka zimi i tendencija pomeranja temperaturnog minimuma ka februaru, a maksimuma ka avgustu.

4.2. Koncentracija polena u vazduhu tokom godine

Od prirode biljne vrste i od parametara spoljašnje sredine kao što su temperatura, vlažnost vazduha, insolacija itd., zavisi kada, koliko dugo i koja količina polena će se naći u vazduhu. Posmatrane tri biljne vrste u ovom radu imaju različite sezone cvetanja i oslobađanja polenovog praha. Sezona ambrozije je od 7. do 9. meseca tokom godine, sezona breze od 3. do 5. i sezona trave od 4. do 10.

Na sledećim slikama, koje su preuzete sa sajta www.polleninfo.org, dat je prikaz količine polena u Evropi u različitim vremenskim periodima za posmatrane biljke. Prikaz je usklađen prema jačini očekivanih alergijskih simptoma. Bela boja prikazuju delove Evrope gde nema rizika za pojavu simptoma usled izloženosti datom tipu polena, zelena predstavlja oblasti sa malom količinom polena, a od žute do tamno crvene boje su prikazane oblasti sa umereno do veoma visokom povišenom količinom polena.

Prvo ćemo posmatrati ambroziju.



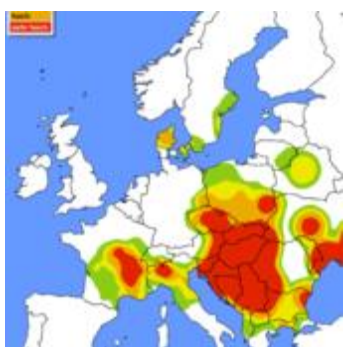
Slika 4.3 Ambrozija 12.-22. jul



Slika 4.4 Ambrozija 23.-31. jul



Slika 4.5 Ambrozija 1.-11. avgust



Slika 4.6 Ambrozija 1.-10. septembar



Slika 4.7 Ambrozija 21.-30. septembar



Slika 4.8 Ambrozija 21.-30. oktobar

Njena sezona kreće krajem jula, najveća koncentracija je početkom septembra, da bi se krajem istog meseca znatno smanjila. Neznatna količina se zadržava u vazduhu i u oktobru.

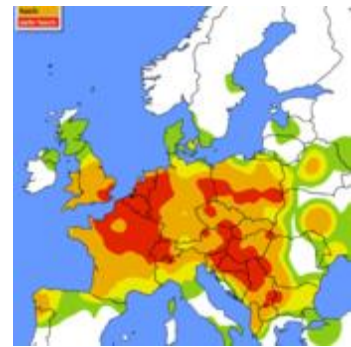
Ukoliko posmatramo polen breze imaćemo drugačiju situaciju. Cvetanje breze kreće sredinom marta, a sredinom aprila koncentracija polena dostiže svoj maksimum. Nakon toga koncentracija se smanjuje i krajem maja u području Srbije gde se nalaze posmatrani gradovi nemamo opasnosti od alergijskih reakcija.



Slika 4.9 Breza 1.-11. mart



Slika 4.10 Breza 12.-22. mart



Slika 4.11 Breza 23.-31. mart



Slika 4.12 Breza 11.-20. april



Slika 4.13 Breza 12.-22. maj



Slika 4.14 Breza 23.-31. maj

Sezona polena trave je najduža ukoliko uporedimo sa prostale dve biljne vrste koje posmatramo. Polen je javlja već u aprilu, koncentracija se postepeno povećava i dostiže svoj maksimum u junu, a zatim se postepeno smanjuje i sezona završava krajem septembra ili početkom oktobra. Na sledećim slikama možemo da vidimo kretanje koncentracije polena trave na opisan način.



Slika 4.15 Trava 11. – 20. april



Slika 4.16 Trava 12. – 22. maj



Slika 4.17 Trava 1. – 10. jun



Slika 4.18 Trava 12. – 22. jul



Slika 4.19 Trava 23. – 31. avgust



Slika 4.20 Trava 23. – 31. septembar

4.3. Metode za popunjavanje nedostajućih podataka

U toku merenja i prikupljanja podataka dešava se da u celoj bazi imamo podatke koji nedostaju. Do takvih „rupa“ u podacima može da dodje usled kvara na mernom uređaju, problema sa strujnim napajanjem ili nekog drugog problema koji rezultuje neizmerenom koncentracijom polena za određeni deo dana, ceo dan, a ponekad, u zavisnosti od uređaja kojim se meri koncentracija, možemo imati i nedostauće podatke za sedam ili više dana.

Nedostajući podaci mogu vrlo negativno da utiču na konstrukciju modela za predviđanje određenih vrednosti, a samim tim i na dobijanje netačnih predikcija. Zbog toga su nam tačni podaci osnova za dalji rad. Da bi smo mogli predvideti koncentraciju polena za bilo koji dan sa najvećom mogućom tačnošću nedostajuće podatke je potrebno na određeni način popuniti.

Podaci nad kojima je sprovedeno glavno istraživanje su mereni na teritoriji Novog Sada. Bazu podataka čine dvočasovne koncentracije polena, merene svakoga dana u periodu od 2.2.2000. do 30.9.2017. godine. Dakle, u bazi podataka bi trebalo da se nalazi 78900 vrednosti. Međutim, među njima se nalazio veliki broj nedostajućih podataka.

U prethodnim, sličnim, istraživanjima testirane su različite metode za interpolaciju nedostajućih podataka. Od svih metoda linerana interpolacija se pokazala kao nabolji metod, pa je ona korištena za interpolaciju i u ovom istraživanju kako za Novi Sad tako i za ostale gradove.

4.4. Statističke karakteristike podataka

Pre nego što nastavimo dalji rad sa podacima koje imamo za posmatrane gradove ispitaćemo neke njihove statističke osobine. Ispitivanje je vršeno na prosečnim dnevnim koncentracijama polena, a osobine koje su ispitane su sledeće:

- Raspodela
- Korelacija
- Autokorelacija

Pod prvom ispitanom osobinom, raspodelom, podrazumeva se ispitivanje da li podaci imaju normalnu raspodelu ili ne. To je učinjeno testiranjem vrednosti koeficijenta asimetrije, koeficijenta ekscesa i testiranjem obe vrednosti odjednom pomoću Jarque i Bera testa. Testiranje je vršeno na nivou značajnosti 0,05 i 0,01. Način testiranja je detaljnije opisan u prethodnom poglavlju.

Za sva četiri posmatrana grada, na osnovu tri navedena testa i oba nivoa značajnosti dobijen je isti rezultat. Pretpostavka da podaci imaju normalnu raspodelu se odbacuje. Dakle, koncentracije polena ni za jednu od tri biljke nemaju normalnu raspodelu.

Druga navedena osobina, korelacija, testirana je pomoću Pirsonove, Spirmanove i Kendalove korelacije. Pošto smo za sve tri mere dobili slične rezultate, predstavice samo rezultate dobijene korišćenjem Pirsonovog koeficijenta korelacije. Više o ovom koeficijentu se nalazi u prethodnom poglavlju.

Testiranje korelacije izvršeno je za celu seriju podataka za Novi Sad, tj. za period od 2000. do 2017. godine. Dobijeni rezultati se nalaze u sledećoj tabeli:

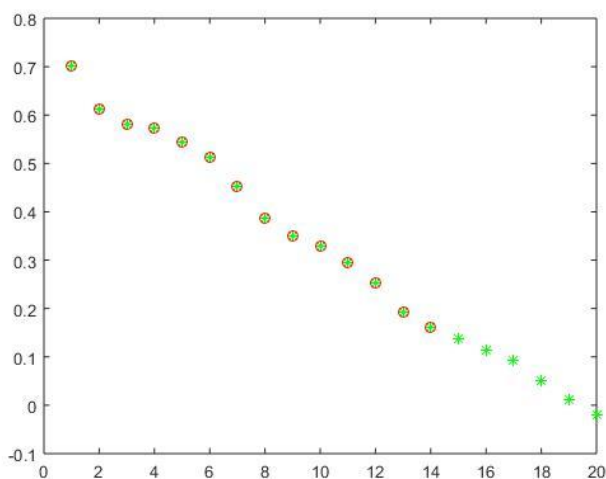
	Polen/ m^3	Temp - min	Temp - max	Temp - prosečna	Vlažnost vazduha	Padavine	Brzina vetra
Polen/ m^3	1	0.222	0.201	0.202	- 0.075	0.005	- 0.052
Temp_min	0.222	1	0.909	0.952	- 0.438	0.060	- 0.138
Temp_max	0.201	0.909	1	0.982	- 0.667	- 0.069	- 0.222
Temp_prosečna	0.202	0.952	0.982	1	- 0.606	- 0.027	- 0.207
Vl. vazduha	- 0.075	- 0.438	- 0.667	- 0.606	1	0.254	- 0.005
Padavine	0.005	0.060	- 0.069	- 0.027	0.254	1	0.121
Brzina vetra	- 0.052	- 0.138	- 0.222	- 0.207	- 0.005	0.121	1

Tabela 4.1 Rezultati dobijeni pomoću Pirsonovog testa korelacije

Možemo da zaključimo da osim temperature, za koju se očekuje da bude u visokoj korelaciji, jedina druga značajna korelacija je negativna korelacija između temperature i vlažnosti vazduha.

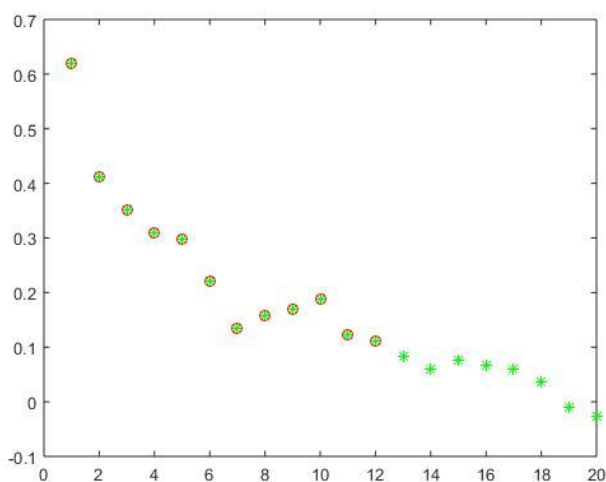
Poslednja osobina, autokorelacija, ispitana je na dva načina koja su takođe opisana u prethodnom poglavlju. Prvi način testira povezanost količine koncentracije polena u jednom danu sa količinom koja je izmerena u vazduhu u prethodnih 20 dana. Testiranje se vrši za svaki od prethodnih dana posebno. Drugi način testiranja nam omogućava da proverimo da li postoji autokorelacija između više članova niza odjednom. Rezultat ovog testiranja je da ne postoji povezanost sa koncentracijama izmerenim u prethodnih 20 dana ili da postoji povezanost, ali ne znamo sa kojim tačno od prethodnih dana. Oba načina testiranja su vršena na nivou značajnosti 0,05 i 0,01. Takođe, testiranje je vršeno samo za one podakte koji se odnose na sezonu odgovarajuće biljke u toku godine. Oni podaci koji ne pripadaju sezoni nisu razmatrani jer je tada koncentracija polena uglavnom nula ili jako blizu nule.

Drugi od navedenih testova je za svaki od testiranih gradova za sve tri biljke pokazao da postoji povezanost sa nekim od prethodnih 20 dana. Nakon toga korišten je prvi test koji daje detaljnije rezultate. Konačno, nakon izvršene analize za sva četiri posmatrana grada, dobijeni su slični rezultati pa su zbog preglednosti rada u nastavku prikazani samo grafici koji se odnose na testove sprovedene za Novi Sad na nivou značajnosti 0.05.



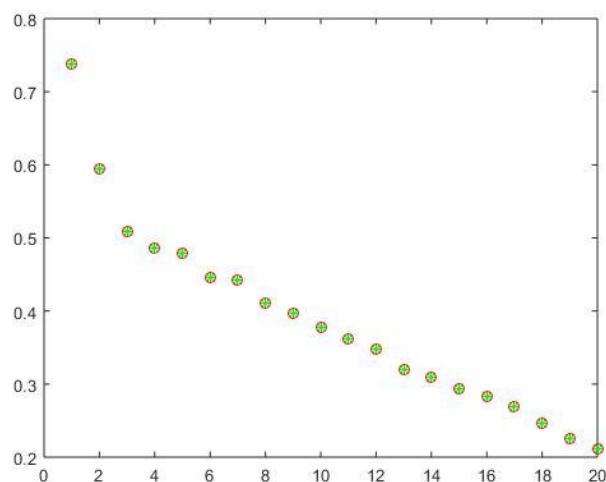
Grafik 4.1 Autokorelacija prosečnih dnevnih vrednosti polena ambrozije u Novom Sadu

Na Grafiku 4.1 možemo da primetimo da postoji statistički značajna korelacija sa prethodnih 14 dana ukoliko posmatramo prosečne dnevne vrednosti polena ambrozije u Novom Sadu.



Grafik 4.2 Autokorelacija prosečnih dnevnih vrednosti polena breze u Novom Sadu

Za brezu imamo malo drugačije vrednosti koje su prikazane na Grafiku 4.2. Autokorelacija je statistički značajna sa prethodnih 12 dana.



Grafik 4.3 Autokorelacija prosečnih dnevnih vrednosti polena trave u Novom Sadu

U slučaju trave, autokorelacija je statistički značajna sa svih 20 prethodnih dana. Grafik 4.3 daje detaljniji uvid u pojedinačne vrednosti autokorelacije.

Testiranje autokorelacije je izvršeno i pomoću serije logaritamskih prinosa $\{r_t\}_{t=1}^{T-1}$ koji se dobijaju na sledeći način:

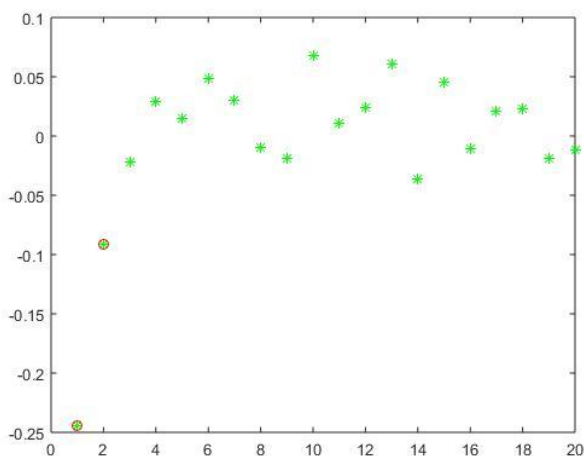
$$r_t = \ln\left(\frac{y_t - y_{t-1}}{y_{t-1}} + 1\right) = \ln\frac{y_t}{y_{t-1}} = \ln y_t - \ln y_{t-1}$$

pri čemu vremenska serija $\{y_t\}_{t=1}^T$ predstavlja vremensku seriju prosečnih dnevnih koncentracija polena u sezoni za odgovarajuću biljku ali uvećanu za 1, tj.

$$y_t = \textit{presečna dnevna koncentracija} + 1,$$

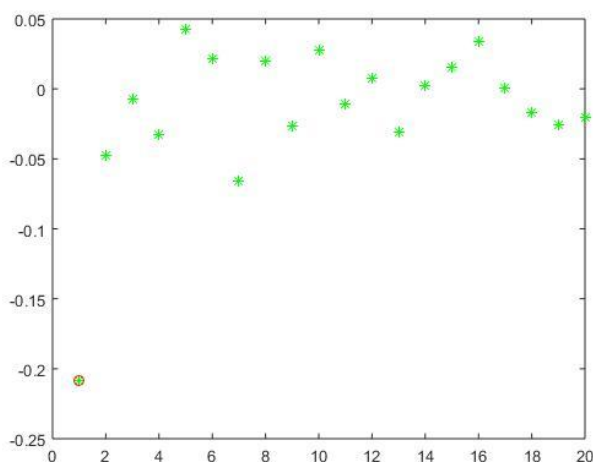
za svako $t = 1, \dots, T$. Na taj način izbegavamo pojavljivanje nule u logaritamskoj funkciji.

Za sva četiri posmatrana grada dobijeni su slični rezultati, pa i grafici, ukoliko ih poredimo na osnovu biljaka pa ćemo, kao i u prethodnom slučaju, prikazati samo grafike koji se odnose na Novi Sad i testiranje koje je sprovedeno na nivou značajnosti 0,05.



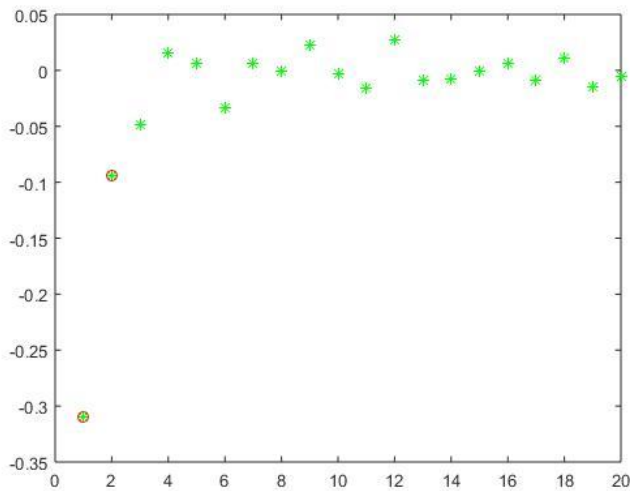
Grafik 4.4 Autokorelacija logaritamskih prinosa prosečnih dnevnih vrednosti polena ambrozije u Novom Sadu

Na Grafiku 4.4 koji se odnosi na ambroziju uočavamo da postoji statistički značajna autokorelacija sa prethodne dve vrednosti, ali u ovom slučaju autokorelacija je negativna.



Grafik 4.5 Autokorelacija logaritamskih prinosa prosečnih dnevnih vrednosti polena breze u Novom Sadu

Autokorelacija logaritamskih prinosa koji se odnose na brezu statistički je značajna samo sa prethodnom vrednošću i pri tom autokorelacija je negativna što možemo da vidimo na Grafiku 4.5.



Grafik 4.6 Autokorelacija logaritamskih prinosa prosečnih dnevnih vrednosti polena trave u Novom Sadu

Za travu takođe imamo negativne vrednosti statistički značajne autokorelacije. Na odgovarajućem Grafiku 4.6 možemo da ih uočimo. Autokorelacija je statistički značajna samo sa prethodne 2 vrednosti.

Testirane statističke osobine omogućavaju detaljniji uvid u međusobni odnos podataka o koncentraciji polena, kao i u njihov odnos sa drugim parametrima u prirodi koji mogu uticati na kretanje polena i njegovu vrednost u atmosferi na određenom geografskom položaju. Podaci o autokorelaciji mogu biti od značaja za kratkoročnu prognozu. Kako je predviđanje koncentracije polena veoma kompleksno, podaci o dugoročnoj prognozi takođe mogu biti od velikog značaja za kratkoročnu prognozu u smislu odgovarajućih ulaznih parametara u modele za kratkoročne predikcije. U ovom radu ćemo se zadržati na modelima za dugoročne predikcije, te u nastavku navodimo preostale modele koji su korišteni za predikcije, a potom i njihove rezultate.

4.5. NOVI metodi

U prethodnom poglavlju naveli smo STARI metod koji predstavlja relevantni metod za ovo istraživanje. Sada ćemo navesti tri nova metoda koja su definisana u ovom istraživanju i koja sadrže određene modifikacije u odnosu na STARI metod sa ciljem poboljšanja rezultata, odnosno dobijanja preciznijih predikcija.

Navešćemo još jednom potrebne oznake:

- r_i^k - koncentracija polena u danu i godine k
- r^k - vektor koncentracija u godini k
- stvarne koncentracije (PRAVE) - r
- predikcije - \hat{r}
- skup svih godina za koje imamo podatke – G
- testiranje - $G_k = G \setminus \{k\}$

Testiranje sve vrši na taj način sto jednu od godina iz skupa G izbacimo iz baze podataka i pretpostavimo da predviđamo njene vrednosti. Stvarne vrednosti izmerene za tu godinu nam služe da na osnovu kriterijuma procenimo koji metod daje najbolje predikcije.

Metod NOVI koristi sledeću formulu za računanje predikcije koncentracije polena u danu i godine k :

$$\hat{r}_i^k = \frac{1}{|G_k|} \sum_{j \in G_k} (\omega^T a_i^j + c),$$

gde je $a_i^j = (r_{i-b}^j, \dots, r_i^j, \dots, r_{i+b}^j)^T$ isto kao u STAROM metodu, vektor stvarnih koncentracija za godinu $j \in G_k$. Broj dana b se određuje na način koji je naveden za STARI metod. Za NOVE metode utvrđeno je da daju najbolje predikcije ukoliko izaberemo da je $b = 9$.

Vektori ω i c dobijaju se rešavanjem sledećeg problema optimizacije bez ograničenja

$$\min_x f(x) := \frac{1}{|G|} \sum_{k \in G} \frac{\|\hat{r}^k(x) - r^k\|^2}{\bar{r}^k \sqrt{366}},$$

$$x = (\omega; c).$$

Dakle, rešavanjem ovog problema dobijamo vektor ω i broj c za koje je NRMSE najmanja, odnosno, pošto je funkcija $f(x)$ diferencijabilna rešavamo problem najmanjih kvadrata. Metod koja je korišten za rešavanje ovog problema je Levenberg-Marquardt metod [5].

Sledećom modifikacijom dobijamo metod NOVI1, čija formula za računanje predikcija izgleda ovako:

$$\hat{r}_i^k = \frac{1}{|G_k|} \sum_{j \in G_k} (\omega^T \text{sort}(a_i^j) + c).$$

Dakle, razlika u odnosu na metod NOVI je u tome što su članovi vektora a_i^j sortirani u neopadajućem poredku. Vektor ω i broj c dobijamo na isti način kao u metodu NOVI. Ukoliko pak ω i c odaberemo na sledeći način:

- $\omega = (0, \dots, 1, \dots, 0)$, gde se 1 nalazi na i -tom mestu,
- $c = 0$

metod NOVI1 se svodi na metod STARI, tj. možemo primetiti da je metod NOVI1 generalizacija metoda STARI.

Treća modifikacija, odnosno metod NOVI2, u sebi sadrži odgovarajuće podatke o temperaturi, a formula je sledeća:

$$\hat{r}_i^k = \sum_{j \in G_k} g_j^k (\omega^T \text{sort}(a_i^j) + c).$$

Ovaj metod je takođe generalizacija prethodnih, tj. za:

- $\omega = (0, \dots, \frac{1}{|G_k|}, \dots, 0)$, gde se $\frac{1}{|G_k|}$ nalazi na i -tom mestu i $c = 0$ svodi se na metod STARI
- $g_j = (\frac{1}{|G_k|}, \dots, \frac{1}{|G_k|}, \dots, \frac{1}{|G_k|})$ svodi se na metod NOVI1.

Za biljke koje posmatramo poznati su sledeći podaci:

	predsezona	sezona
Ambrozija	1. – 6. Meseca (182 dana)	7. – 9. meseca
Breza	1. – 2. Meseca (60 dana)	3. – 5. meseca
Trava	1. – 3. Meseca (91 dan)	4. – 10. meseca

Sa t_i^k označavaćemo prosečnu dnevnu temperaturu u danu i godine k . Vektor $t^k(p) = (t_1^k, \dots, t_p^k)^T$ predstavlja vektor prosečnih temperatura u predsezoni i različitih je dimenzija za različite biljke. Ideja ovog modela jeste da godinu, koju smo izostavili iz baze podataka i za koju želimo predvideti vrednosti polena pomoću modela, uporedimo na osnovu temperature sa ostalim godinama koje su preostale u bazi. To radimo tako što prvo formiramo sledeći koeficijent:

$$s_j^k = \|t^k(p) - t^j(p)\|^{-1}, j \in G_k.$$

Što je veća sličnost između godina j i k , to je koeficijent s_j^k veći. Sledeći korak jeste normiranje s_j^k :

$$g_j^k = \frac{s_j^k}{\sum_{i \in G_k} s_i^k} \in (0,1)$$

Na taj način dobijamo težinske koeficijente u metodu NOVI2 koji su formirani na osnovu podataka o temperaturi za Novi Sad.

Preostala 3 grada, Vrbas, Sombor i Zrenjanin, se nalaze na takvom geografskom položaju i relativno blizu Novog Sada da se koefijenti koji su dobijeni za Novi Sad mogu koristiti i u predikcijama za ostale gradove. Dnevna temperatura između posmatranih gradova se ne razlikuje značajno, odnosno očuvava se ista sličnost između godina na osnovu temperature za preostala 3 grada kao i za Novi Sad i zbog toga nije potrebno koeficijente računati za svaki grad posebno.

U nastavku rada iznecemo rezultate navedenih metoda, kao i njihovu efikasnost u dugoročnom predviđanju koncentracije polena ambrozije, breze i trave na teritoriji posmatranih gradova. Metode smo poredili na osnovu kriterijuma navedenih u prethodnom poglavlju. Dakle, celokupno istraživanje, formiranje metoda i koeficijenata vezanih za temperaturu, odabir kriterijuma, itd. je sprovedeno na osnovu podataka o koncentraciji polena na teritoriji Novog Sada. Zatim smo te iste metode i način njihovog poređenja primenili na ostale gradove.

5. Rezultati

5.1. Novi Sad

Dugoročna prognoza koncentracije polena za ambroziju, brezu i travu na teritoriji grada Novi Sad, kao i na teritoriji ostalih gradova, vršena je pomoću tri metoda: STARI, NOVI1 i NOVI2. Metode smo poredili na osnovu 6 kriterijuma: NRMSE, PROC, Spirmanovo ρ , NRMSE_S, PROC_S i Spirmanovo ρ_S . Podaci o dvočasovnim koncentracijama beleženi su u periodu od 2000. do 2017. godine i ukupan broj podataka nad kojim je sprovedeno glavno istraživanje iznosi 78900 za svaku od biljaka.

Ukoliko posmatramo predikcije ambrozije, metodi NOVI1 i NOVI2 po svim kriterijumima daju bolje predikcije od metoda STARI (Tabela 5.1). Rezultati za ova dva nova metoda se razlikuju samo kod jednog kriterijuma, odnosno metod NOVI2 ima manji procenat podbacivanja u sezoni od metoda NOVI1, pa na osnovu toga možemo reći da je za predviđanje koncentracije polena ambrozije najbolje koristiti metod NOVI2. Za preostale dve biljke imamo drugačiju situaciju. Za predikcije breze STARI metod ima najmanju srednje kvadratnu grešku, međutim isti metod ima najveći procenat podbacivanja i najmanje Spirmanovo ρ bilo da posmatramo sve predikcije ili samo predikcije u sezoni. Dakle, različiti kriterijumi daju suprotne rezultate i zbog toga ne možemo jednoznačno izdvojiti najbolji metod u slučaju predviđanja koncentracije polena breze. Takođe, ukoliko posmatramo rezultate kriterijuma za travu ne možemo izdvojiti najbolji metod jer za svaki kriterijum različit metod daje najbolje rezultate. Dakle, u slučaju breze i trave potrebna je detaljnija analiza rezultata. Odabir metoda se vrši u zavisnosti od toga za koju biljnu vrstu pravimo predikcije i kojem kriterijumu pridajemo više značaja.

	Ambrozija					
	NRMSE	PROC	ρ	NRMSE _S	PROC _S	ρ_S
STARI	2,46	0,21	0,91	4,90	0,42	0,88
NOVI1	2,43	0,19	0,94	4,84	0,36	0,92
NOVI2	2,43	0,19	0,94	4,84	0,35	0,92
	Breza					
	NRMSE	PROC	ρ	NRMSE _S	PROC _S	ρ_S
STARI	5,41	0,11	0,77	10,80	0,35	0,75
NOVI1	5,57	0,09	0,78	11,11	0,27	0,77
NOVI2	5,57	0,09	0,79	11,11	0,28	0,79
	Trava					
	NRMSE	PROC	ρ	NRMSE _S	PROC _S	ρ_S
STARI	1,62	0,27	0,87	2,12	0,44	0,86
NOVI1	1,60	0,24	0,92	2,09	0,39	0,91
NOVI2	1,59	0,25	0,91	2,08	0,40	0,91

Tabela 5.1 Poređenje metoda – tabela proseka za Novi Sad

Poređenje metoda možemo izvršiti i na drugačiji način (Tabela 5.2), tj. na osnovu verovatnoća uspešnosti metoda za određenu biljnu vrstu. Dakle, da bi izračunali potrebne verovatnoće posmatramo rezultate svih metoda na osnovu istih 6 kriterijuma, ali za svaku godinu posebno. U slučaju Novog Sada krećemo od 2000. godine i kriterijuma NRMSE. Onom metodu koji za datu biljnu vrstu ima najmanju vrednost NRMSE, odnosno onom koji daje najbolje predikcije za ovu godinu i po ovom kriterijumu, dodeljujemo vrednost 1, dok ostala dva metoda dobijaju vrednost nula. Postupak ponavljamo za preostale kriterijume. Zatim sve ponovimo za preostale godine. Sledeći korak je da prebrojimo koliko puta, odnosno za koliko godina je svaki od metoda, posebno po svakom od kriterijuma, bio najbolji i dobijene vrednosti podeljimo sa ukupnim brojem godina. Na taj način dobijamo verovatnoće uspešnosti za svaki od metoda na osnovu pomenutih kriterijuma. Drugim rečima, dobijamo vrednosti koje nam govore kolika je verovatnoća da ćemo dobiti dobre predikcije pomoću jednog od metoda u odnosu na preostala dva. Za pojedine kriterijume zbir verovatnoća nije 1 jer se dešava da više metoda ima iste vrednosti pa umesto samo jednom, vrednost 1 dodeljujemo u korist dva ili čak sva tri metoda.

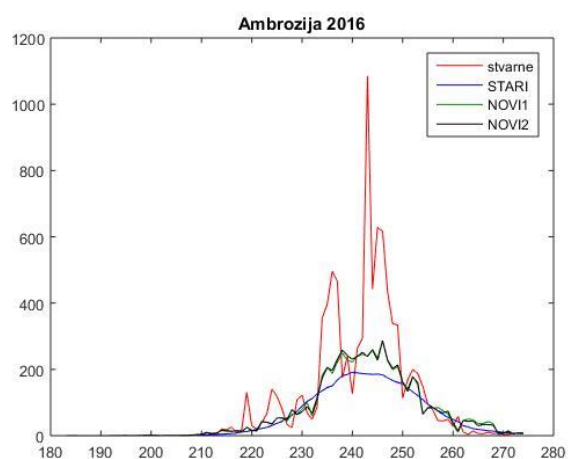
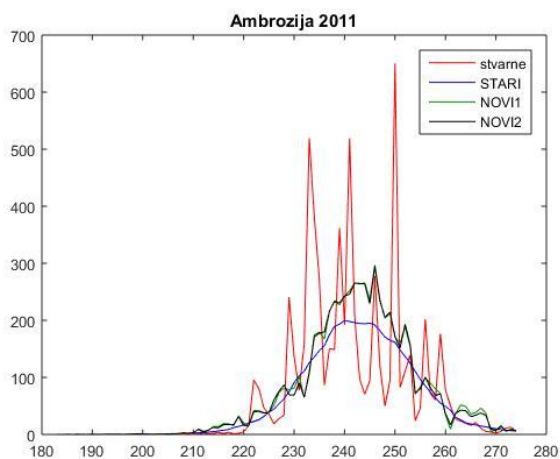
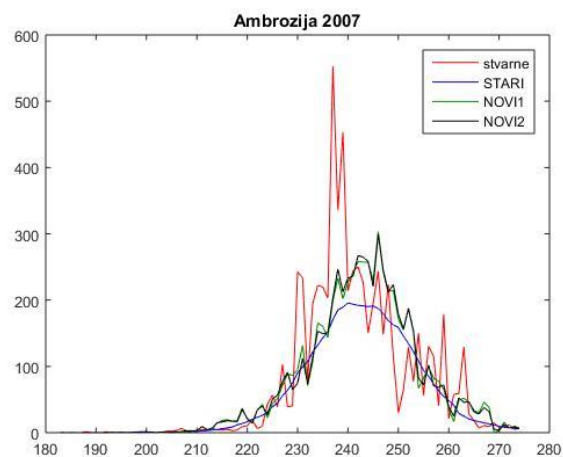
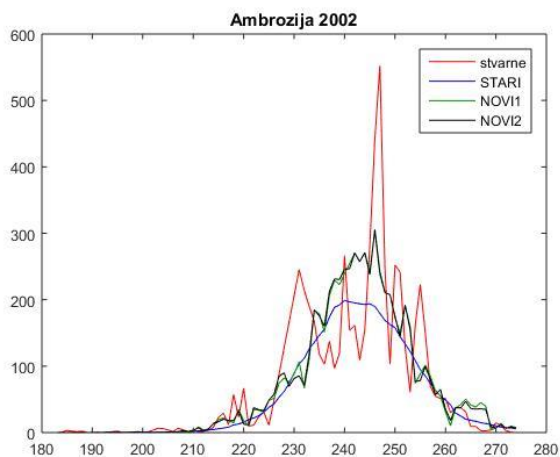
Ovakvo poređenje, kao ni prethodno, ne izdvaja ni jedan metod kao najbolji za sve tri vrste biljaka. Dakle, ponovo je potrebna detaljnija analiza i odluka o izboru metoda zavisi od više parametara. Da bi na neki način mogli uporediti metode izračunali smo prosečne vrednosti verovatnoća uspešnosti svakog metoda na osnovu jednog kriterijuma za sve tri biljke (kolona ABT). Ukoliko posmatramo kriterijume NRMSE i $NRMSE_S$ metod NOVI2 se izdvaja kao najbolji, PROC i $PROC_S$ izdvajaju metod NOVI1, dok Spirmanovo ρ i Spirmanovo ρ_S izdvajaju različite metode, tj. NOVI 2 i NOVI1 respektivno. Poslednje tri kolone Tabele 3.2., izuzev prve vrste, pokazuju srednje vrednosti verovatnoca uspešnosti svakog metoda za svaku biljku pojedinačno na osnovu svih 6 kriterijuma. Dakle, ovakav način poređenja rezultata omogućava nam kombinovanje različitih kriterijuma. Uočavamo da metod NOVI2 daje najbolje predikcije za ambroziju i brezu, dok za travu to čini metod NOVI1. Na kraju kada izračunamo uspešnost predikcija modela za sve tri biljke zajedno na osnovu svih 6 kriterijuma dobijemo da metod NOVI1 ima najbolju ocenu, odnosno najveću verovatnoću da će predvideti dobre vrednosti. Međutim, kao što je napomenuto ovakav rezultat trebamo uzimati sa oprezom i uvek izvršiti detaljniju analizu. Odabir metoda svakako treba vršiti u zavisnosti od biljne vrste za koju želimo izračunati dugoročne predikcije, kao i od kriterijuma koji smo odabrali za poređenje metoda.

		ABT	A	B	T
NRMSE	STARI	0,30	0,33	0,22	0,33
	NOVI1	0,31	0,39	0,39	0,17
	NOVI2	0,39	0,28	0,39	0,50
PROC	STARI	0,00	0,00	0,00	0,00
	NOVI1	0,61	0,39	0,67	0,78
	NOVI2	0,56	0,89	0,50	0,28
ρ	STARI	0,02	0,00	0,06	0,00
	NOVI1	0,48	0,50	0,17	0,78
	NOVI2	0,50	0,50	0,78	0,22
NRMSE _s	STARI	0,30	0,33	0,22	0,33
	NOVI1	0,31	0,39	0,39	0,17
	NOVI2	0,39	0,28	0,39	0,50
PROC _s	STARI	0,02	0,06	0,00	0,00
	NOVI1	0,63	0,33	0,78	0,78
	NOVI2	0,48	0,83	0,33	0,28
ρ_s	STARI	0,00	0,00	0,00	0,00
	NOVI1	0,54	0,61	0,17	0,83
	NOVI2	0,46	0,39	0,83	0,17
		ABT	A	B	T
STARI		0,10	0,12	0,08	0,11
NOVI1		0,48	0,44	0,43	0,58
NOVI2		0,46	0,53	0,54	0,32

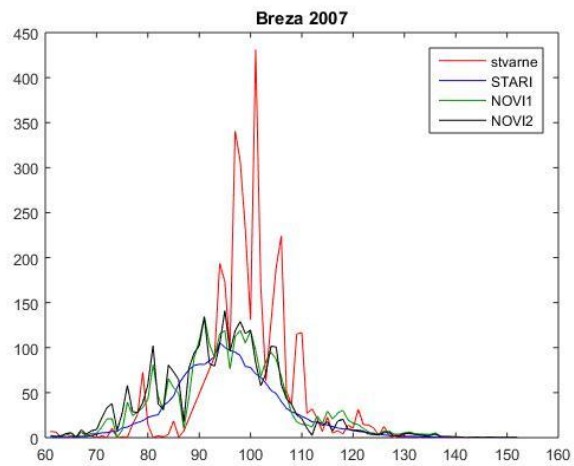
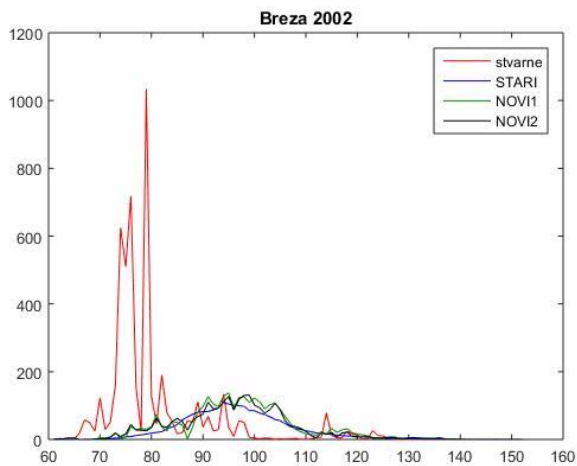
Tabela 5.2. Verovatnoće uspešnosti metoda

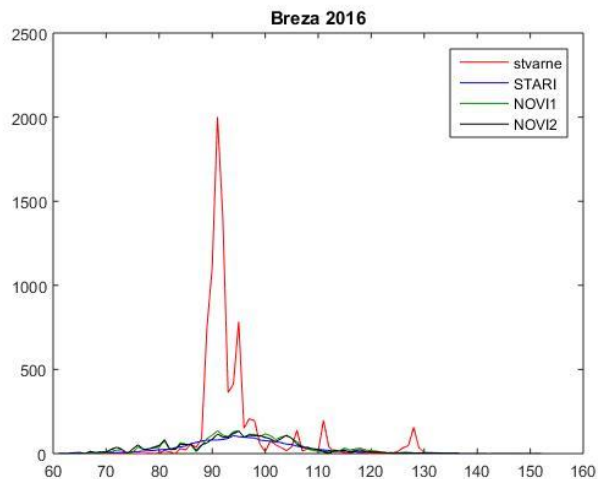
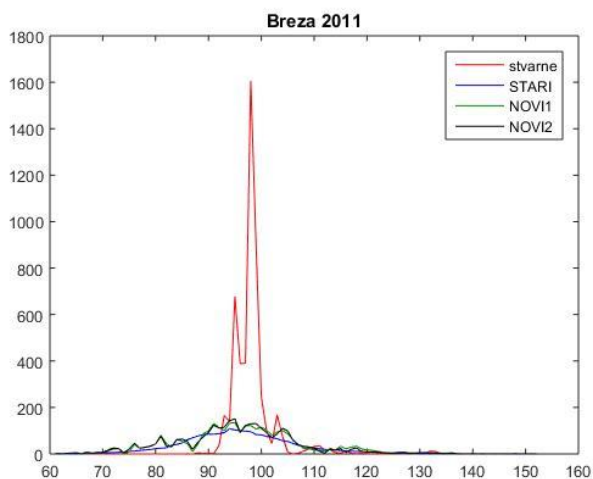
Grafici koji slede u nastavku rada daju prikaz stvarnih vrednosti koncentracije polena i vrednosti predviđenih pomoću metoda. Zbog preglednosti rada, za svaku biljnu vrstu prikazaćemo grafike za 4 godine koje su odabrane na slučajan način. To su 2002., 2007., 2011. i 2016. godina. Na graficima možemo da primetimo koliko dobro koji metod prati stvarne vrednosti. Možemo uočiti da je najteže predvideti iznenadne skokove koncentracije polena. Problem u njihovom predviđanju nastaje zbog toga što veliki broj faktora u atmosferi utiče na to koliko polena će se naći u atmosferi na određenom geografskom prostoru u određenom vremenskom periodu i iznenadna promena tih faktora dovodi do iznenadnih i nepredvidljivih skokova koncentracije polena.

Prvo ćemo prikazati grafike za ambroziju. Uočavamo da stvarne vrednosti variraju u zavisnosti od toga koju godinu posmatramo. Za određene godine imamo samo po jedan skok na grafiku, dok za neke imamo znatno veći broj istih. Maksimalna prosečna dnevna vrednost koncentracije polena ambrozije na teritoriji grada Novi Sad zabeležena je 28.08.2001. godine i iznosila je $2105/m^3$ polenovih zrna.

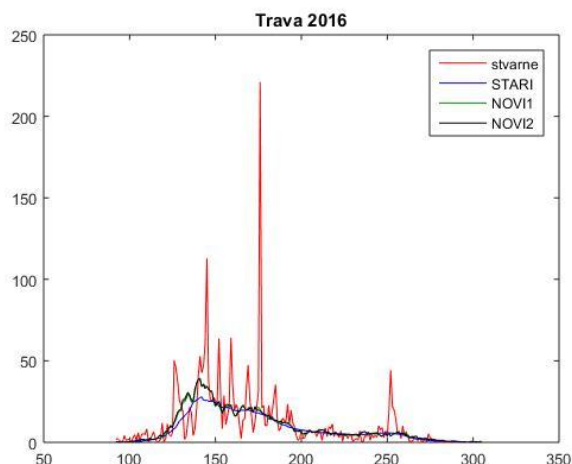
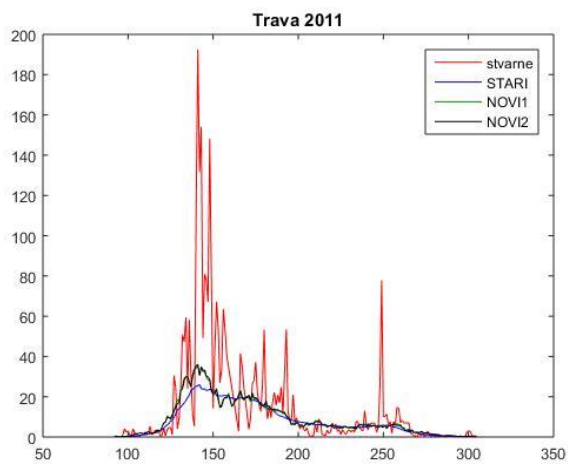
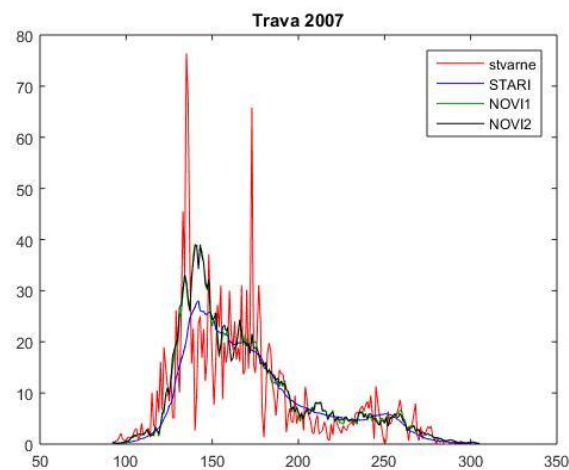
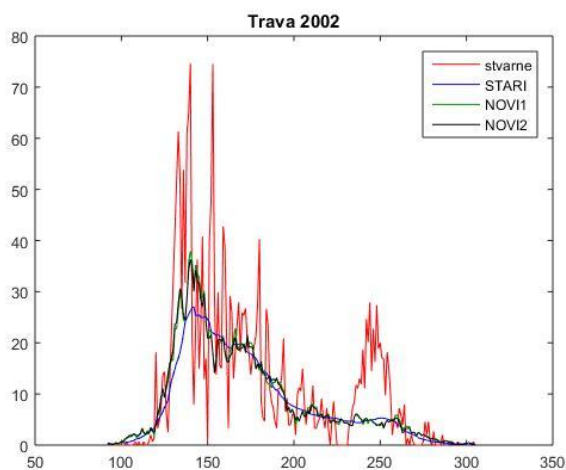


Sljedeća 4 grafika daju prikaz stvarnih i predviđenih koncentracije polena breze. Maksimalna prosečna dnevna vrednost koncentracije polena za ovu biljku izmerena je 22.03.2014. godine i iznosila je $2941/m^3$ polenovih zrna.





Poslednja 4 grafika predstavljaju stvarne i predvidjene koncentracije polena trave. Za nju je maksimalna prosečna dnevna vrednost koncentracije polena izmerena 24.06.2016. i iznosila je $221/m^3$.



5.2. Vrbas

Podaci o koncentraciji polena za ambroziju, brezu i travu u Vrbasu beleženi su od 2009. do 2017. godine. Pošto se radi o dvočasovnim koncentracijama, ukupan broj podataka iznosi 39444 za svaku od biljaka. Na isti način, kao i za Novi Sad, dobijene su predikcije koncentracije polena za ove tri biljke i u Vrbasu. Efikasnost metoda smo poredili na osnovu pomenutih 6 kriterijuma, čiji se rezultati nalaze u Tabeli 5.3 i Tabeli 5.4.

Posmatramo prvo Tabelu 5.3. Za predikcije koncentracije polena ambrozije izdvajaju se dva metoda, NOVI1 i NOVI2. Ukoliko odaberemo kao kriterijum za poređenje NRMSE i $NRMSE_S$ onda metod NOVI1 ima bolje rezultate, odnosno ima manju srednje kvadratnu grešku gledano na nivou cele godine ili samo za sezonu. Ukoliko posmatramo Spirmanovo ρ i ρ_S oba metoda imaju iste vrednosti, kao i za procenat podbacivanja u toku cele godine. Međutim, procenat podbacivanja je za nijansu manji u toku sezone ukoliko se koristi metod NOVI2. Ukoliko želimo odabrati metod za predviđanje koncentracije breze, ponovo je potrebno da se odlučimo koji kriterijum će nam biti najvažniji jer kao ni kod ambrozije nemamo jednoznačno određen najbolji metod. NOVI 2 ima najbolje rezultate ukoliko posmatramo NRMSE, $NRMSE_S$, Spirmanovo ρ i ρ_S , dok je procenat podbacivanja i procenat podbacivanja u sezoni najmanji za metod NOVI1. Za travu je situacija takođe specifična. Metodi NOVI1 i NOVI2 imaju skoro identične rezultate. Razlika je jedino kod procenta podbacivanja u toku sezone gdje manju vrednost ima metod NOVI2, tako da jedno na osnovu toga možemo reći da je za predviđanje koncentracije polena trave najbolje koristiti ovaj metod. Iako pri odabiru metoda trebamo uzimati u obzir kriterijum na osnovu koga poredimo metode, vidimo da za svaku od posmatranih biljaka NOVI metodi daju bolje rezultate od metoda STARI pa možemo da kažemo da su modifikovani metodi efikasniji za dobijanje predikcija.

	Ambrozija					
	NRMSE	PROC	ρ	$NRMSE_S$	$PROC_S$	ρ_S
STARI	2,53	0,18	0,88	5,04	0,43	0,86
NOVI1	2,33	0,16	0,93	4,64	0,33	0,92
NOVI2	2,34	0,16	0,93	4,66	0,32	0,92
	Breza					
	NRMSE	PROC	ρ	$NRMSE_S$	$PROC_S$	ρ_S
STARI	5,07	0,20	0,71	10,00	0,39	0,72
NOVI1	4,96	0,12	0,78	9,74	0,23	0,80
NOVI2	4,92	0,13	0,79	9,69	0,24	0,81
	Trava					
	NRMSE	PROC	ρ	$NRMSE_S$	$PROC_S$	ρ_S
STARI	1,40	0,32	0,90	1,83	0,48	0,88
NOVI1	1,36	0,29	0,94	1,76	0,42	0,93
NOVI2	1,36	0,29	0,94	1,76	0,41	0,93

Tabela 5.3 Poređenje metoda – tabela proseka za Vrbas

Metode možemo uporediti i na osnovu toga kolika je verovatnoća uspešnosti za svaki od njih da će predvideti dobre vrednosti. Ovi podaci se nalaze u Tabeli 5.4. Različiti kriterijumi za različite biljke preferiraju drugačije metode i zbog toga moramo biti pažljivi pri analizi rezultata i uzimati sve navedeno u obzir.

Posmatrajmo prvo rezultate za ambroziju posebno za svaki od kriterijuma. NRMSE, NRMSE_S i Spirmanovo ρ_S izdvajaju metod NOVI1 kao najbolji za predviđanje koncentracije polena ove biljke, dok preostala tri kriterijuma izdvajaju metod NOVI2. Ukoliko pak posmatramo prosečne vrednosti na osnovu svih kriterijuma, odnosno izvršimo kombinaciju kriterijuma (poslednje tri kolone tabele), uočavamo da metod NOVI1 ima najveću vrednost, tj. najveću verovatnoću da će uspešno predvideti tražene vrednosti. Za brezu imamo drugačije rezultate. Kriterijumi PROC i PROC_S izdvajaju metod NOVI1 kao najbolji, dok preostali kriterijumi izdvajaju metod NOVI2. Pošto je za veći broj kriterijuma metod NOVI2 imao veće vrednosti, kada iskombinujemo svih 6 upravo je NOVI2 taj koji ima najveću verovatnoću uspešnosti. Metod za predviđanje koncentracije polena trave je najjednostavnije odabrati. Na osnovu svih kriterijuma metod NOVI2 ima najbolje rezultate.

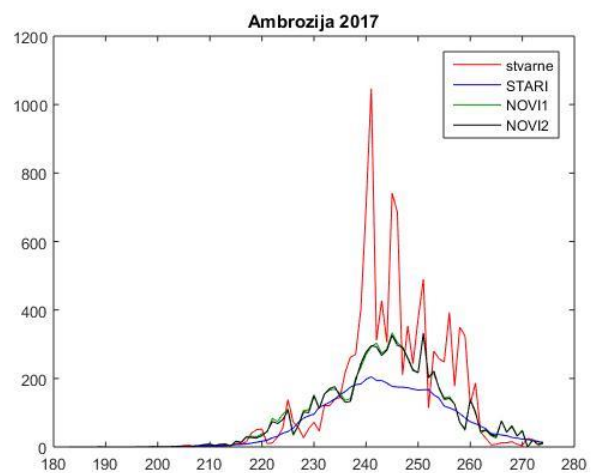
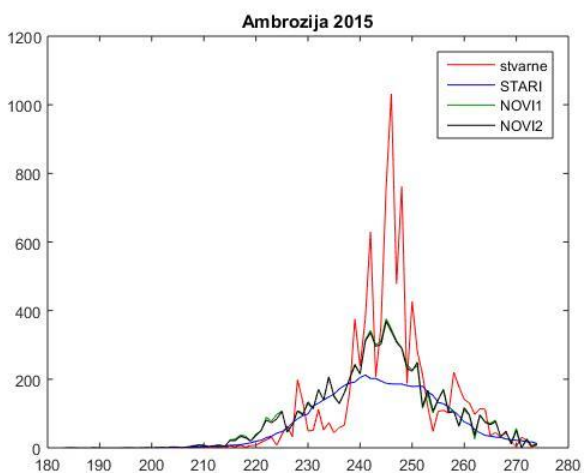
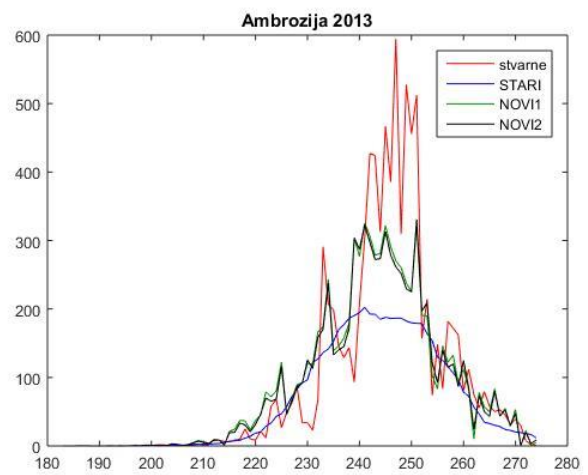
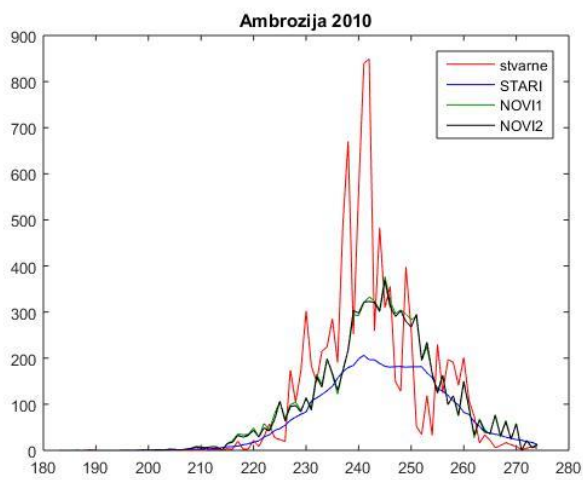
Ukoliko posmatramo rezultate tako da želimo samo jedan od metoda da koristimo za predviđanje koncentracije polena za sve tri biljke (kolona ABT) ponovo imamo situaciju kao za travu. Na osnovu svih kriterijuma metod NOVI2 ima najbolje vrednosti. Dakle, kada izvršimo kombinaciju svih metoda za sve tri biljke, konačan rezultat nam ukazuje da je najbolje koristiti metod NOVI2 (poslednje tri vrste kolone ABT). Ali kao što je već napomenuto, metod je bolje birati za svaku biljku pojedinačno.

		ABT	A	B	T
NRMSE	STARI	0,15	0,11	0,22	0,11
	NOVI1	0,41	0,78	0,11	0,33
	NOVI2	0,44	0,11	0,67	0,56
PROC	STARI	0,00	0,00	0,00	0,00
	NOVI1	0,59	0,44	0,78	0,56
	NOVI2	0,67	0,78	0,44	0,78
ρ	STARI	0,00	0,00	0,00	0,00
	NOVI1	0,33	0,44	0,11	0,44
	NOVI2	0,67	0,56	0,89	0,56
NRMSE _S	STARI	0,15	0,11	0,22	0,11
	NOVI1	0,41	0,78	0,11	0,33
	NOVI2	0,44	0,11	0,67	0,56
PROC _S	STARI	0,00	0,00	0,00	0,00
	NOVI1	0,56	0,56	0,67	0,44
	NOVI2	0,67	0,89	0,33	0,78
ρ_S	STARI	0,00	0,00	0,00	0,00
	NOVI1	0,41	0,56	0,22	0,44
	NOVI2	0,59	0,44	0,78	0,56
		ABT	A	B	T
STARI		0,05	0,04	0,07	0,04
NOVI1		0,45	0,59	0,33	0,43
NOVI2		0,58	0,48	0,63	0,63

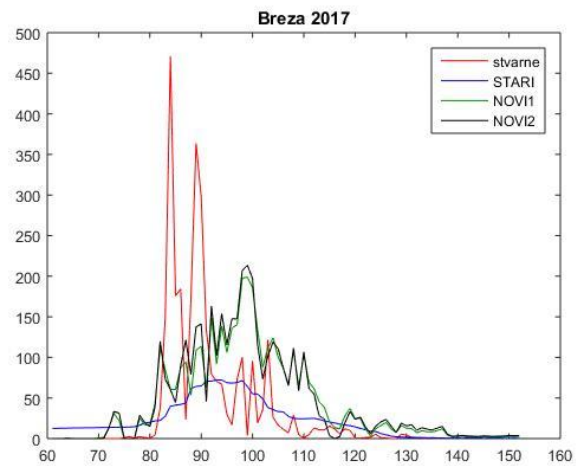
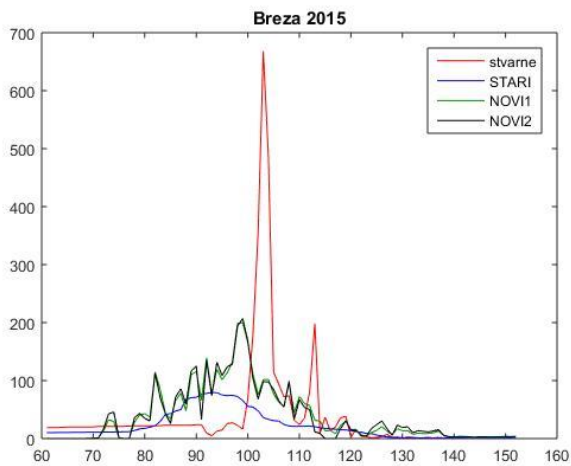
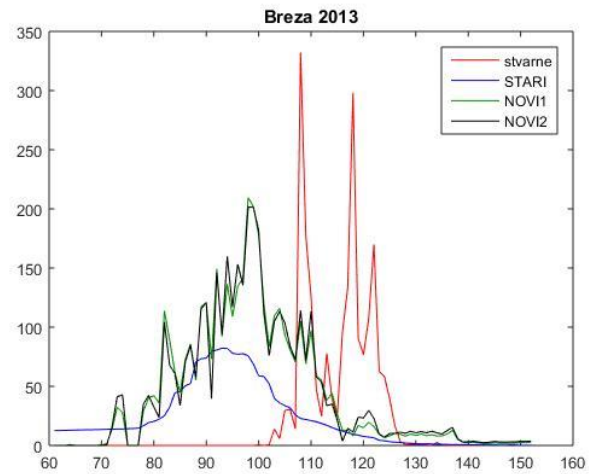
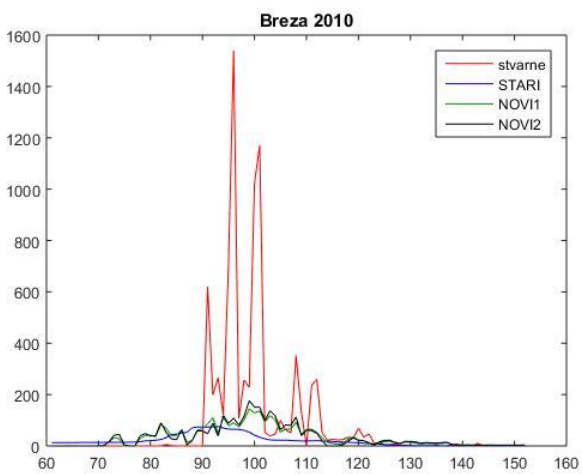
Tabela 5.4 Verovatnoće uspešnosti metoda

U nastavku ćemo prikazati grafike na kojima se vizuelno može oceniti i primetiti koliko dobro predikcije pomoću različitih metoda prate kretanje stvarnih vrednosti. Kao i za Novi Sad na slučajan način smo izabrali 4 godine i samo za njih ćemo prikazati grafike. To su 2010., 2013., 2015. i 2017. godina.

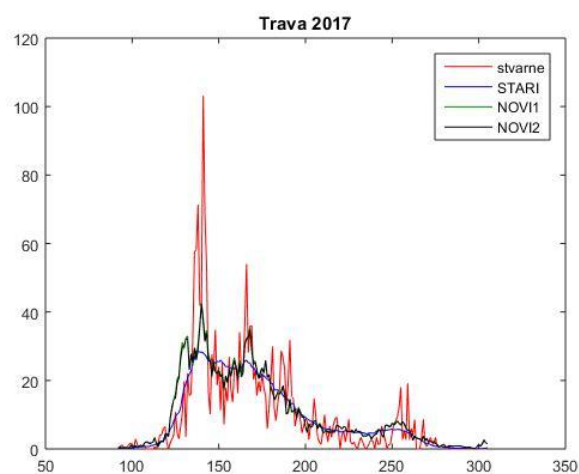
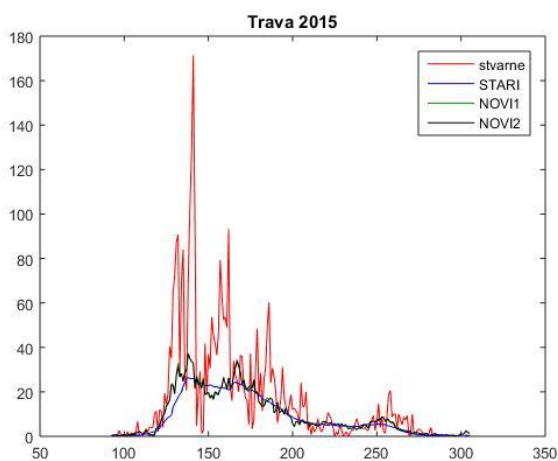
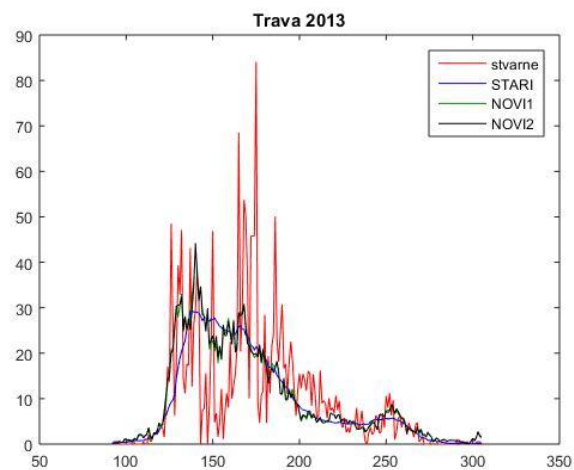
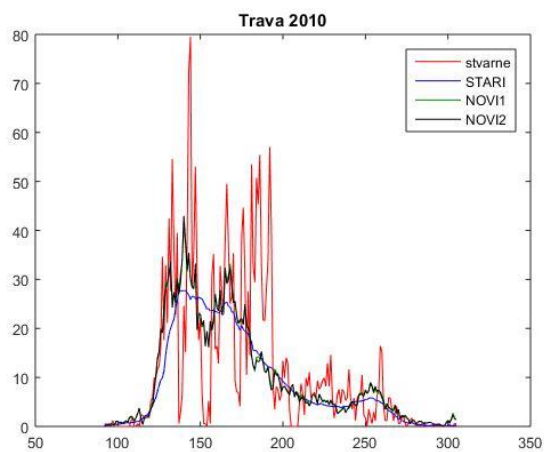
Prvo ćemo vidjeti kako grafici izgledaju za ambroziju. Stvarne vrednosti, njihovi iznenadni skokovi kao i periodi sa maksimalnom i minimalnom koncentracijom u toku sezone variraju u zavisnosti od toga koju godinu posmatramo. Maksimalna prosečna dnevna vrednost koncentracije polena ambrozije na teritoriji Vrbasa zabeležena je 28.08.2011. godine i iznosila je $1234/m^3$ polenovih zrna.



Sledeća 4 grafika prikazuju stvarne i predviđene koncentracije polena breze. Maksimalna prosečna dnevna vrednost koncentracije polena za ovu biljku izmerena je 30.03.2012. godine i iznosila je $1894/m^3$ polenovih zrna. Možemo da primetimo da je 2013. godine sezona breze, odnosno period kada se njen polen nalazi u atmosferi, počela kasnije nego ostalih godina zbog drugih atmosferskih faktora. Kako metodi koriste ostale godine za predviđanje koncentracije polena primećujemo da predikcije ne prate dobro stvarne vrednosti. Pravljenje modela je upravo zbog toga jako kompleksno jer veliki broj faktora u atmosferi utiče na početak sezone kao i na dalje kretanje i količinu polena u vazduhu.



Stvarne i predviđene koncentracije polena trave za odabrane 4 godine predstavljene su sledećim graficima. Za nju je maksimalna prosečna dnevna vrednost koncentracije polena izmerena 20.05.2015. i iznosila je $171/m^3$.



5.3. Sombor

Podaci o koncentraciji polena ambrozije, breze i trave u Somboru beleženi su od 2008. do 2017. godine. Ukupan broj podataka iznosi po 43836 za svaku biljku posebno, jer je koncentracija kao i kod ostalih gradova beležena na svaka 2 sata. Tabela 5.5. nam daje pregled efikasnosti svakog od metoda ukoliko ih uporedimo na osnovu pomenutih 6 kriterijuma. Podaci od verovatnoći uspešnosti metoda na osnovu istih kriterijuma se nalaze u Tabeli 5.6.

Posmatramo prvi način poređenja i Tabelu 5.5. Za predikcije koncentracije polena ambrozije po kriterijumima NRMSE i NRMSE_S kao najbolji metod se izdvaja NOVI2. Ukoliko nam je pak više važnije koliki je procenat podbacivanja predikcija kako za celu godinu, tako i za sezonu, onda je metod NOVI1 taj koji treba izabrati jer on ima najmanju vrednost po ovom kriterijumu. Ukoliko posmatramo Spirmanovo ρ i ρ_S vidimo da pomenuta 2 metoda imaju iste vrednosti tako da odabir vršimo na osnovu nekog od prethodna 2 kriterijuma. Za predikcije koncentracije polena breze metodi NOVI1 i NOVI2 su bolji od metoda STARI, imaju iste vrednosti po svim kriterijumima osim za NRMSE i NRMSE_S gde metod NOVI1 ima manje vrednosti pa se on može izdvojiti kao najbolji od sva tri metoda za ovu biljku. Kod trave je slična situacija kao i za brezu. Modifikovani metodi su bolji po svim kriterijumima od STAROG metoda, imaju iste vrednosti po svim kriterijumima osim za NRMSE i NRMSE_S gde metod NOVI2 ima bolje rezultate. Na osnovu toga možemo reći da je metod NOVI2 najbolje koristiti za računanje dugoročnih predikcija koncentracije polena trave. Kao i za prethodne gradove, uočavamo da su rezultati bolji za modifikovane metode, ali svakako trebamo vršiti detaljniju analizu pri odabiru metoda jer on svakako zavisi od biljne vrste za koju vršimo predikcije.

	Ambrozija					
	NRMSE	PROC	ρ	NRMSE _S	PROC _S	ρ_S
STARI	1,89	0,22	0,95	3,78	0,44	0,94
NOVI1	1,83	0,20	0,97	3,65	0,38	0,97
NOVI2	1,82	0,21	0,97	3,62	0,39	0,97
	Breza					
	NRMSE	PROC	ρ	NRMSE _S	PROC _S	ρ_S
STARI	9,96	0,12	0,57	13,02	0,32	0,69
NOVI1	6,14	0,09	0,81	11,91	0,28	0,77
NOVI2	6,16	0,09	0,81	11,95	0,28	0,77
	Trava					
	NRMSE	PROC	ρ	NRMSE _S	PROC _S	ρ_S
STARI	1,63	0,27	0,93	2,13	0,45	0,88
NOVI1	1,60	0,24	0,94	2,09	0,41	0,93
NOVI2	1,59	0,24	0,94	2,08	0,41	0,93

Tabela 5.5 Poređenje metoda – tabela proseka za Sombor

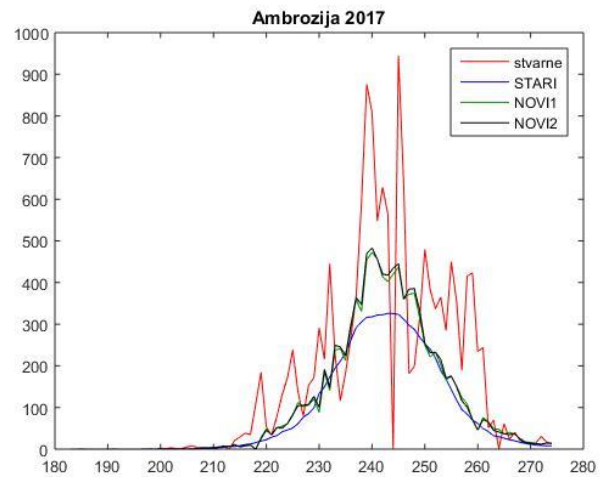
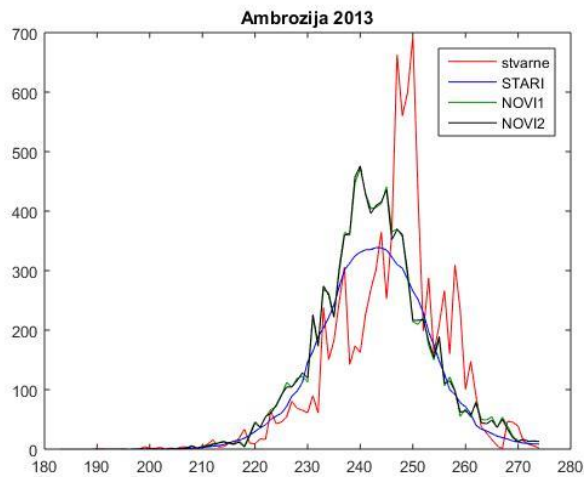
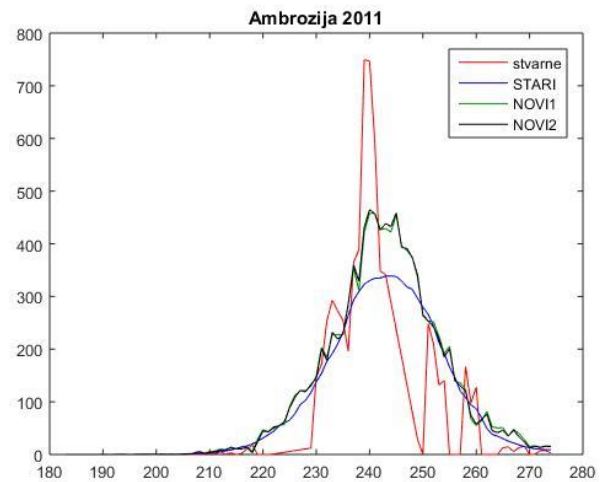
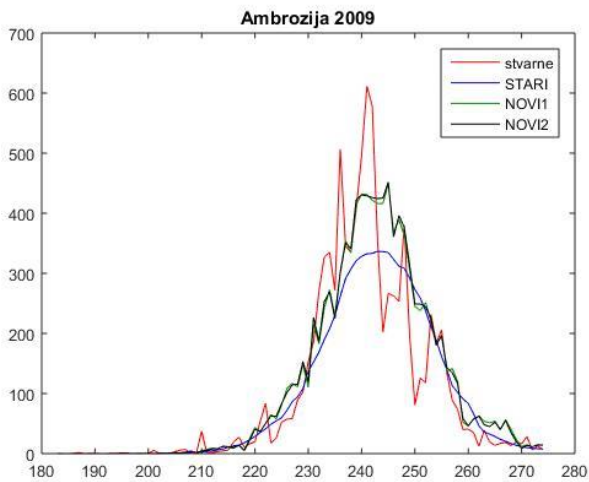
Drugi način poređenja metoda je pomoću verovatnoća uspešnosti datih u Tabeli 5.6. Za predviđanje koncentracije polena ambrozije na osnovu kriterijuma NRMSE i PROC_S najbolje je koristiti metod NOVI2, dok je na osnovu preostala 4 kriterijuma to metod NOVI1. Kada izvršimo kombinaciju kriterijuma (poslednje 3 vrste tabele) dobijamo očekivan rezultat, odnosno metod NOVI1 ima najveću verovatnocu uspešnosti. Za brezu metod NOVI1 ima najveću verovatnoću uspešnosti na osnovu kriterijuma NRMSE i PROC_S. Na osnovu kriterijuma Spirmanovo ρ i ρ_S najuspešniji je metod NOVI2, dok za preostala dva kriterijuma PROC i NRMSE_S imamo iste vrednosti za po dva metoda NOVI1, NOVI2 i STARI, NOVI1, respektivno. Dakle, u zavisnosti od toga kojem kriterijumu pridajemo više značaja odabir metoda može da bude drugačiji. Ukoliko izvršimo kombinaciju kriterijuma, rezultat je kao za prethodnu biljku, tj. metod NOVI1 ima najveću verovatnoću uspešnosti. Za treću biljku, travu, imamo najjasnije rezultate. Po svim kriterijumima trebalo bi izabrati metod NOVI2, pa i kombinacija kriterijuma izdvaja isti kao metod sa najvećom verovatnoćom uspešnosti da će dati bolje predikcije nego preostala dva metoda. U koloni ABT kombinovani su rezultati za sve tri biljke pa se može analizirati koji metod je najbolji ukoliko želimo odabrati samo jedan za predviđanje koncentracija polena svih biljaka. Na osnovu kriterijuma NRMSE, NRMSE_S, PROC i PROC_S metod NOVI1 ima najveće vrednosti, odnosno najveću verovatnoću uspešnosti. Preostala dva kriterijuma Spirmanovo ρ i ρ_S izdvajaju metod NOVI2. Kada pogledamo krajnji rezultat, ukoliko izvršimo kombinaciju svih biljaka i svih kriterijuma metod NOVI2 ima najveću verovatnoću uspešnosti.

		ABT	A	B	T
NRMSE	STARI	0,21	0,30	0,13	0,20
	NOVI1	0,43	0,20	0,75	0,40
	NOVI2	0,36	0,50	0,13	0,40
PROC	STARI	0,00	0,00	0,00	0,00
	NOVI1	0,71	0,90	0,63	0,60
	NOVI2	0,50	0,20	0,63	0,70
ρ	STARI	0,04	0,00	0,13	0,00
	NOVI1	0,36	0,60	0,25	0,20
	NOVI2	0,61	0,40	0,63	0,80
NRMSE _S	STARI	0,32	0,30	0,50	0,20
	NOVI1	0,36	0,20	0,50	0,40
	NOVI2	0,32	0,50	0,00	0,40
PROC _S	STARI	0,11	0,00	0,38	0,00
	NOVI1	0,61	0,70	0,50	0,60
	NOVI2	0,43	0,30	0,25	0,70
ρ_S	STARI	0,04	0,00	0,13	0,00
	NOVI1	0,29	0,60	0,25	0,00
	NOVI2	0,68	0,40	0,63	1,00
		ABT	A	B	T
STARI		0,12	0,10	0,21	0,07
NOVI1		0,46	0,53	0,48	0,37
NOVI2		0,48	0,38	0,38	0,67

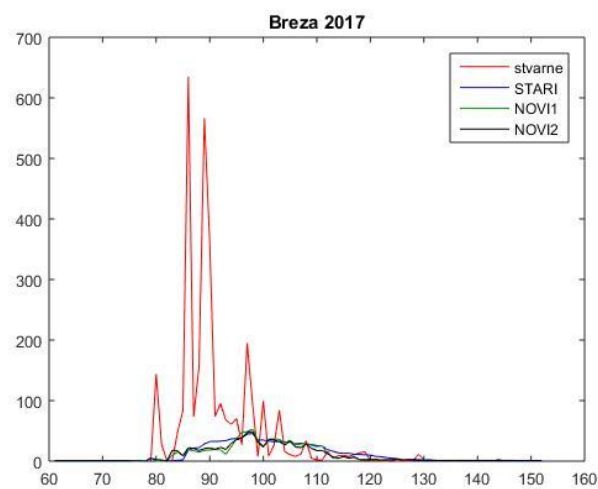
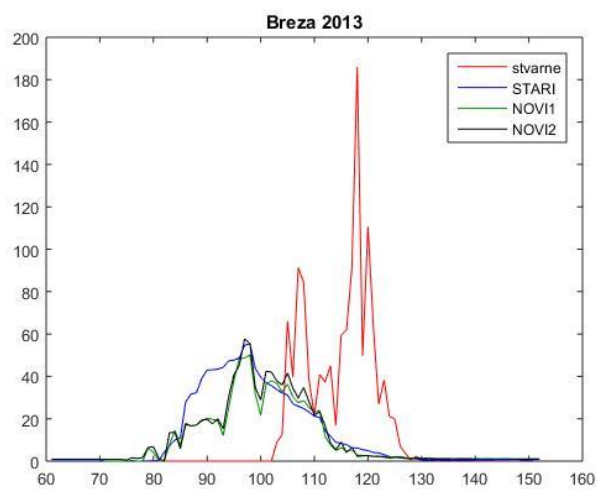
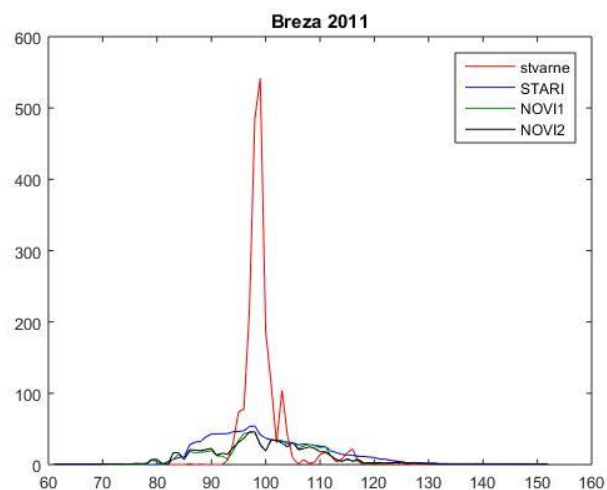
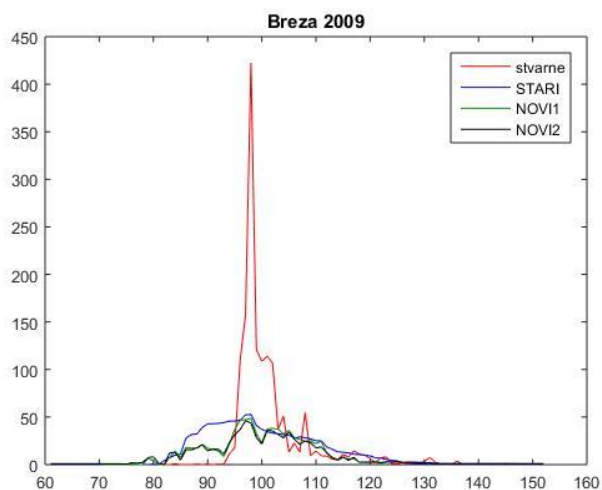
Tabela 5.6 Verovatnoće uspešnosti metoda

Na graficima koji se nalaze u nastavku rada možemo da uporedimo kretanje stvarnih vrednosti koncentracije polena sa onima koje smo predvideli pomoću metoda STARI, NOVI1 i NOVI2. Na slučajan način su izabrane 4 godine i samo za njih ćemo prikazati grafike za sve tri biljke. To su 2009.,2011.,2013. i 2017. godina.

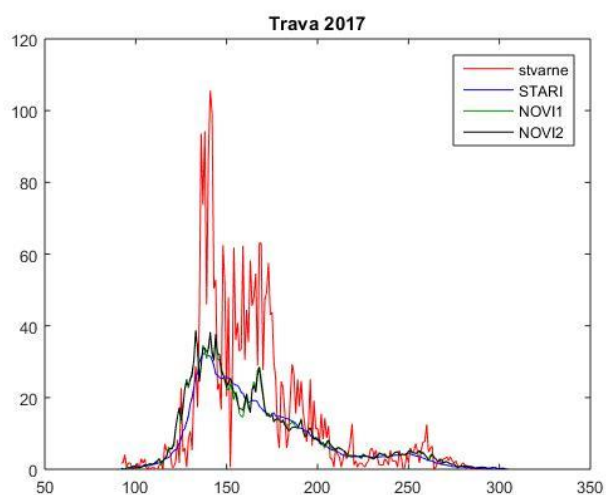
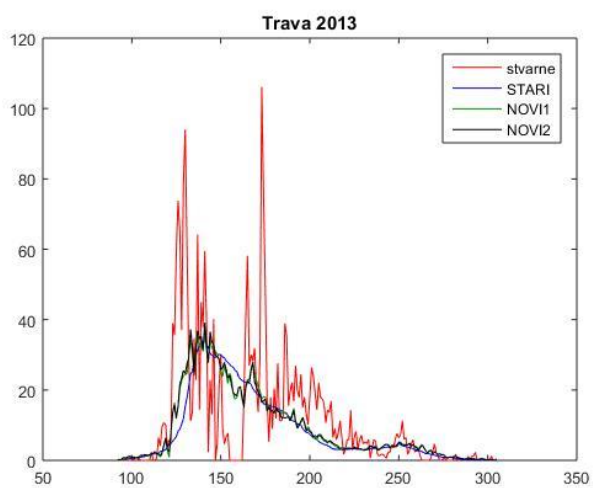
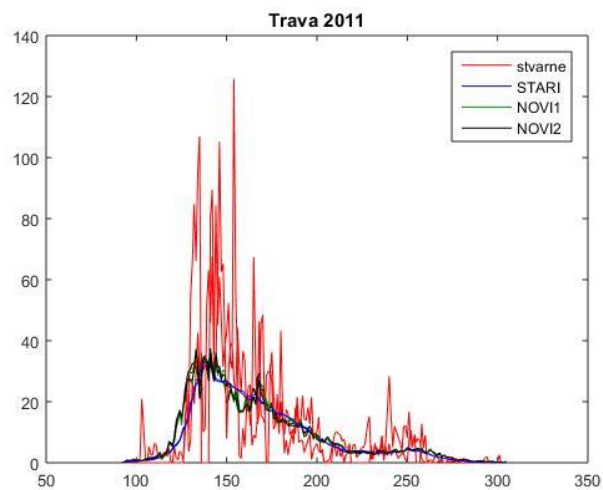
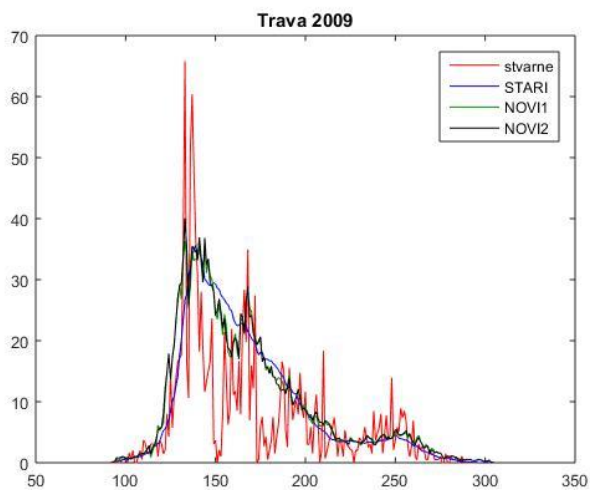
Kao i do sada prvo prikazujemo grafike za ambroziju. Oni se naravno razlikuju u zavisnosti od posmatrane godine. Maksimalna prosečna dnevna vrednost koncentracije polena ove biljke na teritoriji Sombora zabeležena je 31.08.2016. godine i iznosila je $1027/m^3$ polenovih zrna.



Stvarne i predviđene koncentracije polena breze možemo da uporedimo na sledećim graficima. Maksimalna prosečna dnevna vrednost koncentracije polena za ovu biljku izmerena je 04.04.2012. godine i iznosila je $1451/m^3$ polenovih zrna.



Naredna 4 grafika su vezana za treću biljku, travu. Za nju je maksimalna prosečna dnevna vrednost koncentracije polena izmerena 02.06.2008. i iznosila je $126/m^3$.



5.4. Zrenjanin

Za razliku od ostalih gradova, za Zrenjanin umesto za tri imamo podatke samo za dve biljne vrste. Koncentracija polena merena je samo za ambroziju i travu. Period u kome su beležene dvočasovne dnevne vrednosti je od 2008. do 2017. godine. Dakle, za svaku biljku imamo po 43836 podataka. Kao i za prethodne gradove metode za dobijanje dugoročnih predikcija koncentracije polena poredili smo na osnovu istih 6 kriterijuma, a rezultati su prikazani u Tabeli 5.7 i Tabeli 5.8.

Detaljnija analiza pri odabiru metoda je potrebna i u slučaju ovog grada. Odabir kriterijuma za poređenje metoda je prvi korak za koji se moramo opredeliti. Posmatramo najpre rezultate za ambroziju. Ukoliko poređenje vršimo na osnovu procenata podbacivanja metod NOVI2 ima najbolje rezultate. Ukoliko nam je pak važniji rezultat za Spirmanovo ρ i ρ_S , metodi NOVI1 i NOVI2 imaju isti rezultat ali bolji od metoda STARI. Istu situaciju imamo kod kriterijuma NRMSE, dok najmanju vrednost NRMSE_S ima metod NOVI2. Rezultati za travu su malo drugačiji. Na osnovu kriterijuma Spirmanovo ρ i ρ_S ne možemo se opredeliti ni za jedan metod jer sva tri imaju isti rezultat. Takođe istu vrednost NRMSE_S, kao i isti procenat podbacivanja u togu godine i samo u toku sezone imaju metodi NOVI1 i NOVI2, ali bolji od metoda STARI. Samo za kriterijum NRMSE imamo poretak na osnovu rezultata i kao najbolji metod se izdvaja NOVI1.

	Ambrozija					
	NRMSE	PROC	ρ	NRMSE _S	PROC _S	ρ_S
STARI	2,39	0,17	0,90	4,77	0,42	0,87
NOVI1	2,26	0,13	0,93	4,51	0,29	0,92
NOVI2	2,26	0,12	0,93	4,50	0,28	0,92
	Trava					
	NRMSE	PROC	ρ	NRMSE _S	PROC _S	ρ_S
STARI	1,49	0,27	0,93	1,95	0,44	0,92
NOVI1	1,45	0,23	0,93	1,90	0,38	0,92
NOVI2	1,46	0,23	0,93	1,90	0,38	0,92

Tabela 5.7 Poređenje metoda – tabela proseka za Zrenjanin

Odabir metoda možemo vršiti i na osnovu verovatnoća uspešnosti. Posmatrajući rezultate za ambroziju vidimo da za različite kriterijume imamo različite rezultate. Tako npr. za kriterijume NRMSE i $NRMSE_S$ imamo iste vrednosti, i oni izdvajaju NOVI1 i NOVI2 kao bolje metode od metoda STARI. Zatim za PROC i $PROC_S$ najveću verovatnoću uspešnosti ima metod NOVI2. Spirmanovo ρ izdvaja metod NOVI1, a Spirmanovo ρ_S dva metoda, NOVI1 i NOVI2. Kombinacija svih 6 kriterijuma ukazuje da je najbolje koristiti metod NOVI2, ukoliko se odlučimo za izbor metoda na ovaj način. Kod trave imamo sličnu situaciju kao u prethodnoj tabeli, tj. na osnovu kriterijuma Spirmanovo ρ i ρ_S ne možemo doneti odluku o izboru metoda jer sva tri imaju iste vrednosti. Posmatrajući NRMSE i $NRMSE_S$ izbor bi bio metod NOVI2, dok PROC i $PROC_S$ izdvajaju metod NOVI1. Kada pogledamo kombinaciju svih kriterijuma najveću verovatnoću uspešnosti ima metod NOVI2.

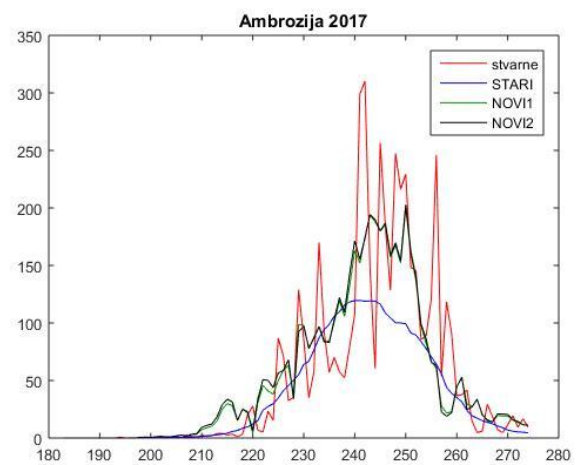
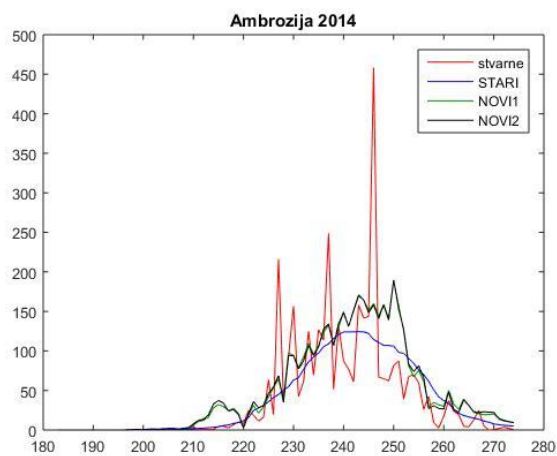
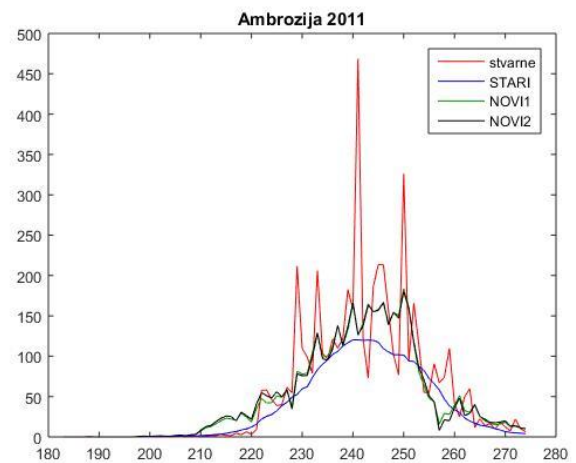
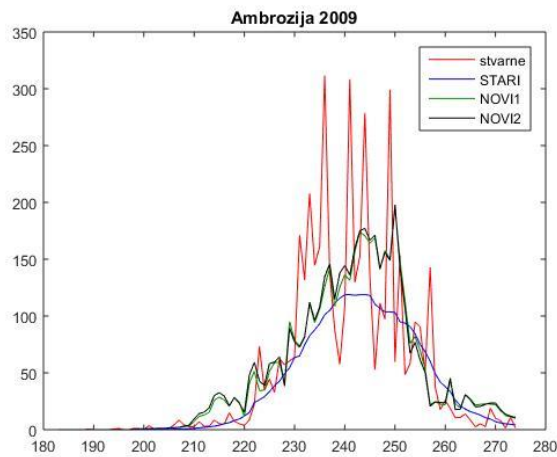
Ukoliko bi želeli koristiti samo jedan metod za predviđanje koncentracija polena obe biljke potrebno je posmatrati rezultate u kolini AT. Kriterijumi NRMSE, $NRMSE_S$, PROC i $PROC_S$ izdvajaju metod NOVI2, dok Spirmanovo ρ izdvaja metod NOVI1. Za Spirmanovo ρ_S metodi NOVI1 i NOVI2 imaju istu vrednost, ali veću nego metod STARI. Kombinacija rezultata svih kriterijuma za obe biljke ide u korist metodu NOVI2. Za njega imamo najveću krajnju verovatnoću uspešnosti. Međutim, kao i za prethodne gradove potrebno je napomenuti da je odabir metoda potrebno vršiti u zavisnosti od toga za koju biljku zelimo predvideti koncentracije polena i kojem kriterijumu pridajemo najviše značaja.

		AT	A	T
NRMSE	STARI	0,21	0,20	0,22
	NOVI1	0,32	0,40	0,22
	NOVI2	0,47	0,40	0,56
PROC	STARI	0,00	0,00	0,00
	NOVI1	0,63	0,60	0,67
	NOVI2	0,68	0,80	0,56
ρ	STARI	0,16	0,00	0,33
	NOVI1	0,47	0,60	0,33
	NOVI2	0,37	0,40	0,33
$NRMSE_S$	STARI	0,21	0,20	0,22
	NOVI1	0,32	0,40	0,22
	NOVI2	0,47	0,40	0,56
$PROC_S$	STARI	0,00	0,00	0,00
	NOVI1	0,63	0,50	0,78
	NOVI2	0,68	0,80	0,56
ρ_S	STARI	0,16	0,00	0,33
	NOVI1	0,42	0,50	0,33
	NOVI2	0,42	0,50	0,33
		AT	A	T
STARI		0,12	0,07	0,19
NOVI1		0,46	0,50	0,43
NOVI2		0,52	0,55	0,48

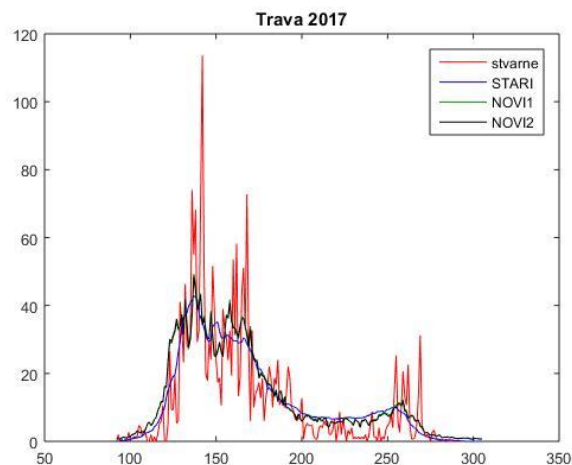
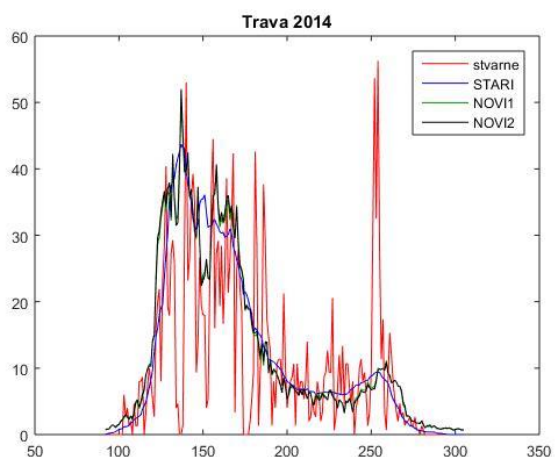
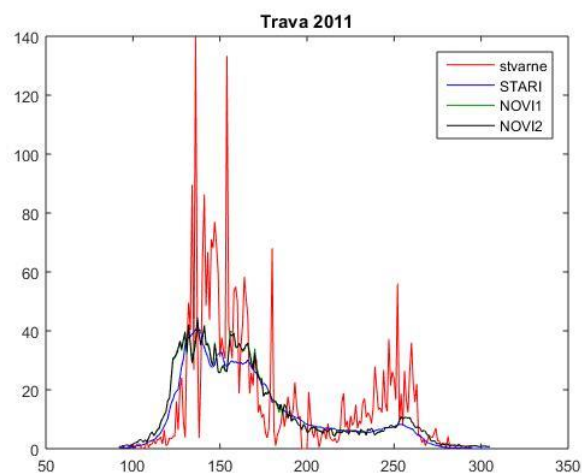
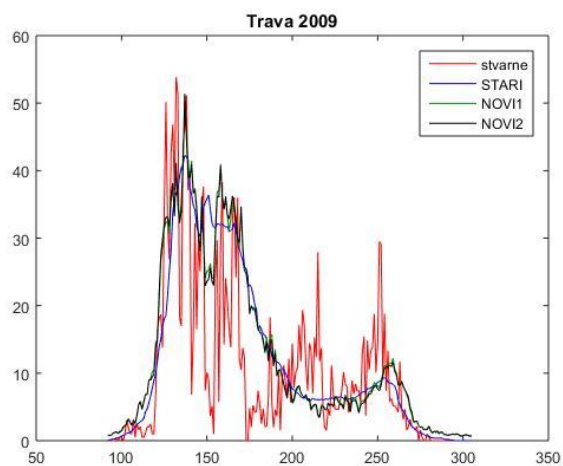
Tabela 5.8 Verovatnoće uspešnosti metoda

Sledi grafički prikaz stvarnih vrednosti i vrednosti predviđenih pomoću metoda. Kao i za prethodne gradove, radi preglednosti rada, na slučajan način smo izabrali 4 godine i samo za njih ćemo prikazati grafike. To su 2009., 2011., 2014. i 2017. godina.

Prvo ćemo vidjeti kako grafici izgledaju za ambroziju. Maksimalna prosečna dnevna vrednost koncentracije polena ambrozije na teritoriji grada Zrenjanina zabeležena je 30.08.2016. godine i iznosila je $745/m^3$ polenovih zrna.



Stvarne i predviđene koncentracije polena trave za odabrane 4 godine predstavljene su sledećim graficima. Za nju je maksimalna prosečna dnevna vrednost koncentracije polena izmerena 24.06.2016. i iznosila je $364/m^3$.



Zaključak

U ovom radu predstavljena su tri metoda za formiranje dugoročnih predikcija koncentracije polena ambrozije, breze i trave. Relevantni metod je preuzet iz rada [3]. Zatim su izvršene modifikacije datog metoda i dobijena su dva nova. Glavno ispitivanje metoda i analiza rezultata sprovedena je na osnovu istorijskih podataka o koncentraciji polena za posmatrane biljke na teritoriji Novog Sada. Poređenje metoda vršeno je pomoću 3 kriterijuma koja su primenjena na predikcije za celu godinu, a zatim na deo godine koji predstavlja sezonu za odgovarajuću biljku. Zatim su pomenuti metodi primenjeni na istorijske podatke o koncentraciji polena izmerene na teritoriji Vršca, Sombora i Zrenjanina, a poređenje je vršeno na isti način kao i za Novi Sad.

Rezultati koji su dobijeni mogu se tumačiti na različite načine. U zavisnosti od grada i biljne vrste, kao i od kriterijuma kojem pridajemo najviše značaja ili kombinacije istih možemo izabrati različite metode za formiranje dugoročne prognoze koncentracije polena. Međutim, od tri metoda koja su predstavljena u radu, odabir se svodi na jedan od dva nova metoda. Dakle, modifikacije su donele poboljšanje rezultata odnosno omogućile formiranje preciznijih predikcija od relevantnog metoda.

Dugoročna predikcija koncentracije polena nije jednostavna jer zavisi od velikog broja faktora i pojava koje se dešavaju u prirodi. Neke od njih možemo pratiti i predvideti, a neke ne. Zbog toga ovaj problem predstavlja izazov i ostavlja prostor za dodatnu modifikaciju metoda. Takođe, dugoročna prognoza se može koristiti i kao ulazni niz podataka za dobijanje kratkoročnih prognoza za koje očekujemo da će biti mnogo preciznije. To predstavlja jednu od ideja za buduća istraživanja.

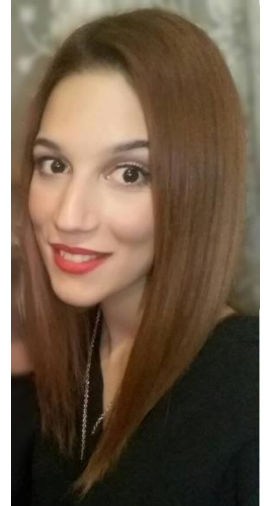
Osobama koje su alergične na polen i dugoročna i kratkoročna prognoza mogu biti od velikog značaja. Planiranje putovanja, godišnjih odmora i odgovarajuće terapije postaje olakšano i to je jedan od načina da se unapredi kvalitet života sa alergijom. Kako se procenjuje da danas u svetu svaka treća osoba ima alergijske reakcije na polen i da je broj alergičnih svake godine u porastu, potreba za ovakvom vrstom prognoze je sve veća.

Literatura

- [1] Brockwell, Peter, J., Davis, Richard, A.,(2002). Introduction To Time Series and Forecasting: Springer Science+Business Media New York
- [2] Chris, Chatfield, (2003). The Analysis of Time Series: An Introduction, Sixth Edition: Chapman and Hall/CRC
- [3] Šikoparija, B., Marko, O., Panić, M., Jakovetić, D., Radišić, P., (2018). How to prepare a pollen calendar for forecasting daily pollen concentrations of Ambrosia, Betula and Poaceae?: Springer Science + Business Media B.V., part of Springer Nature
- [4] Tsay, R. S. (2005). Analysis of Financial Time Series. 2. izd. Njujork: A John Wiley & Sons, Inc.
- [5] J. Nocedal, S. J. Wright, (1999). Numerical Optimization, Springer
- [6] Österreichische Lungenunion, Alergija brošura
- [7] Zlatko J. Kovačić, (1995). Analiza vremenskih serija, Beograd
- [8] www.nspolen.com
- [9] www.alergijanapolen.rs
- [10]www.polleninfo.org

Biografija

Jelena Samardžija je rođena 29.01.1994. godine u Bosanskom Petrovcu, Bosna i Hercegovina. Osnovnu školu „Nikola Tesla” u Derventi završava 2009. godine kao nosilac Vukove diplome. Potom, takođe u Derventi, upisuje opšti smer gimnazije, koju završava 2013. Studije Prirodno-matematičkog fakulteta u Novom Sadu, smer Primenjena matematika, modul matematika finansija, upisuje iste godine, i uspešno ih završava septembra 2016. Iste godine upisuje master akademske studije smer Primenjena matematika. Zaključno sa junskim rokom 2018. godine položila je sve ispite i stekla uslov za odbranu master rada.



**UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI
FAKULTET
KLJUČNA DOKUMENTACIJSKA
INFORMACIJA**

Redni broj:

RBR

Identifikacioni broj:

IBR

Tip dokumentacije: *monografska dokumentacija*

TD

Tip zapisa: *tekstualni štampani materijal*

TZ

Vrsta rada: *master rad*

VR

Autor: *Jelena Samardžija*

AU

Mentor: *dr Nataša Krklec Jerinkić*

MN

Naslov rada: *Optimizacioni metodi za dugoročne predikcije koncentracije polena glavnih alergena u Vojvodini*

NR

Jezik publikacije: *srpski (latinica)*

JP

Jezik izvoda: *s/e*

JI

Zemlja publikovanja: *Republika Srbija*

ZP

Uže geografsko područje: *Vojvodina*

UGP

Godina: 2019.

GO

Izdavač: *autorski reprint*

IZ

Mesto i adresa: *Novi Sad, Trg Dositeja Obradovića 4*

MA

Fizički opis rada: *5 poglavlja, 61 stranica, 21 slika, 9 tabela, 50 grafika*

FO

Naučna oblast: *matematika*

NO

Naučna disciplina: *optimizacioni metodi*

ND

Ključne reči: *polen, alergija, ambrozija, breza, trava, oprtimizacioni metodi, vremenske serije*

PO

UDK

Čuva se: *u biblioteci Departmana za matematiku i informatiku, Prirodno-matematičkog fakulteta, u Novom Sadu*

ČU

Važna napomena:

VN

Izvod: *U master radu su pedstavljeni optimizacioni metodi za dugoročno predviđanje koncentracije polena ambrozije, breze i trave na teritoriji sledeća četiri grada u Vojvodini: Novi Sad, Vrbas, Sombor i Zrenjanin. Za formiranje predikcija korišteni su podaci o dvočasovnim dnevnim koncentracijama polena ambrozije, breze i trave u Novom Sadu za period od 2000. do 2017. godine, u Vrbasu za period od 2009. do 2017., zatim u Somboru za period od 2008. do 2017. i u Zrenjanu takođe za period od 2008. do 2017. godine, ali samo za ambroziju i travu. Za dobijanje predikcija korištena su tri optimizaciona metoda čiji rezultati su poređeni na osnovu različitih kriterijuma u cilju odabira najboljeg metoda.*

IZ

Datum prihvatanja teme od strane NN veća:

DP

Datum odbrane:

DO

Članovi komisije:

KO

Predsednik: *dr Nataša Krejić, redovni profesor*

Član: *dr Nataša Krklec Jerinkić, docent*

Član: *dr Dušan Jakovetić, docent*

**UNIVERSITY OF NOVI SAD
FACULTY OF SCIENCES
KEY WORD DOCUMENTATION**

Accession number:

ANO

Identification number:

INO

Document type: *monograph type*

DT

Type of record: *printed text*

TR

Contents code: *Master thesis*

CC

Author: *Jelena Samardžija*

AU

Mentor: *Nataša Krklec Jerinkić, PhD*

MN

Title: *Optimization methods for long-term prediction of pollen concentration of major allergens in Vojvodina*

XI

Language of text: *Serbian (latin)*

LT

Language of abstract: *s/e*

LA

Country of publication: *Republic of Serbia*

CP

Locality of publication: *Vojvodina*

LP

Publication year: *2019*

PY

Publisher: *author's reprint*

PU

Publ. place: *Novi Sad, Trg Dositeja Obradovića 4*

PP

Physical description: *5 chapters, 61 pages, 21 pictures, 9 tables, 50 graphics*

PD

Scientific field: *mathematics*

SF

Scientific discipline: *optimization methods*

SD

Key words: *pollen, allergy, ambrosia, betula, poaceae, optimization methods, time series*

UC

Holding data: *Department of Mathematics and Informatics' Library, Faculty of Sciences, Novi Sad*

HD

Note:

N

Abstract: In this master thesis are presented the optimization methods for long-term predictions of the concentration of pollen of ambrosia, birch and grass on the territory of the following four cities in Vojvodina: Novi Sad, Vrbas, Sombor and Zrenjanin. For the formation of the prediction were used the data on two-hour daily concentrations of ambrosia, betula and poaceae fields in Novi Sad for the period from 2000 to 2017, in Vrbas for the period from 2009 to 2017, then in Sombor for the period from 2008 to 2017 and in Zrenjanin also for the period from 2008 to 2017, but only for ambrosia and grass. Three optimization methods were used to obtain the predictions, the results of which were compared on the basis of different criteria in order to select the best method.

AB

Accepted by the Scientific Board on:

ASB

Defended:

DE

Thesis defended board:

DB

President: *Nataša Krejić, PhD, full professor*

Member: *Nataša Krklec Jerinkić, PhD, assistant professor*

Member: *Dušan Jakovetić, PhD, assistant professor*