

Sveučilište u Zagrebu  
Prirodoslovno-matematički fakultet  
Geofizički odsjek

**Valentino Neduhal**

**Primjena analize empirijskih ortogonalnih  
funkcija na određivanje nestacionarnosti  
vremenskih nizova**

**Diplomski rad**

Zagreb, 2020.

Sveučilište u Zagrebu  
Prirodoslovno-matematički fakultet  
Geofizički odsjek

Diplomski sveučilišni studij Fizika – Geofizika  
Smjer: Meteorologija i fizička oceanografija

**Valentino Neduhal**

Diplomski rad

**Primjena analize empirijskih ortogonalnih  
funkcija na određivanje nestacionarnosti  
vremenskih nizova**

Voditelj diplomskog rada: doc. dr. sc. Željko Večenaj

Zagreb, 2020.

Zahvaljujem se doc. dr. sc. Željku Večenaju na ideji, te mentorstvu i izrazitom strpljenju tijekom izrade ovog rada. Zahvalio bi se roditeljima, braći, baki, djedu te ostaloj obitelji na srdačnoj pomoći i podršci koju su mi pružali tijekom godina provedenih na studiju i ranije! Zahvalio bih se djevojci, što je vjerovala u mene i u trenucima kada sam ja sumnjao. Vi ste me nosili kada nisam mogao hodati, te je ovaj rad, podjednako vaš kao što je i moj.

# Sažetak

Statistička stacionarnost je vrlo važan koncept u statističkoj teoriji turbulencije. Parametrizacije površinskog sloja, koje povezuju turbulentne tokove s gradijentima srednje vrijednosti preko Monin-Obukovljeve teorije sličnosti, pretpostavljaju stacionarnost turbulentnog procesa. U literaturi ne postoji konsenzus o jedinstvenom kriteriju na temelju kojeg bi odredili da li je vremenski niz (realizacija procesa) stacionaran ili nestacionaran, no postoji mnogo metoda razvijenih na temelju različitih pretpostavki. To ostavlja prostora za dodatna istraživanja na tom području dajući upravo i motivaciju ovog rada.

Metoda empirijskih ortogonalnih funkcija (EOF) omogućuje dekompoziciju vremenskog niza na svojstvene ortogonalne modove. Rezultat primjene EOF analize na vremenski niz je matrica čiji je svaki stupac jedna ortogonalna funkcija (matricu tih funkcija također nazivamo EOF) i matrica čiji je svaki stupac jedna glavna komponenta (matricu glavnih komponenti nazivamo PC). Množenjem  $n$ -tog stupca matrice EOF i  $n$ -tog stupca matrice PC dobiva se  $n$ -ti mod EOF dekompozicije vremenskog niza. Također, moguće je odrediti udio svakog moda u varijabilnosti početnog niza.

Kako bi prepoznali na koji način EOF analiza reagira na nestacionarnost unutar nizova, primijenjena je u tri koraka na četiri različita tipa vremenskih nizova. U trećem koraku, EOF analiza je primijenjena na realnim vremenskim nizovima. Oni su dobiveni mjerenjima brzina vjetra i sonične temperature na tornju blizu Masleničkog mosta, sjeverno od Zadra u razdoblju od 9. listopada 2015. do 9. listopada 2016. za vrijeme epizoda bure. Na temelju rezultata tih primjena doneseno je nekoliko važnih zaključaka. Najvažniji od njih je da EOF analiza doista može prepoznati nestacionarnost statističkih momenata prvog i drugog reda vremenskih nizova. Štoviše, uočene su razlike u matricama EOF, PC i raspodjelama udjela modova dobivenih primjenom EOF analize na nizove nestacionarnog prvog momenta, te na nizove nestacionarnog drugog momenta. Na temelju donesenih zaključaka može se naslutiti da postoji mogućnost (uz adekvatno definirane kriterije i metodologiju) razvitka EOF metode, koja bi uz prepoznavanje nestacionarnosti niza, mogla raspoznati i porijeklo te nestacionarnosti (je li to nestacionarnost momenta prvog ili drugog reda). To ostavlja pregršt prostora za daljnja istraživanja.

Ključne riječi: EOF analiza, nestacionarnost, određivanje nestacionarnosti, analiza empirijskih ortogonalnih funkcija

# Implementation of the empirical orthogonal functions analysis to determine nonstationarity of time series

## Abstract

Statistical stationarity is a very important concept in the statistical theory of turbulence. Surface layer parameterizations, that relate the turbulent fluxes to mean gradients through Monin–Obukhov similarity theory, assume the stationarity of the turbulent processes. There is no consensus in the literature on a single criterion to determine whether a time series (process realization) is stationary or non-stationary, but there are many methods developed based on different assumptions. This leaves room for additional research in the area, giving the motivation for this work.

The empirical orthogonal function (EOF) analysis allows the decomposition of a time series into different orthogonal modes. The result of applying EOF analysis to a time series is a matrix whose each column is an orthogonal function (the matrix of these functions is also called EOF) and a matrix whose each column is one principal component (the matrix of principal components is called PC). It is also possible to determine the share of each mode in the variability of the initial time series.

To identify how EOF analysis responds to non-stationarity within time series, it was applied in 3 steps to 4 different types of time series. In the third step, EOF analysis is applied on time series of wind speed and sonic temperature measured near Maslenica bridge, north of the city of Zadar in the period from October 9, 2015, to October 9, 2016, during events of bora. Based on the results of these applications, several important conclusions have been drawn. The most important of these is that EOF analysis can indeed recognize the non-stationarity of the first and second-order statistical moments. Moreover, differences between EOF, PC matrices and the distributions of shares of modes obtained by applying EOF analysis to the series with non-stationary first-moment and the series with non-stationary second-moment, were observed. Based on the conclusions reached, it can be suggested that there is a possibility (with adequately defined criteria and methodology) of developing an EOF method, which, by recognizing the non-stationarity of a series, could also recognize the origin of that non-stationarity (whether it is the non-stationarity of the first or second-order moment). This leaves plenty of room for further research.

Keywords: EOF analysis, non-stationarity, determination of non-stationarity, empirical orthogonal function analysis

# Sadržaj

<b>1. Uvod</b>	<b>1</b>
<b>2. Teorijska podloga</b>	<b>2</b>
2.1. Statistička stacionarnost . . . . .	2
2.2. Test obrnutog rasporeda (RAT metoda) . . . . .	3
2.3. Test fluktuiranja oko medijana (RUT metoda) . . . . .	4
2.4. Mahrt-ova metoda (M98 metoda) . . . . .	5
2.5. EOF analiza . . . . .	6
2.6. Proces autoregresije . . . . .	9
<b>3. Umjetni podaci</b>	<b>10</b>
<b>4. Primjena EOF analize na umjetne vremenske nizove</b>	<b>13</b>
4.1. Primjena EOF analize na nizove s bijelim šumom . . . . .	13
4.2. Primjena EOF analize na nizove izvedene iz AR(1) procesa . . . . .	27
<b>5. Realni podaci</b>	<b>34</b>
<b>6. Primjena EOF analize na realne vremenske nizove</b>	<b>35</b>
<b>7. Zaključak</b>	<b>38</b>
<b>8. Daljnja istraživanja</b>	<b>40</b>
<b>Literatura</b>	<b>41</b>

# 1. Uvod

Statistička stacionarnost (u daljnjem tekstu, stacionarnost) je vrlo važan koncept kada je riječ o analizi vremenskih nizova, i njihovoj primjeni u mnogim granama znanosti. Formalna definicija stacionarnosti dana je nešto kasnije. Neformalno, stacionarnost vremenskog niza je posljedica činjenice da se statistička svojstva procesa koji generira spomenuti niz ne mijenjaju u vremenu. Važno je primijetiti da to ne znači da se sam vremenski niz ne mijenja u vremenu, već da se način na koji se on mijenja, ne mijenja u vremenu.

U statističkoj teoriji turbulencije stacionarnost je od posebnog značaja. Parametrizacije površinskog sloja koje povezuju turbulentne tokove s gradijentima srednje vrijednosti preko Monin-Obukovljeve teorije sličnosti pretpostavljaju stacionarnost kao posljedicu ergotske hipoteze (npr. Stul, 1988). Cijela teorija dakle počiva na usrednjavanju ansambla realizacija svakog turbulentnog procesa. Kako je takvo usrednjavanje nemoguće postići u atmosferi jer je realizacija svakog procesa neponovljiva, mora se pretpostaviti kako su vremenski srednjaci jedne realizacije turbulentnog procesa upravo srednjaci ansambla njegovih realizacija. To je ostvarivo samo kod stacionarnih procesa (npr. Bendat i Piersol, 1986). Ukoliko vremenski niz (realizacija turbulentnog procesa) nije stacionaran, tada je vremenski srednjak nekog njegovog svojstva loša aproksimacija srednjaka ansambla realizacija tog procesa. To uvodi pogrešku i nepouzdanost u statistiku turbulencije. Stacionarnost je dakle, nužan uvjet za razmatranje turbulencije standardnom statističkom teorijom turbulencije pa je stoga vrlo važno izbaciti nestacionarne vremenske nizove iz njenog razmatranja.

Pojava nestacionarnosti u atmosferi je vrlo česta. Ona može primjerice nastati kao posljedica dnevnog ciklusa hlađenja i grijanja Zemljine površine, što proizvodi trend u vremenskim nizovima i promjenjivost varijance. Utjecaj nestacionarnosti u obradi podataka moguće je minimizirati "izrezivanjem" i zanemarivanjem nestacionarnih dijelova iz vremenskog niza (Mahrt, 1998). Kako bi se mogli izrezati određeni dijelovi niza, nužno je imati kriterij na temelju kojeg bi odredili stacionarnost ili nestacionarnost pojedinog podniza. Kako u literaturi ne postoji konsenzus o jedinstvenom kriteriju na temelju kojeg bi se to moglo odrediti, određivanje nestacionarnosti preostaje otvoreno za daljnja istraživanja. Iz te činjenice upravo proizlazi motivacija ovog rada.

U ovom radu proučit ćemo mogućnost uporabe EOF analize u detektiranju nestacionarnosti u vremenskim nizovima primjenjujući je na slijed vremenskih nizova čija su statistička svojstva poznata. Nakon zaključka da EOF analiza uistinu reagira na postojanje nestacionarnosti u vremenskom nizu, postavljamo ju u perspektivu tako što na istom setu realnih podataka primjenjujemo EOF analizu paralelno s tri od prije poznate metode za određivanje nestacionarnosti te uspoređujemo njihove rezultate.

## 2. Teorijska podloga

### 2.1. Statistička stacionarnost

Stacionarnost je statističko svojstvo stohastičkog procesa ili vremenskog niza, koje govori da je proces ostvario svojstveno statistički ravnotežno stanje u kojem mu se raspodjela ne mijenja značajno. Stacionarnost garantira da su svojstva procesa ili vremenskog niza poput srednjaka, varijance, autokovarijance (...) konstantna u vremenu. S obzirom na to da li su sve karakteristike (svi statistički momenti) ili samo dio njih (npr. momenti prvog i drugog reda) nepromjenjivi u vremenu, može se govoriti o različitim tipovima stacionarnosti: stacionarnost u užem smislu/potpuna stacionarnost (svi momenti su konstantni u vremenu) i slaba stacionarnost/stacionarnost u širem smislu/stacionarnost do reda  $m$  (prvih  $m$  momenta konstantni u vremenu).

Za stohastički proces (ili vremenski niz)  $(X_t)_{t \in \mathbb{Z}}$  kažemo da je potpuno stacionaran ako je zajednička funkcija gustoće vjerojatnosti od  $(X_{t_1}, \dots, X_{t_k})$  identična funkciji gustoće od  $(X_{t_1+\tau}, \dots, X_{t_k+\tau})$  za svaki  $\tau$ , gdje je  $k$  proizvoljni pozitivni cijeli broj (Myers, 1989).

Ekvivalentno potpunoj stacionarnosti je tvrdnja da se funkcija gustoće vjerojatnosti procesa ne mijenja u vremenu. Takva definicija stacionarnosti je prerestriktivna. Procesi u prirodi u pravilu ne pokazuju takvo ponašanje, stoga pribjegavamo manje strogoj definiciji stacionarnosti.

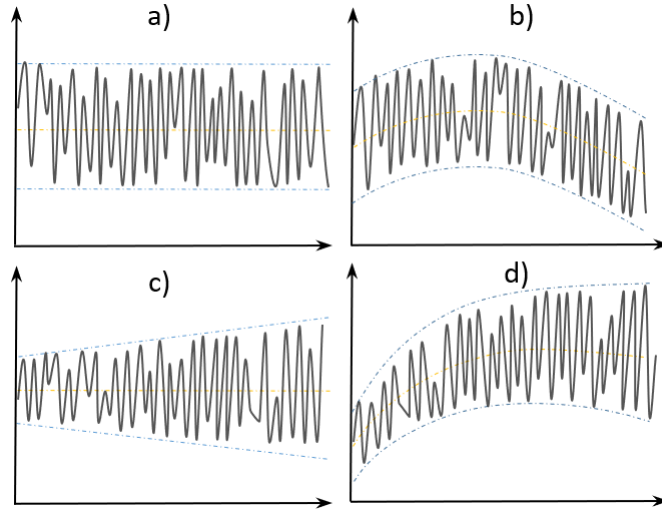
Stohastički proces (ili vremenski niz)  $(X_t)_{t \in \mathbb{Z}}$  je stacionaran drugog reda (stacionaran u užem smislu) ako zadovoljava sljedeće uvjete (Myers, 1989):

1.  $\mathbb{E}[X_t] = \mathbb{E}(X_{t+\tau}) = \mu, \forall \tau \in \mathbb{N}$
2.  $\mathbb{E}[(X_t - \mu)^2] = \text{Var}(X_t) < \infty$
3.  $\mathbb{E}[(X_{t_1} - \mu)(X_{t_2} - \mu)] = \mathbb{E}[(X_{t_1+\tau} - \mu)(X_{t_2+\tau} - \mu)] = \text{Cov}(X_{t_1}, X_{t_2}) = f(t_2 - t_1), \forall \tau, t_1, t_2 \in \mathbb{N}.$

Treći uvjet vrijedi za svaki  $t_1$  i  $t_2$  što znači da vrijedi i za  $t_1 = t_2 = t$ . Tada izraz pod brojem tri postaje  $\mathbb{E}[(X_t - \mu)(X_t - \mu)] = \mathbb{E}[(X_t - \mu)^2] = f(t - t) = f(0) = \textit{konst.} \iff \text{Var}(X_t) = \textit{konst.} \forall t \in \mathbb{N}$ . To implicira da je varijanca stacionarnog procesa drugog reda također konstantna u vremenu kao i srednjak. Važno je naglasiti da se sve do sada spomenuto odnosilo na stacionarnost stohastičkog procesa. U ovom radu proučava se stacionarnost vremenskog niza, dakle konačno duge realizacije procesa, no zaključci za stacionarnost nizova su jednaki onima za stacionarnost procesa. Upravo je stacionarnost drugog reda ona čije postojanje najčešće želimo utvrditi i na koju se ovaj rad fokusira. Dakle, kada se niz proglasi stacionarnim, podrazumijevamo da su



mu srednja vrijednost te varijanca konstantni ili bolje rečeno - dovoljno malo variraju u vremenu da ih se može proglasiti konstantnim (slika 2.1 a)). Niz se može proglasiti nestacionarnim iz tri različita razloga: varijabilne srednje vrijednosti (slika 2.1 b)), varijabilne varijance (slika 2.1 c)) te varijabilnosti varijance i srednje vrijednosti (slika 2.1 d)).



Slika 2.1. Shematski prikaz niza stacionarnog srednjaka i stacionarne varijance (a), nestacionarnog srednjaka i stacionarne varijance (b), stacionarnog srednjaka i nestacionarne varijance (c), nestacionarnog srednjaka i varijance (d).

U nastavku su predstavljene tri postojeće metode za određivanje stacionarnosti vremenskih nizova. Opis istih preuzet je iz Večenaja i de Wekkera (2015).

## 2.2. Test obrnutog rasporeda (RAT metoda)

Pretpostavimo da se niz  $Q_i, i = 1, 2, 3, \dots, N$  sastoji od  $N$  međusobno nezavisnih opažanja slučajne varijable  $Q$ , što je u ovom radu neki određeni statistički moment. Svako opažanje  $Q_i$  iz niza, od  $i = 1$  do  $i = N - 1$ , uspoređuje se sa svakim sljedećim opažanjem u nizu. Dakle, opažanje  $Q_i$  se uspoređuje s opažanjima  $Q_{i+1}, Q_{i+2}, \dots, Q_N$  što je ukupno  $N - i$  parova  $(Q_i, Q_j)$  gdje je  $j = i + 1, i + 2, \dots, N$  za svaki  $i$ . Ukoliko za par vrijedi  $Q_i > Q_j$ , tada oni čine obrnuti raspored (eng. *reverse arrangement*, od kuda i ime testa - *reverse arrangement test*, skraćeno RAT) te svakom takvom paru pridružuje se vrijednost  $a_{ij} = 1$ . Svakom  $Q_i < Q_j$  pridružuje se  $a_{ij} = 0$ . Možemo definirati sljedeću veličinu:

$$A = \sum_{i=1}^{N-1} \sum_{j=i+1}^N a_{ij}, \quad (2.1.)$$

koja je, uz pretpostavku da niz  $Q_i$  ne posjeduje trend, zapravo slučajna varijabla s očekivanom vrijednošću  $\mu_A$  i varijancom  $\sigma_A^2$ . Prema Stuart et al. (1963) i Bendat i

Piersol (1986) tada vrijedi:

$$\mu_A = \frac{N(N-1)}{4}, \quad (2.2.)$$

$$\sigma_A^2 = \frac{N(2N+5)(N-1)}{72}. \quad (2.3.)$$

Null hipoteza da u nizu nema trenda može biti odbijena na nivou signifikantnosti  $\alpha$  ukoliko  $A$  upada unutar kritičnog intervala  $[A_d, A_g]$ , gdje su vrijednosti  $A_d$  i  $A_g$  dane u tablici 2.1. Dakle, ako test za neki niz daje vrijednost parametra  $A$  tako da vrijedi  $A_d < A < A_g$  za taj niz RAT metoda tvrdi da ne posjeduje trend.

Tablica 2.1. Kritične vrijednosti parametara metoda RAT, RUT i M98 definirane u poglavljima 2.2, 2.3 i 2.4.

	RAT	RUT	M98
kritične	$A_g = 272$	$R_g = 21$	$NR_{krit} = 2$
vrijednosti	$A_d = 162$	$R_d = 10$	

### 2.3. Test fluktuiranja oko medijana (RUT metoda)

Pretpostavimo da se niz  $Q_i$   $i = 1, 2, 3, \dots, N$  sastoji od  $N$  međusobno nezavisnih opažanja slučajne varijable  $Q$  (statističkog momenta), te posjeduje medijan  $m$ . Neka je  $N_G^+$  broj grupa uzastopnih vrijednosti  $Q_i$  iznad medijana, a  $N_G^-$  broj grupa uzastopnih vrijednosti  $Q_i$  ispod medijana. Ako niz ne posjeduje trend (null hipoteza), tada zbroj tih dviju grupa jest slučajna varijabla. Onda vrijedi:

$$R = N_G^+ + N_G^-, \quad (2.4.)$$

s očekivanjem

$$\mu_R = \frac{N}{2} + 1 \quad (2.5.)$$

i varijancom

$$\sigma_R^2 = \frac{N(N-2)}{4(N-1)}. \quad (2.6.)$$

Null hipotezu da niz  $Q_i$  nema trend odbacuje se na nivou signifikantnosti  $\alpha$  ako je parametar  $R$  unutar intervala  $[R_d, R_g]$ . Granice stacionarnosti  $R_d$  i  $R_g$  dane su u tablici 2.1.

## 2.4. Mahrt-ova metoda (M98 metoda)

Mahrt-ova metoda (Mahrt, 1998) bazira se na pretpostavci da su statistički momenti stacionarnog vremenskog niza gaussovski raspodijeljeni. Iz te pretpostavke slijedi da standardne devijacije statističkih momenata, izračunate na manjim segmentima niza, mogu predvidjeti standardnu devijaciju statističkih momenata na dužem dijelu vremenskog niza. Metoda podrazumijeva računanje parametra  $NR$  koji predstavlja omjer standardne devijacije određenog statističkog momenta između više segmenata vremenskog niza i srednje standardne devijacija podataka unutar segmenta. Početni vremenski niz duljine  $N$  se dijeli na  $I$  podnizova gdje svaki ima  $J$  segmenata. Računaju se prvi i drugi momenti tih segmenata  $Q_{ij}$  za  $i = 1, \dots, I$  i  $j = 1, \dots, J$ , nakon čega se računaju srednje vrijednosti momenata svakog podniza

$$SR(i) = \frac{1}{J} \sum_{j=1}^J Q_{ij}, \quad (2.7.)$$

što daje ukupno  $I$  srednjih vrijednosti. U idućem koraku, računa se srednja vrijednost srednjih vrijednosti svih podnizova

$$SR' = \frac{1}{I} \sum_{i=1}^I \left( \frac{1}{J} \sum_{j=1}^J Q_{ij} \right), \quad (2.8.)$$

te devijacije svakog segmenta od te srednje vrijednosti

$$Q'_{br}(i) = \frac{1}{J} \sum_{j=1}^J (Q_{ij} - SR'). \quad (2.9.)$$

Sada je moguće izračunati standardnu devijaciju takvog niza (devijacija između podnizova),

$$\sigma_{Q_{br}} = \sqrt{\frac{1}{I-1} \sum_{i=1}^I Q_{br}'^2(i)}. \quad (2.10.)$$

Za svaki segment  $J$  moguće je izračunati standardnu devijaciju unutar svakog segmenta

$$\sigma_{Q_{wr}}(i) = \sqrt{\frac{1}{J-1} \sum_{j=1}^J (Q_{ij} - SR(i))^2}. \quad (2.11.)$$

Srednja vrijednost standardnih devijacija svih  $I$  podnizova je tada (Bendat i Piersol, 1986)

$$\bar{\sigma}_{Q_{wr}} = \frac{1}{I} \sum_{i=1}^I \sigma_{Q_{wr}}(i) \quad (2.12.)$$

$$RE = \frac{\bar{\sigma}_{Q_{wr}}}{\sqrt{J}}. \quad (2.13.)$$

Omjer standardne devijacije statističkog momenta između više podnizova i srednje standardne devijacije segmenata unutar podnizova tada je:

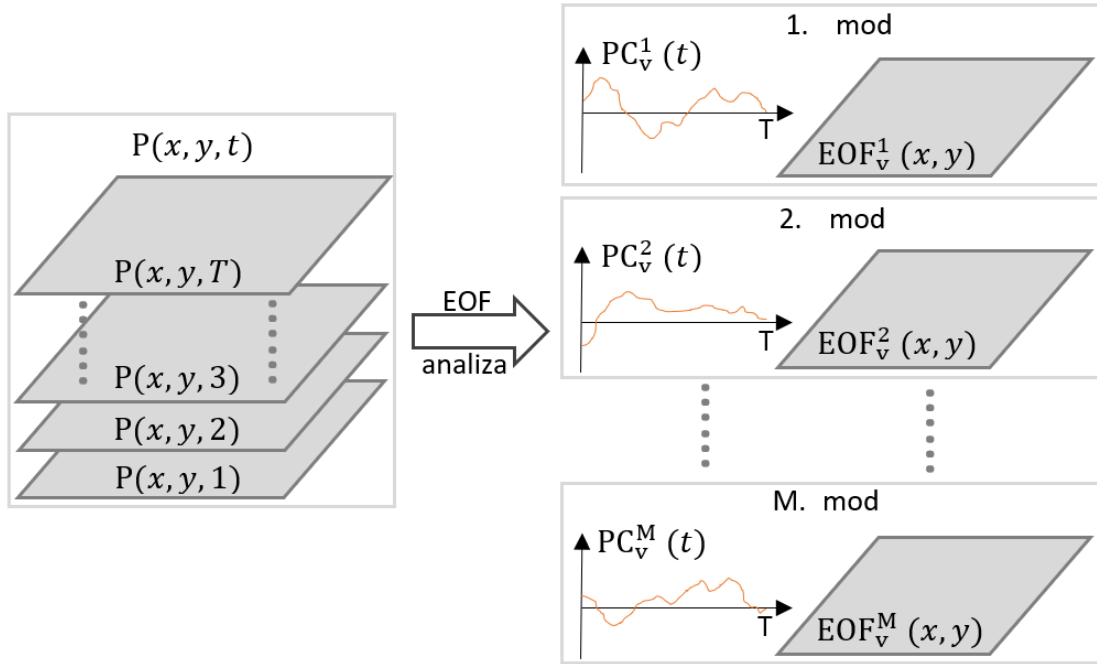
$$NR = \frac{\sigma_{Q_{br}}}{RE}. \quad (2.14.)$$

Za stacionarne statističke momente  $NR$  danog vremenskog niza bi trebao biti  $\sim 1$ . Kako bi se kriterij ublažio, kao gornja granica stacionarnosti uzima se kritična vrijednost 2 (dana u tablici 2.1).

## 2.5. EOF analiza

Analiza empirijskih ortogonalnih funkcija ili EOF (eng. *Empirical Orthogonal Function*) se idejno jednostavno može objasniti kao dekompozicija seta podataka na svojstvene ortogonalne modove (slika 2.2). Koristi se kao metoda za analiziranje varijabilnosti skalarnog polja podataka. Ona je čisto matematički koncept koji je zbog svoje univerzalnosti svoju primjenu pronašao u raznim područjima znanosti. U geofizičkim znanostima koristi se na poljima fizikalnih varijabli (površinska temperatura, tlak ...) koje imaju vremenski razvoj (npr. Hannachi, 2004; Preisendorfer i Mobley, 1988). Rezultati primjene EOF analize su uzorci (motivi, eng. *patterns*) varijabilnosti, njihove vremenske varijacije te mjera zastupljenosti svakog moda u ulaznom polju podataka (slika 2.2). Kako literatura nije jednoglasna o terminologiji EOF analize te načinu njene provedbe, ovdje je provedena kako su to učinili Björnsson i Venegas (1997). Rezultati EOF analize provedene na takav način su: matrica kojoj su stupci ortogonalne empirijske funkcije - upravo one su prije spomenuti motivi, matrica kojoj su stupci glavne komponente - vremenski razvoj motiva i dijagonalna matrica kojoj je element na dijagonali svakog stupca proporcionalan udjelu pojedinog moda EOF dekompozicije u početnom polju podataka. U nastavku su navedene oznake koje će biti korištene kroz čitav rad i njihovo značenje. *EOF* - matrica kojoj su stupci ortogonalne empirijske funkcije, *EOF<sub>v</sub>* - ortogonalna empirijska funkcija, *PC* - matrica kojoj su stupci glavne komponente, *PC<sub>v</sub>* - glavna komponenta.

EOF analiza može se matematički provesti na dva načina, koristeći matricu kovarijanci ili koristeći SVD (eng. *Singular Value Decomposition*) metodu. U ovom radu, EOF analiza je provedena uporabom matrice kovarijanci, stoga će ista biti detaljnije objašnjena. Dani izvod je proveden na primjeru proizvoljne geofizičke varijable  $P$  kao što su to učinili Björnsson i Venegas (1997). Mjerenja fizikalne veličine  $P$  prikupljena su na  $N$  lokacija u  $T$  vremenskih koraka. Ti podatci se organiziraju u matricu podataka  $F$ , dimenzija  $T \times N$ . Svaki stupac predstavlja mjerenja provedena na pojedinoj lokaciji,



Slika 2.2. Shematski prikaz rastava polja  $F(x, y, t)$  na  $M$  modova dobivenih EOF analizom.

dok svaki redak predstavlja mjerenja u pripadnom vremenskom koraku. Dakle, ideja je matricu podataka  $F$  rastaviti na  $M$  ortogonalnih EOF modova na sljedeći način:

$$F = \sum_{m=1}^M PC_v^m \cdot (EOF_v^m)^T. \quad (2.15.)$$

Takav rastav je shematski prikazan na Slici 2.2. Nakon konstrukcije matrice podataka  $F$ , računa se matrica kovarijanci

$$Q = \frac{1}{T-1} F^T F. \quad (2.16.)$$

Pretpostavimo nadalje, da je iz svakog stupca  $F$  matrice uklonjena srednja vrijednost tog stupca, ili drugim riječima, da je srednjak svakog stupca jednak nuli. Ta pretpostavka nikako ne utječe na samu provedbu EOF analize, već omogućuje da matricu  $Q$  interpretiramo kao matricu kovarijanci. Matrica  $Q$  je dimenzija  $N \times N$ . Rješavajući problem svojstvenih vrijednosti matrice  $Q$  dobiva se

$$QV = V\Lambda, \quad (2.17.)$$

gdje je  $V$  matrica  $N$  svojstvenih vektora matrice  $Q$ . Svaki stupac matrice  $V$  je jedan svojstveni vektor s pripadnom svojstvenom vrijednošću  $\lambda$  na dijagonali matrice  $\Lambda$  u pripadnom stupcu. Matrica svojstvenih vektora odgovara upravo matrici empirijskih ortogonalnih funkcija,  $EOF = V$ . Svaki svojstveni vektor odgovara jednom  $EOF_v$ -u, a time i modu, dok je njegova pripadna svojstvena vrijednost zapravo mjera udjela

varijabilnosti objašnjene tim modom. Primijetimo da je najveći broj različitih modova jednak broju empirijskih ortogonalnih funkcija u matrici  $EOF$ , tj. broju stupaca  $N$ . Pošto su  $EOF_v$ -ovi u matrici  $EOF$  poredani tako da je u prvom stupcu  $EOF_v$  najzastupljenijeg moda, može se reći da prvih  $M$  modova dovoljno dobro opisuje početno polje podataka  $F$  zanemarujući pri tome ostale modove. U tom slučaju daljnji se proračuni rade s matricom  $EOF$  dimenzija  $N \times M$ , gdje je  $M$  cijeli broj iz intervala  $1 \leq M \leq N$ . Kako svakom pojedinom  $EOF_v$ -u odgovara jedan  $PC_v$ , do njih se dolazi projekcijom matrice podataka na pripadni  $EOF_v$ . Takva se projekcija može vrlo efikasno provesti simultano za sve  $EOF_v$ -ove na način:

$$PC = F \cdot EOF, \quad (2.18.)$$

gdje je  $PC$  matrica ( $T \times M$ ) glavnih komponenata gdje svaki stupac odgovara jednom  $PC_v$ -u. U tom slučaju  $m$ -ti mod može se dobiti vektorskim množenjem  $m$ -tog stupca matrice  $EOF$  s  $m$ -tim stupcem matrice  $PC$ .

$$MOD_m[T \times N] = PC_v^m[T \times 1] \cdot (EOF_v^m)^T[1 \times N]. \quad (2.19.)$$

Postotak objašnjene varijabilnosti matrice podataka  $F$   $m$ -tim modom dan je s:

$$\eta_m = \frac{\lambda_m}{Tr(\Lambda)} \cdot 100\%. \quad (2.20.)$$

Kako su vremenski nizovi od našeg interesa zapravo mjerenja neke fizikalne varijable na jednoj lokaciji u vremenu, matrica  $F$  imala bi dimenzije  $T \times 1$ , jer je  $N = 1$ . To bi vodilo do trivijalnog rezultata s jednim modom. Tom problemu se doskače činjenicom da su u proučavanju turbulencije vremenski nizovi dovoljno dugački (najčešće 30 min) da ih je moguće podijeliti na neki broj ( $N$ ), međusobno nezavisnih podnizova (Večenaj i de Wekker, 2015). Na taj način matrica podataka  $F$  konstruirana je tako da svaki stupac odgovara nezavisnom podnizu originalnog vremenskog niza. Matrica  $F$  je dimenzija  $T/N \times N$ , gdje je  $T/N$  duljina svakog podniza. Iz nje, EOF analiza će nastojati izdvojiti u prvi mod najzastupljeniji motiv vremenske varijacije zajedničke svim podnizovima.

## 2.6. Proces autoregresije

Proces autoregresije (AR) je stohastički proces koji se nerijetko koristi za simuliranje, te prognozu vremenski ovisnih procesa u prirodi. Takav model pretpostavlja da vrijednost varijable u sadašnjem trenutku linearno ovisi o svojim vrijednostima u prošlim trenucima, te o bijelom šumu. Općenita rekurzivna relacija AR(p) modela autoregresije, reda  $p$  glasi :

$$X_t = C + \epsilon_t + \sum_{i=1}^p \phi_i X_{t-i} \quad (2.21.)$$

gdje je  $C$  konstanta,  $\epsilon_t$  je bijeli šum iz  $\mathcal{N}(0, 1)$ ,  $\phi_i$  su koeficijenti autoregresije. Prema Hamilton (1994) rekurzivnu relaciju 2.21 moguće je zapisati u alternativnom obliku:

$$\Phi(L)X_t = C + \epsilon_t \quad (2.22.)$$

Gdje je  $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$  svojstvena jednačba AR procesa reda  $p$ . Kako bi AR(p) model reproducirao stacionaran (u slabom smislu) niz, za korijene  $\lambda_i$  svojstvene jednačbe mora vrijediti  $|\lambda_i| > 1$ . Svaki korijen tada zadovoljava  $0 = 1 - \phi_1 \lambda_i - \phi_2 \lambda_i^2 - \dots - \phi_p \lambda_i^p$ .

### 3. Umjetni podaci

Kako bi se istražila mogućnost primjene EOF analize u svrhu određivanja nestacionarnosti, EOF analiza će u ovom radu biti primijenjena prvo na umjetne nizove čija su statistička svojstva unaprijed poznata. Tek nakon što se na takvim nizovima izdvoje specifičnosti reagiranja EOF analize na nestacionarnost, biti će primijenjena na realne nizove. Turbulencija se proučava na vremenskim nizovima mjerenja dugim 30 minuta (npr. Večenaj i De Wekker, 2015). Mjerenja brzine vjetra i sonične temperature na koja ćemo primijeniti EOF analizu, vrše se frekvencijom od 20 Hz. Vremenski niz dug 30 minuta tada sadrži 36000 podataka. Kako bi ostali konzistentni, umjetni nizovi koji će biti konstruirani sadržavat će također 36000 podataka.

U prvom koraku EOF analizu primjenjujemo na niz bijelog šuma dobiven iz Gaussovog slučajnog procesa srednje vrijednosti 0 i varijance 1:

$$Q \equiv (\epsilon_t)_{t=1, \dots, 36000}, \epsilon_t \sim \mathcal{N}(0, 1). \quad (3.1.)$$

Takav vremenski niz je odabran zbog vrlo lakog konstruiranja te zato što su mu prvi i drugi moment stacionarni. Kako bi proučili kako EOF analiza reagira na srednju vrijednost niza različitu od 0 primjenjuje se na:

$$Q_2 \equiv (\epsilon_t + 2)_{t=1, \dots, 36000}, \epsilon_t \sim \mathcal{N}(0, 1). \quad (3.2.)$$

Vremenske nizove nestacionarnog momenta prvog reda konstruiramo tako da na niz  $Q$  superponiramo 3 vrste hoda. Linearni porast

$$Q_3 \equiv (\epsilon_t + \frac{t}{9000})_{t=1, \dots, 36000}, \epsilon_t \sim \mathcal{N}(0, 1), \quad (3.3.)$$

parabolični hod

$$Q_4 \equiv (\epsilon_t + \frac{1}{400}(\frac{t}{720} - 25)^2)_{t=1, \dots, 36000}, \epsilon_t \sim \mathcal{N}(0, 1), \quad (3.4.)$$

sinusoidalni hod

$$Q_5 \equiv (\epsilon_t + 1.5 \sin(\frac{2\pi t}{36000}))_{t=1, \dots, 36000}, \epsilon_t \sim \mathcal{N}(0, 1). \quad (3.5.)$$

Način na koji EOF analiza reagira na nestacionarnost drugog momenta proučava se također na nizovima izvedenim iz  $Q$ . No umjesto da se na  $Q$  superponira neka funkcija vremena (kao u prethodnim primjerima), sada se niz  $Q$  množi funkcijom vremena. Linearni porast varijance tada se može simulirati nizom

$$Q_6 \equiv (\epsilon_t \cdot \frac{t}{9000})_{t=1, \dots, 36000}, \epsilon_t \sim \mathcal{N}(0, 1), \quad (3.6.)$$

parabolični hod varijance možemo simulirati nizom



$$Q_7 \equiv (\epsilon_t \cdot \frac{1}{400} (\frac{t}{720} - 25)^2)_{t=1, \dots, 36000}, \epsilon_t \sim \mathcal{N}(0, 1), \quad (3.7.)$$

sinusoidalni hod varijance je tada

$$Q_8 \equiv (\epsilon_t \cdot 1.5 \sin(\frac{2\pi t}{36000}))_{t=1, \dots, 36000}, \epsilon_t \sim \mathcal{N}(0, 1). \quad (3.8.)$$

Kako nestacionarnost prvog i drugog momenta istovremeno utječe na rezultate EOF analize proučava se na nizu:

$$Q_9 \equiv (\epsilon_t \cdot (\frac{t^2}{14000 \cdot 36000} + 1/2) + 8 \cdot \tanh(\frac{t - 18000}{7200}) + 15)_{t=1, \dots, 36000}, \epsilon_t \sim \mathcal{N}(0, 1). \quad (3.9.)$$

U idućem koraku, EOF analizu primjenjujemo na niz dobiven iz AR procesa koristeći jednadžbu 2.21. Konkretno, koristimo AR(1) proces :

$$X \equiv (X_t)_{t=1, \dots, 36000}, \quad (3.10.)$$

gdje je

$$X_t = \phi X_{t-1} + \epsilon_t, \epsilon_t \sim \mathcal{N}(0, 1). \quad (3.11.)$$

Za  $\phi$  je odabrana vrijednost  $\phi = 0.995$ . Nizovi dobiveni AR(1) procesom su stacionarni drugog reda. Proučavanje proširujemo tako da iz takvog niza izvedemo nizove nestacionarnog srednjaka i stacionarne varijance, stacionarnog srednjaka i nestacionarne varijance te nestacionarnog srednjaka i nestacionarne varijance. Niz nestacionarnog srednjaka i stacionarne varijance konstruira se superponirajući na niz  $X$  hod u obliku funkcije tangens hiperbolni:

$$X_1 \equiv (X_t + 8 \cdot \tanh(\frac{t - 18000}{7200}) + 15)_{t=1, \dots, 36000}. \quad (3.12.)$$

Niz stacionarnog srednjaka i nestacionarne varijance konstruira se množeći niz  $X$  s kvadratnom funkcijom vremena:

$$X_2 \equiv (X_t \cdot (\frac{t^2}{14000 \cdot 36000} + 1/2))_{t=1, \dots, 36000}. \quad (3.13.)$$

Niz nestacionarnog srednjaka i nestacionarne varijance konstruira se kombinacijom prethodna dva niza:

$$X_3 \equiv (X_t \cdot (\frac{t^2}{14000 \cdot 36000} + 1/2) + 8 \cdot \tanh(\frac{t - 18000}{7200}) + 15)_{t=1, \dots, 36000}. \quad (3.14.)$$

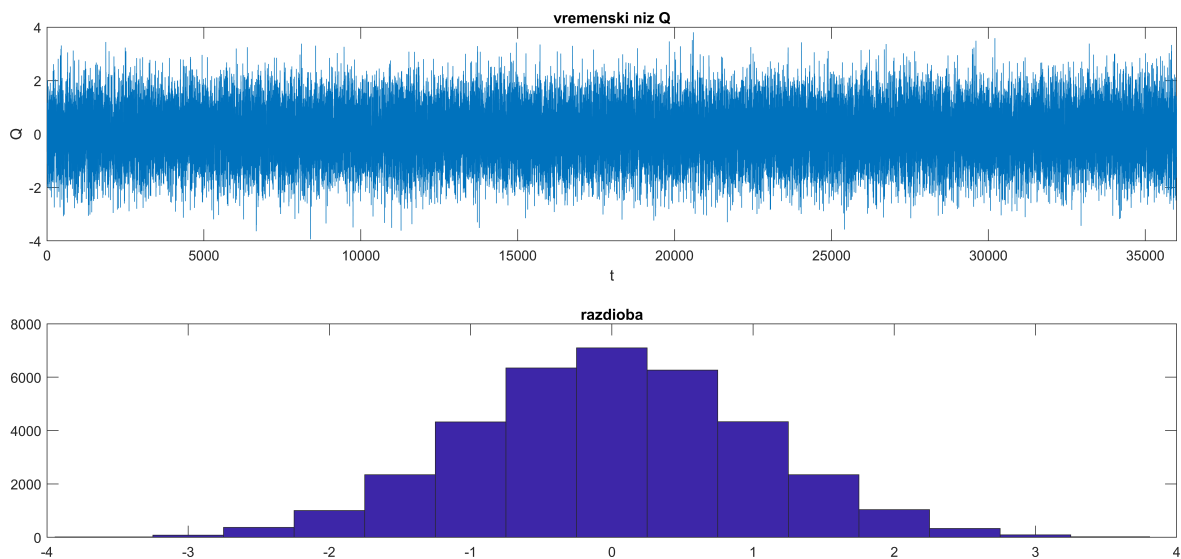
U svim prethodno konstruiranim vremenskim nizovima sve brojčane vrijednosti su

izabrane bez smanjenja općenitosti. Kriterij za odabir upravo takvih konstanti je bio da konstruirani nizovi vizualno dovoljno dobro reprezentiraju naum koji se krije iza njihove konstrukcije. Npr. ako je hod srednjaka niza sinusoidalnog oblika, da on doista bude i vidljiv tijekom crtanja tog niza.

## 4. Primjena EOF analize na umjetne vremenske nizove

### 4.1. Primjena EOF analize na nizove s bijelim šumom

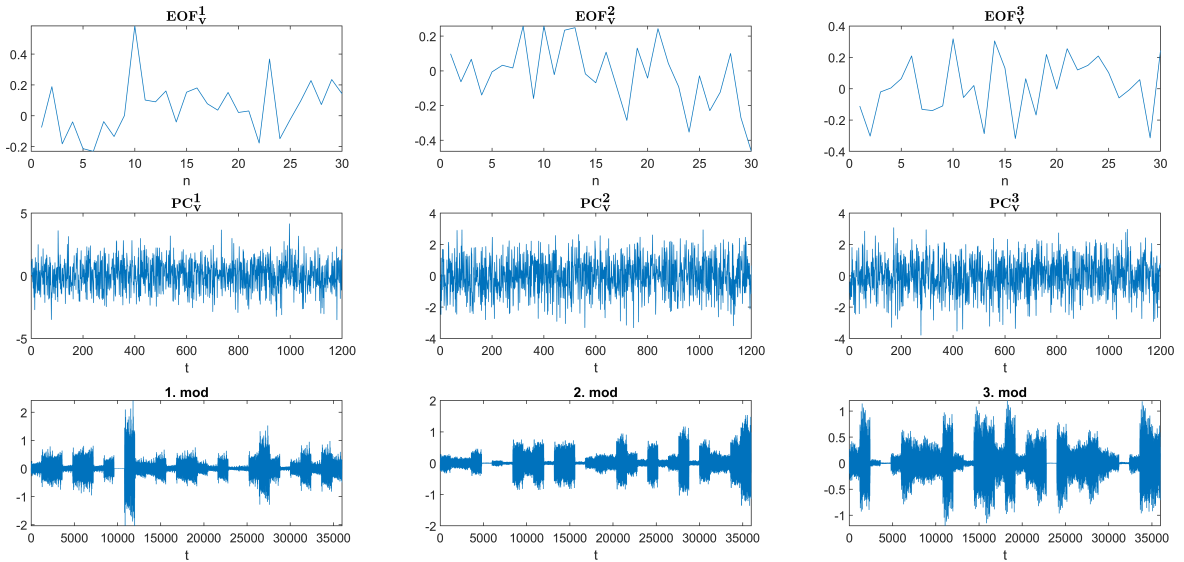
Interpretacija rezultata EOF analize nije trivijalan zadatak (npr. Dommenget i Latif, 2002). Konačni cilj rada je provjeriti da li iz rezultata EOF analize možemo zaključiti da li su nizovi nestacionarni. Kako bi što bolje interpretirali dobivene rezultate EOF analize, Dommenget (2007) predlaže postavljanje svojevrsne null hipoteze. U ovom slučaju pogodna null hipoteza su rezultati EOF analize primijenjene na gotovo savršeno stacionaran niz. Rezultati primjene EOF analize na različite nestacionarne nizove tada se uspoređuju s rezultatima primjene EOF analize na tom stacionarnom nizu, što zaključcima daje na težini. To je motivacija da razmatranje EOF analize u svrhu detekcije nestacionarnosti započnemo upravo od niza bijelog šuma danog jednadžbom 3.1. Razlog odabira duljine niza od 36000 podataka dan je u poglavlju 3. Niz i njegova raspodjela prikazani su na slici 4.1.



Slika 4.1. Gore - vremenski niz  $Q_2$  iz relacije 3.1. Dolje - raspodjela podataka niza  $Q$ .

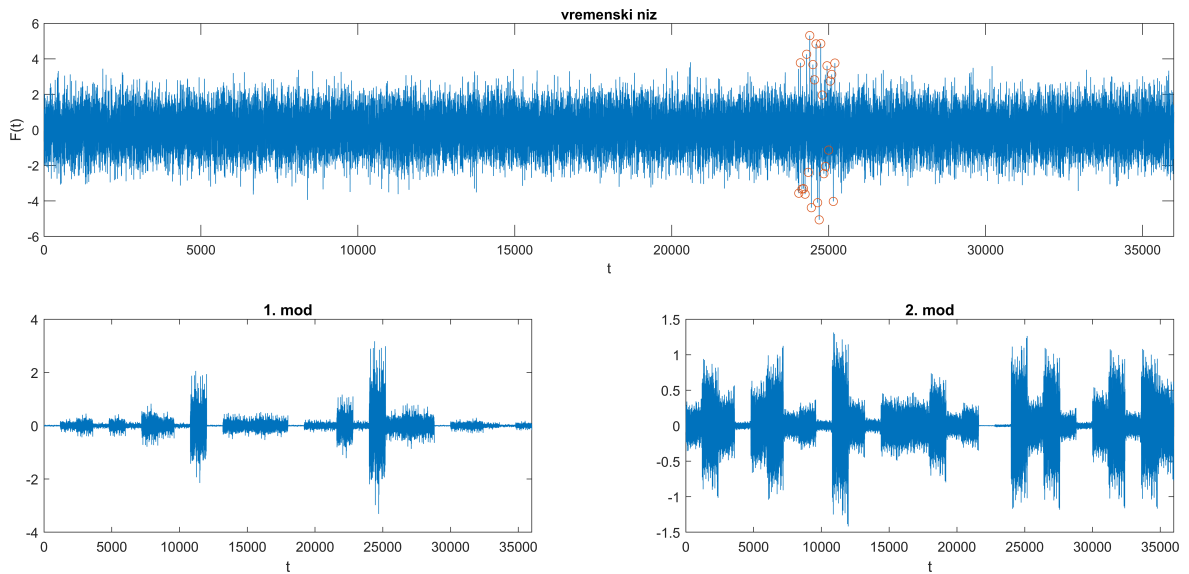
Podsjećamo, EOF analiza se na svim vremenskim nizovima provodi tako da se vremenski niz podijeli na 30 podnizova. Nakon podjele konstruira se matrica podataka tako da je svaki podniz jedan stupac spomenute matrice. Dalje se provodi izvod dan u poglavlju 2.5. Kao rezultat EOF analize dobiva se dakle, 30  $EOF_v$ -ova, 30  $PC_v$ -ova te postotni udjeli pojedinih modova u vremenskom nizu.

Na slici 4.2. prikazana su prva 3  $EOF_v$ -a, odgovarajuća 3  $PC_v$ -a i prva 3 najzasupljenija moda koja oni čine. Možemo primijetiti da su sva 3  $EOF_v$ -a unutar granica  $< -1, 1 >$ . To je posljedica toga što su kao matrica  $EOF$  odabrani ortonormirani svojstveni vektori matrice kovarijanci podataka. Kako se do pojedinih modova



Slika 4.2. Prva 3  $EOF_v$ -a,  $PC_v$ -a i prva 3 moda EOF dekompozicije niza  $Q$  danog izrazom 3.1.

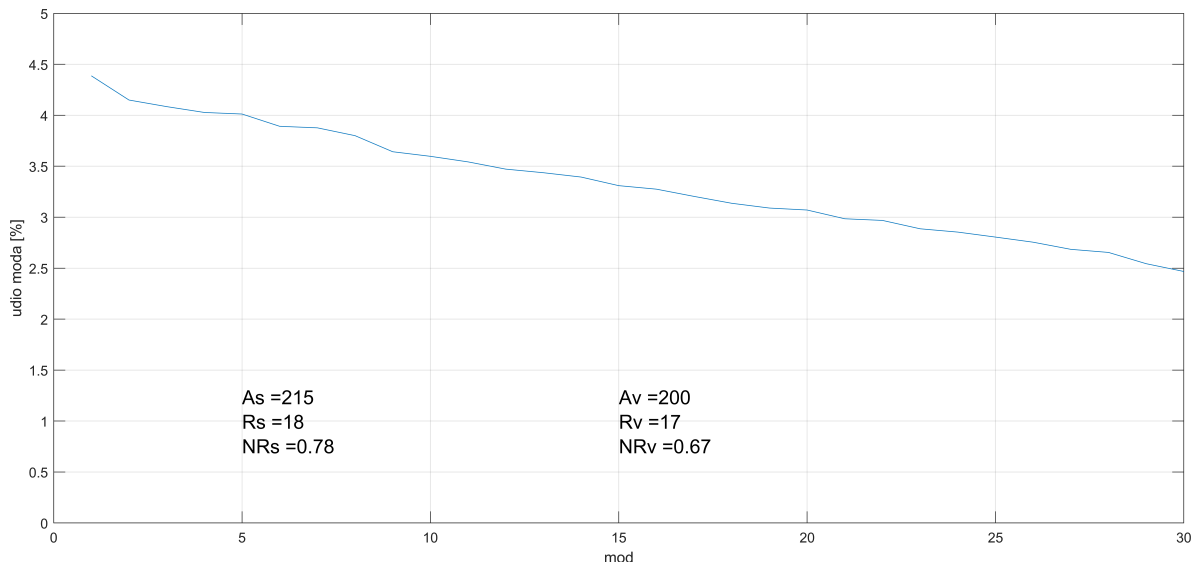
dolazi množenjem pripadnog  $EOF_v$ -a i  $PC_v$ -a, a za  $EOF_v$ -ove je utvrđeno da su normirani, ne čudi da je raspon u kojem varira  $PC_v$  upravo raspon variranja originalnog niza. I  $EOF_v$ -ovi i  $PC_v$ -ovi se doimaju kao da su niz nasumičnih podataka.  $PC_v^1$  se može interpretirati kao dominantni motiv varijacije u vremenu svih 30 podnizova, a  $EOF_v^1$  kao "korekcija" amplitude tog variranja u svakom pojedinom podnizu. Promatrajući prva tri  $EOF_v$ -a i  $PC_v$ -a ne može se uočiti nikakva pravilna struktura ili pravilnost. No gledajući modove koje oni daju, uočavaju se područja koja pojedini mod "uhvati" tj. područja gdje je amplituda varijabilnosti do  $1/2$  amplitude varijabilnosti originalnog



Slika 4.3. Gore - vremenski niz  $Q$  zadan izrazom 3.1 s 24 umjetna podatka prikazana kružićima oko  $t \sim 25000$ . Dolje - prvi i drugi mod EOF dekompozicije gore prikazanog vremenskog niza dobiveni postupkom opisanim u poglavlju 2.5.

niza (npr. područje u 1. modu između 10800 i 12000) i područja u kojima je amplituda varijabilnosti gotovo 0 (npr. područje u 1. modu između 9600 i 10800). Takvo je ponašanje posljedica nastojanja EOF analize da prvim modom uhvati maksimum varijabilnosti niza  $Q$ . To čini tako da prvim modom simulira one najrubnije podatke (koji su u normalnoj raspodjeli najdalje od srednjaka) i to u podnizu gdje je broj rubnih podataka ispod i iznad srednjaka najveća. Drugi mod simulira najrubnije podatke u novom nizu, dobivenom oduzimanjem 1. moda od početnog niza itd... To je relativno lako potvrditi dodamo li u početni niz  $Q$  "umjetne" podatke, koji osobito odstupaju od srednjaka i to u vremenskim koracima koje prvi mod vrlo loše opisuje.

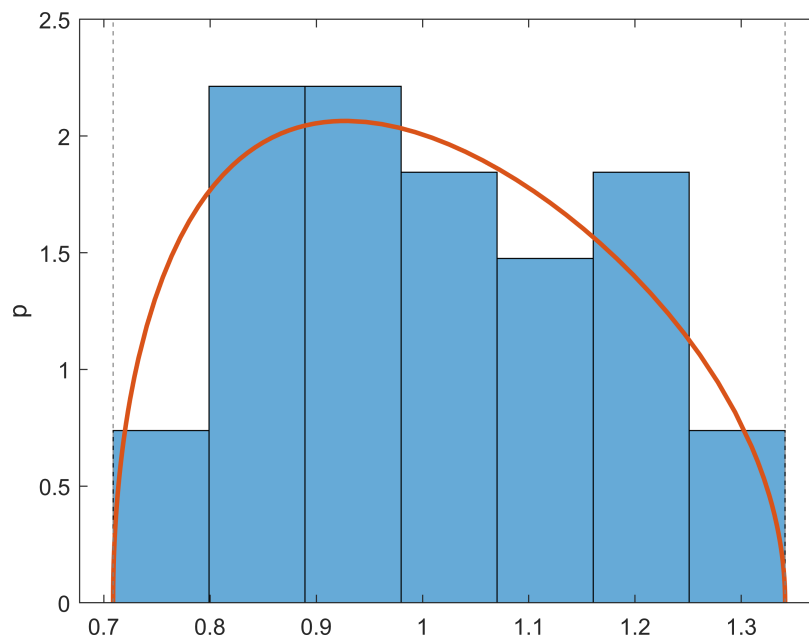
Na slici 4.2. vidimo da prvi mod vrlo loše opisuje niz  $Q$  u vremenskim koracima između 24000 i 25200, stoga oni čine idealno mjesto za umetanje umjetnih podataka. Ako na niz  $Q$  dodamo npr. 24 takva podatka, očekujemo da će EOF dekompozicija takvog niza prvim modom upravo pokušati opisati vremenski period između koraka 24000 i 25200. Slika 4.3 potvrđuje navedeno.



Slika 4.4. Udjeli pojedinih modova EOF dekompozicije u vremenskom nizu  $Q$  (3.1). Prikazani su parametri  $A_s$ ,  $R_s$  i  $NR_s$  dobiveni primjenom RUT, RAT i M98 metoda na statističke momente prvog reda (niz srednjaka) te parametri  $A_v$ ,  $R_v$  i  $NR_v$  dobiveni primjenom RUT, RAT i M98 metoda na statističke momente drugog reda (niz varijanci).

Vratimo se na početni niz  $Q$  i njegovu EOF dekompoziciju. Razmotrimo li udjele varijabilnosti koje pojedini mod opisuje u početnom nizu (slika 4.4), primjećuje se da prvi mod opisuje najviše varijabilnosti početnog niza te monotono opadanje udjela s rednim brojem moda. Na prvi mod otpada  $\sim 4.39\%$  a na trideseti  $\sim 2.47\%$  opisane varijabilnosti. Također, na slici su prikazani parametri  $A_s$ ,  $R_s$  i  $NR_s$  dobiveni primjenom RUT, RAT i M98 metoda na momente prvog reda (niz srednjaka) te parametri  $A_v$ ,  $R_v$  i  $NR_v$  dobiveni primjenom RUT, RAT i M98 metoda na momente drugog reda (niz varijanci). Naravno, kako je niz bijelog šuma konstantne varijance i srednjaka,

usporedbom parametara s kritičnim vrijednostima danim u tablici 2.1., vidimo da te metode nedvojbeno govore da je niz  $Q$  stacionaran. Mali udjeli pojedinih modova u nizu nasumičnih podataka su očekivani. Naime, u nasumičnim podacima EOF analiza ne može izdvojiti nikakve motive zajedničke svim podnizovima. Najbolji pokušaj izdvajanja varijabilnosti je upravo svojevrсни nasumični niz (što je vidljivo u  $PC_v$ -ovima). Kako takav nasumični niz može loše rekonstruirati početni niz, EOF analiza njime nastoji rekonstruirati barem onaj podniz koji posjeduje najveću varijancu (što je upravo podniz s najvećim odstupanjem od srednjaka). Modovi konstruirani iz takvog  $PC_v$ -a i  $EOF_v$ -a očekivano imaju vrlo mali udio u početnom nizu. No, ipak ne možemo biti sigurni da EOF analiza doista nije uhvatila neki nama neprimjetan obrazac u podacima, jer da nije, mogli bismo očekivati da je svaki mod (uniformno) zastupljen sa  $\frac{100\%}{30}$ , a vidimo da je prvi mod značajno više ( $\sim 1.75$  puta više) zastupljen od zadnjeg. U poglavlju 2.5, rečeno je da se udjeli pojedinog moda u početnom nizu računaju iz dobivenih svojstvenih vrijednosti matrice kovarijanci. Udio moda je direktno vezan s pripadnom svojstvenom vrijednosti  $EOF_v$ -a iz kojeg je nastao.



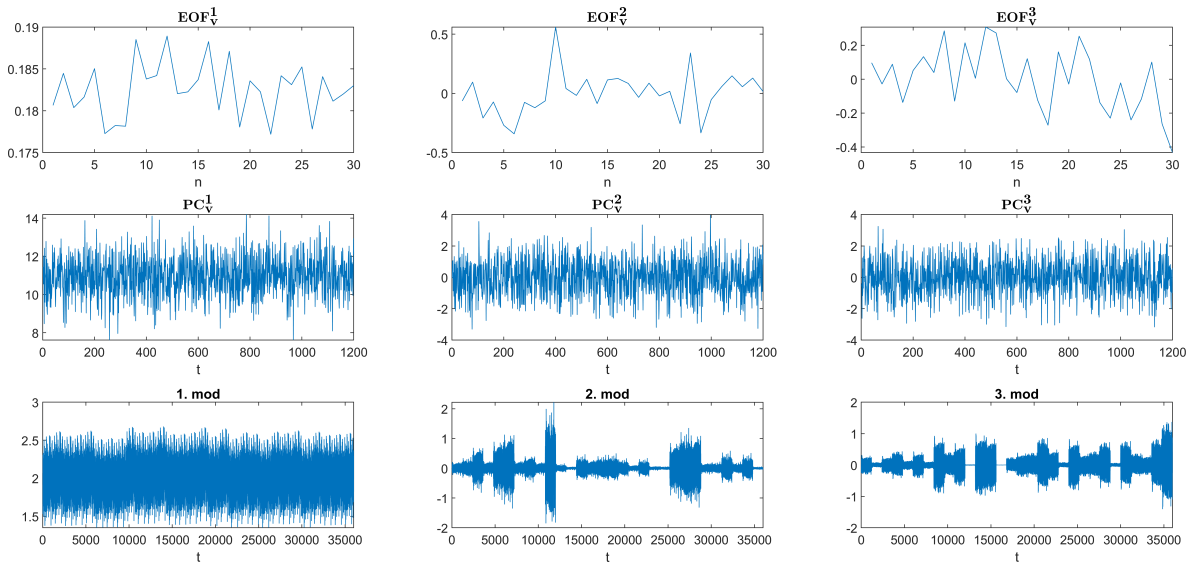
Slika 4.5. Crvena linija - Marčenko–Pastur raspodjela svojstvenih vrijednosti matrice kovarijanci slučajne matrice. Raspodjela je zadana jednađbom 4.1, gdje je  $\gamma = 40$ . Plavo - razdioba svojstvenih vrijednosti matrice kovarijanci niza  $Q$ .

Problemom svojstvenih vrijednosti matrica kovarijance nastalih iz matrica nasumičnih brojeva bavi se teorija slučajnih matrica. To je vrlo zahtjevno područje matematike te daleko premašuje obujam ovog rada, pa ćemo ovdje posegnuti samo za nekoliko bitnijih zaključaka. Jedan od eksplicitnih rezultata koje nam pruža teorija slučajnih matrica je upravo asimptotska raspodjela svojstvenih vrijednosti matrica kovarijanci dimenzija  $N \times N$ , nastalih iz slučajne matrice dimenzija  $T \times N$  (što odgovara našem slučaju i matrici podataka dobivenoj iz  $Q$  (poglavlje 2.5)). Uz pretpostavku da se radi

o velikim dimenzijama slučajne matrice (što je ovdje razumno zadovoljeno) i da vrijedi  $\frac{T}{N} \rightarrow \gamma$ , što je u našem slučaju  $\gamma = \frac{1200}{30} = 40$ , tada raspodjela matrice kovarijanci teži u Marčenko–Pastur raspodjelu (Bejan (2005) str. 28) koja glasi:

$$G(x) = \frac{\gamma}{2\pi x} \sqrt{(b-x)(x-a)}, a \leq x \leq b, \quad (4.1.)$$

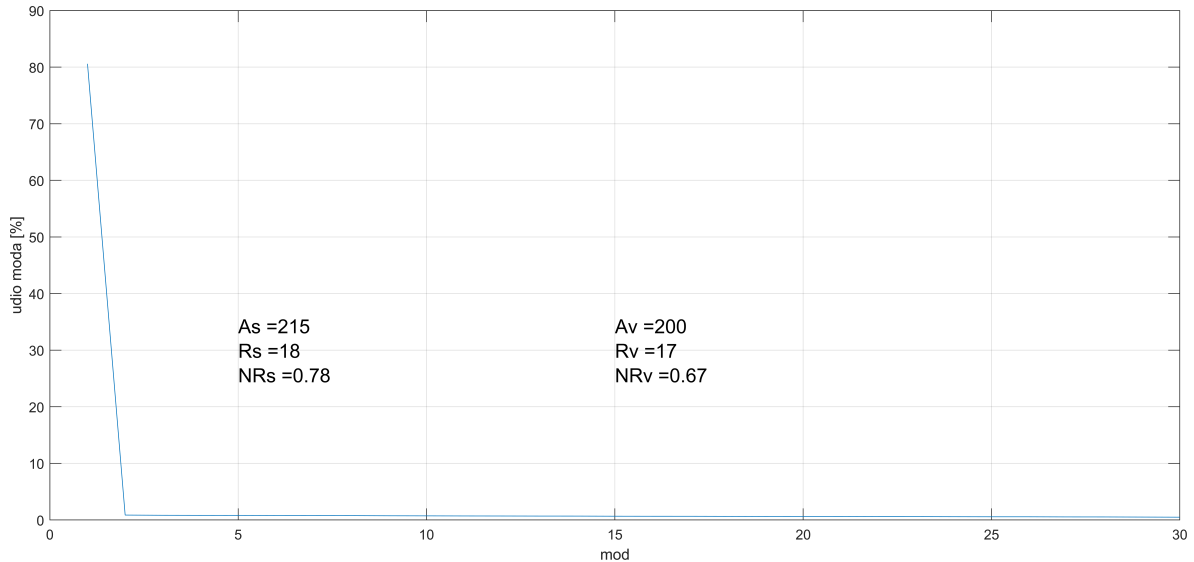
gdje je  $a = (1 - \gamma^{-1/2})^2$ ,  $b = (1 + \gamma^{-1/2})^2$ . Funkcija raspodjele  $G(x)$  te normalizirani histogram svojstvenih vrijednosti dobivenih iz matrice kovarijanci niza  $Q$  prikazani su na slici 4.5. Bez statističkog testiranja može se reći da dobivena raspodjela svojstvenih vrijednosti uistinu dovoljno dobro prati teorijsku raspodjelu. Vidimo da doista postoji "prirodno" raspršenje svojstvenih vrijednosti, od najmanje teorijske vrijednosti od  $a \approx 0.71$  do maksimalne teorijske vrijednosti od  $b \approx 1.34$ . Ono je odgovorno za postojanje dominantnog moda čak i u nizu bijelog šuma. Kako je udio najmanjeg moda proporcionalan najmanjoj svojstvenoj vrijednosti, a udio najvećeg moda proporcionalan najvećoj svojstvenoj vrijednosti, tada bi njihov omjer  $\frac{2.47\%}{4.39\%} \approx 0.563$  trebao biti približno jednak omjeru najmanje i najveće teorijske svojstvene vrijednosti  $\frac{0.71}{1.34} \approx 0.530$ , što vidimo da i jest slučaj. Može se zaključiti da različita zastupljenost modova u nizu  $Q$ , prikazana na slici 4.4, nije posljedica neke prepoznate pravilnosti od strane EOF analize.



Slika 4.6. Gornji red prikazuje redom prva tri  $EOF_v$ -a, srednji red redom prva tri  $PC_v$ -a dok donji red prva 3 moda EOF dekompozicije niza  $Q_2$  zadanog izrazom 3.2 provedene na način opisan u poglavlju 2.5.

Dalje se nameće pitanje, što bi nam dala EOF analiza ukoliko bi početni niz imao srednjak različit od nule. Kako bi smo pokušali odgovoriti na to pitanje, na prethodan niz dodajemo proizvoljnu konstantu  $\mu = 2$  čime tvorimo novi niz  $Q_2$  dan jednadžbom 3.3. Prva tri  $EOF_v$ -a,  $PC_v$ -a niza  $Q_2$  i pripadni im modovi prikazani su na slici 4.6. Razlika s obzirom na prethodni slučaj je odmah vidljiva. Prvi mod nastoji opisati novu

srednju vrijednost dok drugi mod u grubo odgovara prvom modu rastava niza  $Q$ . Treći mod odgovara prijašnjem drugom modu rastava niza  $Q$ . Na slici 4.7 uočljivo je da novi prvi mod objašnjava nešto više od 80% početnog niza, dok na drugi mod otpada tek 0.87% a na trideseti mod 0.49%. Zaključujemo da srednja vrijednost, ako postoji, dominira prvim modom. Variranjem  $\mu$ -a može se primijetiti da udio prvog moda raste

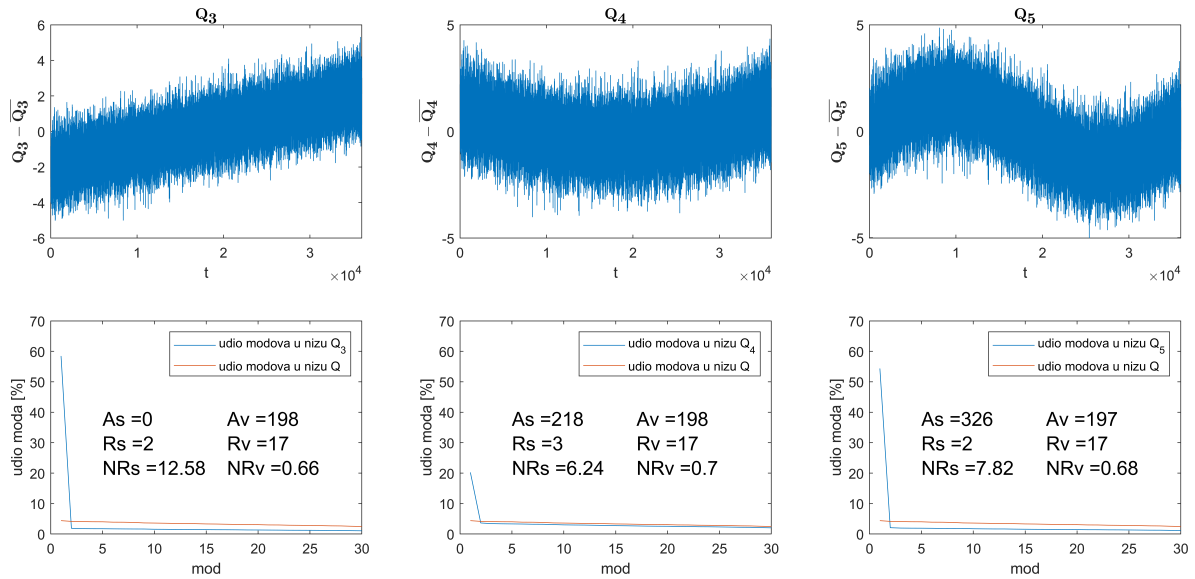


Slika 4.7. Udjeli pojedinih modova EOF dekompozicije u vremenskom nizu  $Q_2$  danog izrazom 3.2. Prikazani su parametri  $A_s$ ,  $R_s$  i  $NR_s$  dobiveni primjenom RUT, RAT i M98 metoda na statističke momente prvog reda (niz srednjaka) te parametri  $A_v$ ,  $R_v$  i  $NR_v$  dobiveni primjenom RUT, RAT i M98 metoda na statističke momente drugog reda (niz varijanci).

porastom omjera  $\mu/\sigma$  i da postane primjetan tek kada  $\mu/\sigma > 0.2$  (nije prikazano). Dakle, možemo reći da je EOF analiza relativno osjetljiva na postojanje konstantnog srednjaka. Na slici 4.7 vidimo parametre RUT, RAT i M98 metoda za srednju vrijednost i varijancu koji se ne razlikuju za nizove sa  $\mu = 2$  i  $\mu = 0$ . Kako postojanje srednje vrijednosti različite od 0 ne utječe na njegovu stacionarnost, a EOF analiza značajno reagira na nju, u svim budućim eksperimentima od nizova će biti oduzeta njihova srednja vrijednost prije nego na njih primijenimo EOF analizu.

Idući korak je razmatranje nizova nestacionarnog srednjaka i stacionarne varijance. Do sada smo promatrali nizove tipa a) sa slike 2.1, a sada nastavljamo s nizovima tipa b). Promotrit ćemo 3 takva niza - s linearnim trendom  $Q_3$  (3.3), parabolničnim hodom  $Q_4$  (jednadžba 3.4) te sinusoidalnim hodom  $Q_5$  (3.5), superponiranim na isti niz bijelog šuma  $Q$ . Nizovi su dani izrazima 3.3, 3.4, 3.5 respektivno. Prikazani su na slici 4.8 zajedno s udjelima pojedinih modova. Na slikama u donjem redu, prikazane su raspodjele udjela pojedinih modova (plave linije) u usporedbi s razdiobom udjela modova u nizu  $Q$  (crvene linije). Također prikazani su i parametri RUT, RAT i M98 metoda za hod srednjaka i hod varijance pripadnih nizova. Udjele modova u nizovima  $Q_3$ ,  $Q_4$  i  $Q_5$  ima smisla uspoređivati s udjelom modova u nizu  $Q$ , jer su oni nastali

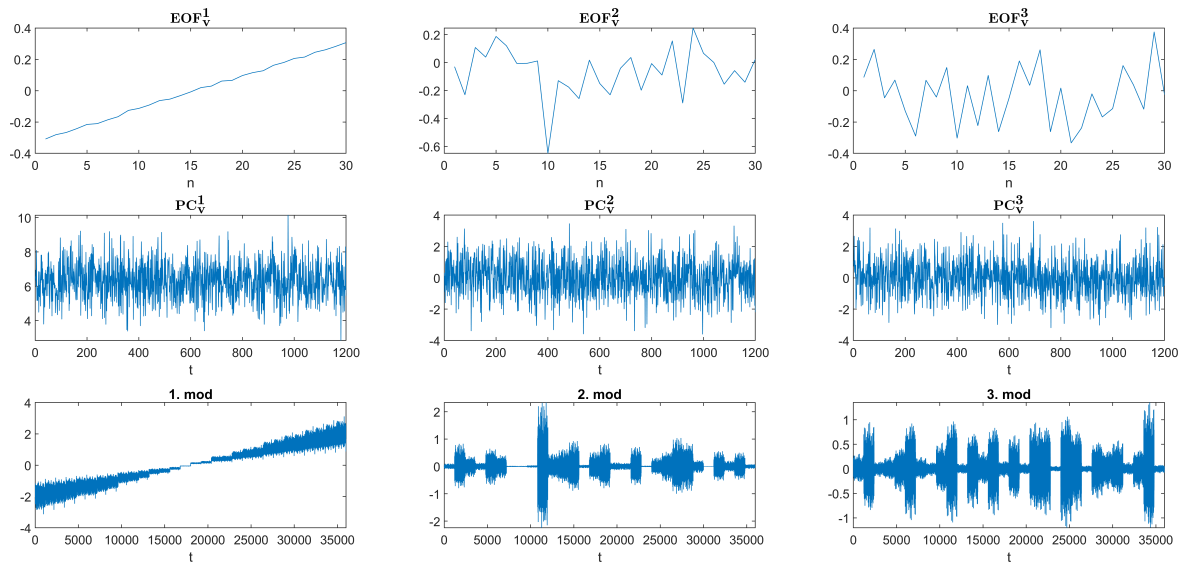




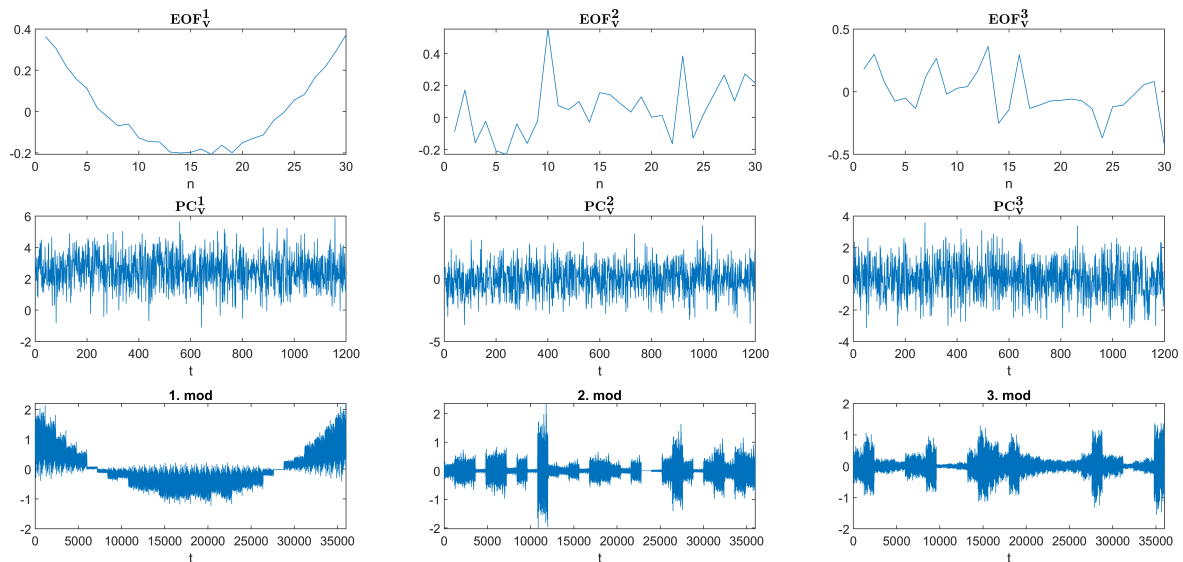
Slika 4.8. Gornji red - nizovi  $Q_3$ ,  $Q_4$  i  $Q_5$  dani izrazima 3.3, 3.4, 3.5 respektivno. Donji red - udjeli pojedinih modova EOF dekompozicije pripadnih nizova (plava linija) u usporedbi sa udjelom pojedinih modova u nizu  $Q$  (crvena linija). Prikazani su parametri  $A_s$ ,  $R_s$  i  $NR_s$  dobiveni primjenom RUT, RAT i M98 metoda na statističke momente prvog reda (niz srednjaka) te parametri  $A_v$ ,  $R_v$  i  $NR_v$  dobiveni primjenom RUT, RAT i M98 metoda na statističke momente drugog reda (niz varijanci) pripadnih nizova.

superponiranjem različitih hodova upravo na niz  $Q$ . Vidljivo je da su hodovi srednjaka nizova  $Q_3$  i  $Q_5$  (relacije 3.3. i 3.5.) proglašeni nestacionarnima jednoglasno od strane sve tri metode (prema  $A_s$ ,  $R_s$ ,  $NR_s$ ), dok su hodovi varijance proglašeni stacionarnima (prema  $A_v$ ,  $R_v$ ,  $NR_v$ ). Za hod srednjaka niza  $Q_4$  (relacija 3.4.) RAT metoda ( $A_s$ ) nije uspjela zaključiti da je nestacionaran, dok su RUT i M98 ( $R_s$ ,  $NR_s$ ) uspješno došle do tog zaključka.

Stacionarnost hoda varijance niza  $Q_4$  prepoznata je od strane sve tri metode. Nemogućnost metode RAT da prepozna nestacionarnost hoda srednjaka niza  $Q_4$  nije iznenađujuće ako imamo u vidu način računanja parametra  $A$  (poglavlje 2.2 - stacionarnost za RAT metodu proizlazi iz simetričnosti hoda). Nadalje, može se primijetiti kako je u sva tri slučaja EOF analiza zamijetila postojanje trenda na način da je povećala udio prvog moda i smanjila udio svih preostalih modova (kako bi zbroj ostao 100%). Stoga je za očekivati da će prvi mod nastojati uhvatiti trend, kao što je u prethodnom primjeru prvi mod nastojao uhvatiti srednju vrijednost niza. Da je to zaista tako, vidljivo je na slikama 4.9, 4.10 i 4.11. U raspodjeli udjela modova u sva tri niza, udio prvog moda je znatno veći od udjela ostalih modova. No, udjeli prvog moda se značajno razlikuju između nizova. Kako je između nizova jedina razlika u obliku superponiranog hoda srednjaka, to je nužno čini uzrokom razlika u udjelima prvog moda. Može se primijetiti da udio ovisi o rasponu hoda (razlici maksimalne i minimalne vrijednosti koju postiže) te o varijanci niza.

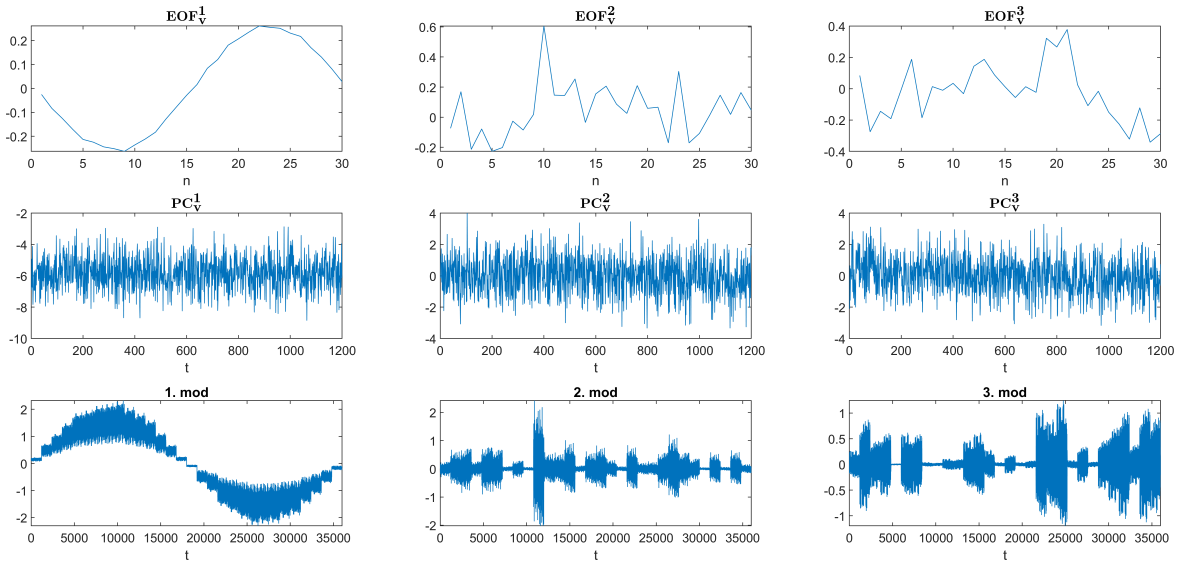


Slika 4.9. Kao na slici 4.6, no za niz  $Q_3$  dan izrazom 3.3.



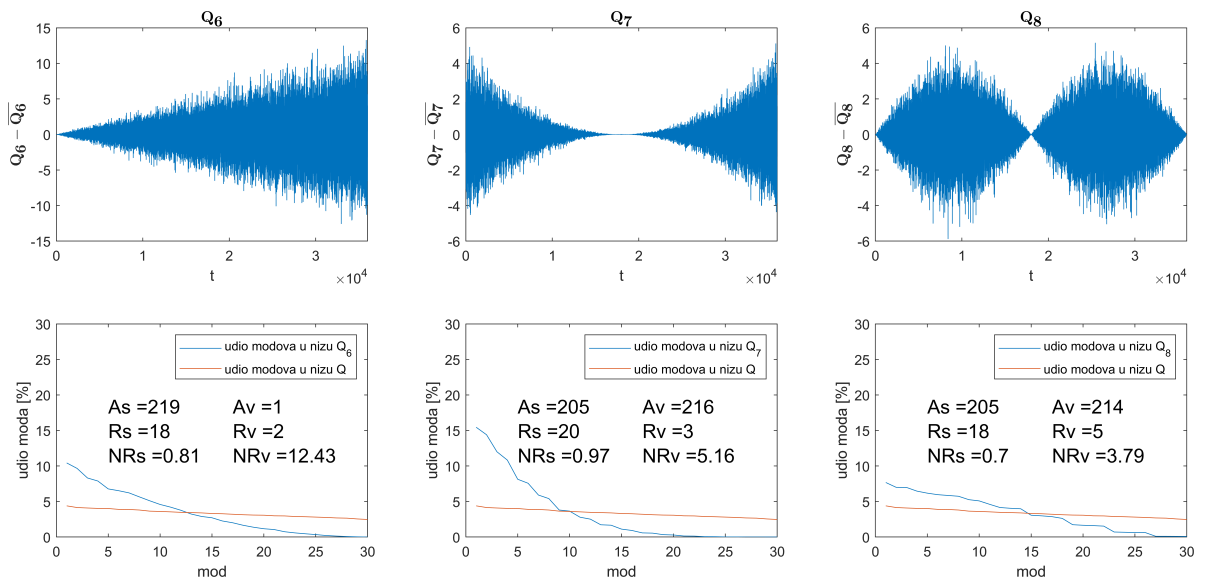
Slika 4.10. Kao na slici 4.6, no za niz  $Q_4$  dan izrazom 3.4.

Zanimljivo je uočiti kako  $EOF_v$ -ovi prvih modova sva tri niza ocrtavaju upravo oblik njihova hoda (do na predznak). Svaka od 30 vrijednosti prvog  $EOF_v$ -a su zapravo normirani srednjaci 30 podnizova na koje rastavljamo početni niz.  $PC_v$ -ovi i dalje zadržavaju oblik nasumičnog niza, no PC 1 sada ima srednjak različit od nule kako bi pri množenju s članovima EOF 1 mogao postići oblik približan hodu niza. Ostali PC-ovi su nasumični nizovi sa srednjakom 0. Nakon što je prvi mod pokupio informaciju o hodu niza (i manji dio njegove nasumičnosti), na ostalim je modovima da objasne preostali nasumični dio niza. Stoga, nije iznenađujuće da drugi mod u sva tri slučaja, jako podsjeća na prvi mod originalnog nasumičnog niza  $Q$ . S obzirom na to da je dio nasumičnosti ostao u prvom modu (prvi mod nije samo čisti hod), ne čudi da se treći modovi poprilično razlikuju međusobno kao i od drugog moda nasumičnog niza. To



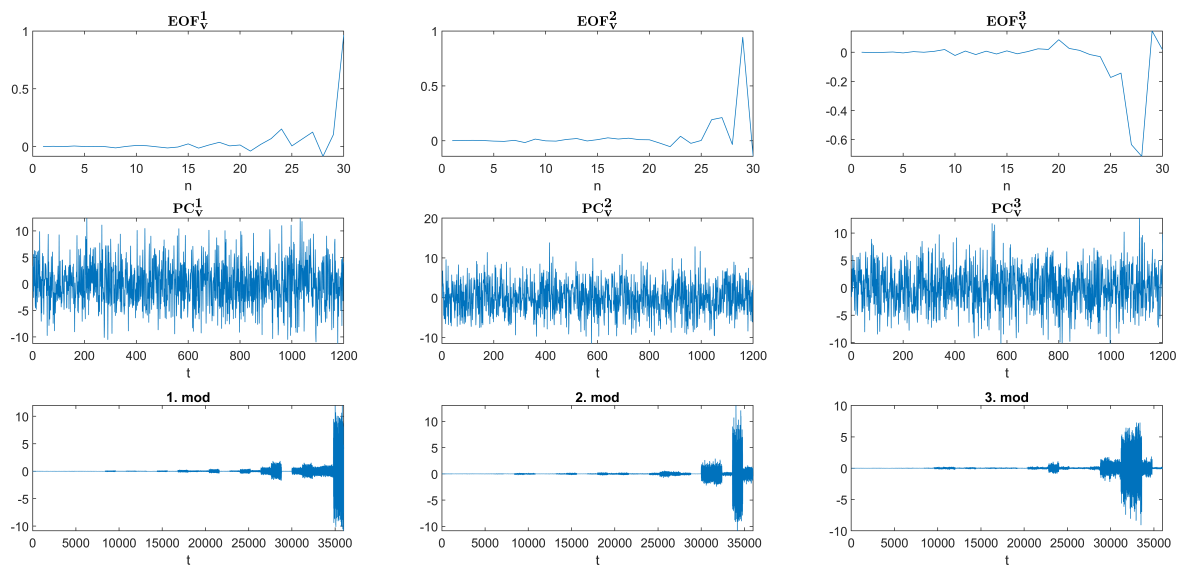
Slika 4.11. Kao na slici 4.6, no za niz  $Q_5$  dan izrazom 3.5.

nam govori da oduzimanjem pripadnih prvih modova od nizova  $Q_3$ ,  $Q_4$  i  $Q_5$ , dobivamo novi niz koji barem približno odgovara nizu  $Q$ . Ponovnom primjenom EOF analize na takve nizove lako je pokazati da udjeli modova u tako konstruiranim novim nizovima odgovaraju upravo stanju prikazanom na slici 4.4 što potvrđuje da su dobiveni nizovi upravo  $Q$ . Do istog zaključka smo mogli doći i bez ponovnog provođenja EOF analize. Kako su modovi EOF dekompozicije ortogonalni, tada iz jednom dobivene dekompozicije prilikom razmatranja možemo samo izbaciti prvi mod, pripadni  $EOF_v$ ,  $PC_v$  te svojstvenu vrijednost i time dobivamo upravo situaciju koju bi dobili primjenom EOF analize na niz  $Q$  ili niz  $(Q_3/Q_4/Q_5 - 1. \text{ mod})$ .

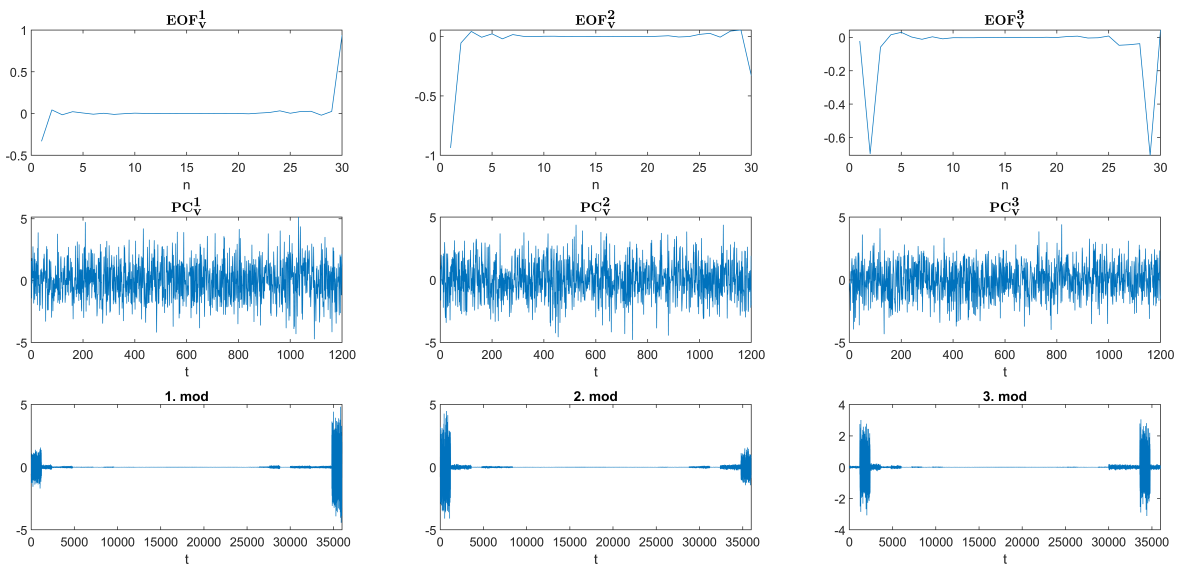


Slika 4.12. Kao na slici 4.8, no za nizove  $Q_6$ ,  $Q_7$  i  $Q_8$ , dane izrazima 3.6, 3.7, 3.8 respektivno.

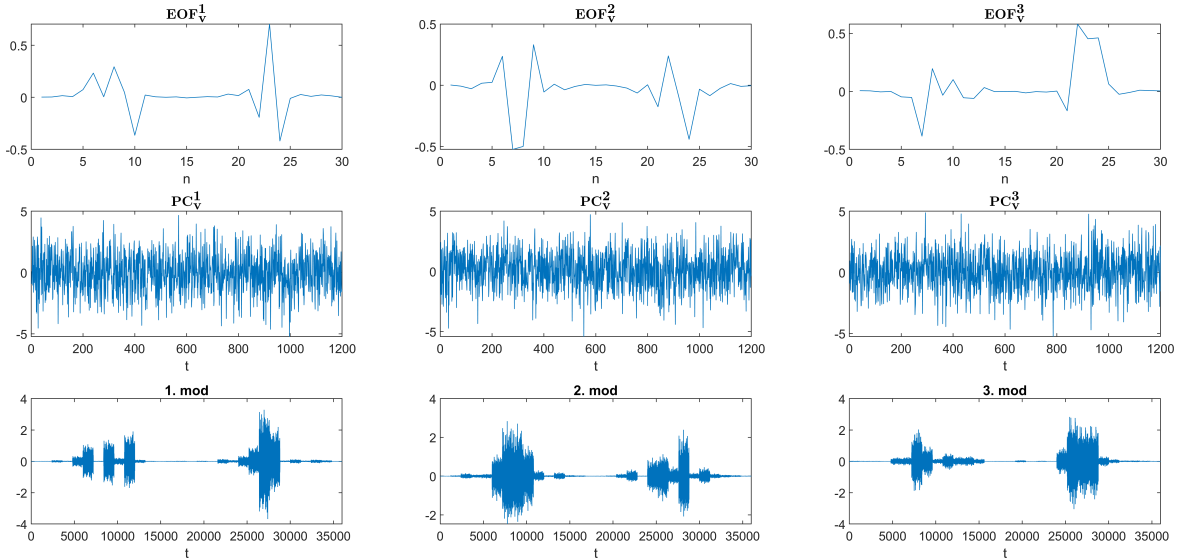
U idućem koraku promatramo nizove tipa c) sa slike 2.1. Takvi nizovi konstruirani su u poglavlju 3 te su dani izrazima 3.6, 3.7, 3.8 ( $Q_6$ ,  $Q_7$ ,  $Q_8$ ). To su nizovi kojima varijanca posjeduje hod u vremenu ali im je srednjak konstantan. Nizovi  $Q_6$ ,  $Q_7$  i  $Q_8$  prikazani su na slici 4.12. Prema parametrima metoda RUT, RAT i M98 vidimo da su sve tri, za hodove sva tri niza, ustvrdile da su stacionarni. Metoda M98 je za varijance sva tri niza zaključila da su nestacionarne, kao i metoda RUT. RAT metoda, s druge strane, je za varijance nizova  $Q_7$  i  $Q_8$  pokazala da su stacionarne (zbog simetričnosti). Udjeli modova EOF dekompozicije se značajno razlikuju za sva tri niza od udjela modova niza  $Q$ . Za razliku od iskakanja samo prvog moda, kao kod nizova s nestacionarnim prvim momentom, ovdje je slučaj da više modova u prvoj polovici naraste na korist onih u drugoj polovici. Razlika između udjela prvog i drugog moda



Slika 4.13. Kao na slici 4.6, no za niz  $Q_6$  dan izrazom 3.6.



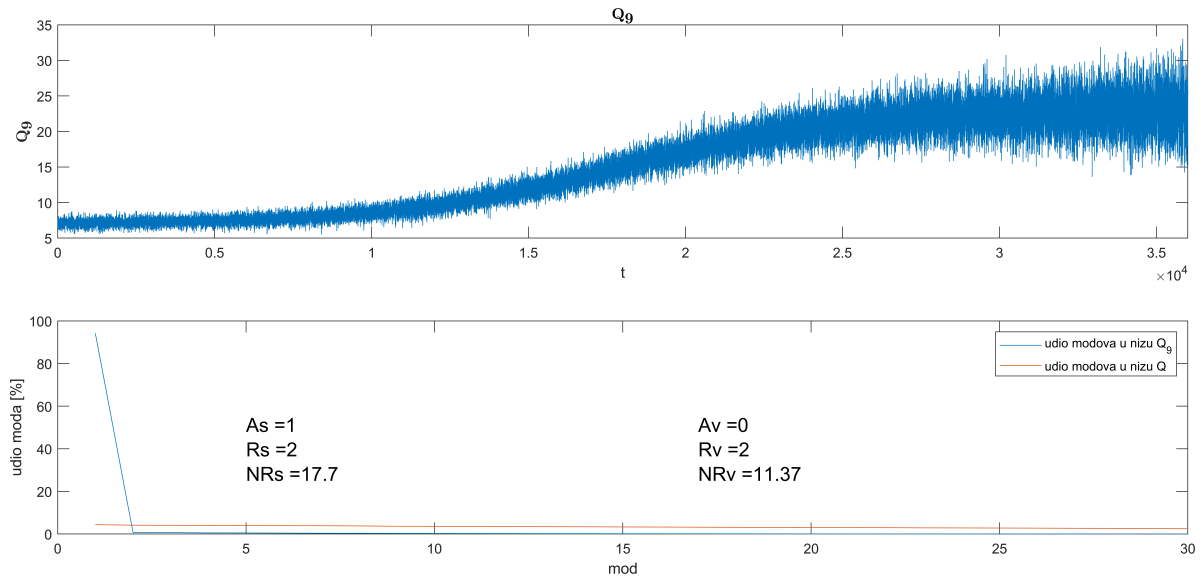
Slika 4.14. Kao na slici 4.6, no za niz  $Q_7$  dan izrazom 3.7.



Slika 4.15. Kao na slici 4.6, no za niz  $Q_8$  dan izrazom 3.8.

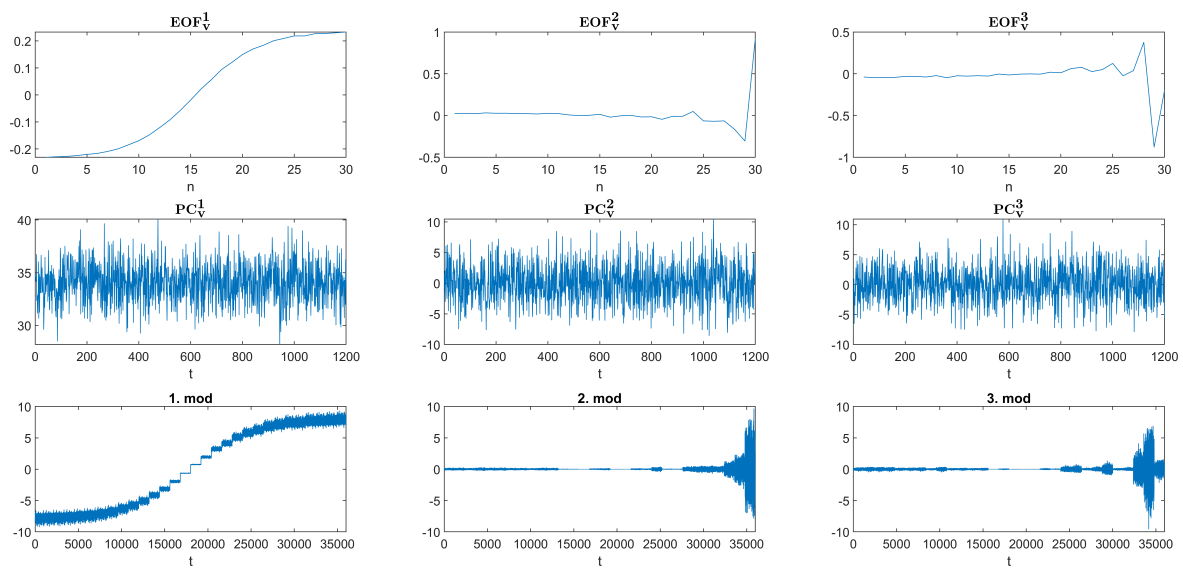
nije toliko drastična. Sa slika 4.13, 4.14 i 4.15 vidljivo je kako prvi modovi nastoje opisati one podnizove početnog niza s najvećom varijancom. Drugi modovi nastoje opisati podnizove s drugom po redu najvećom varijancom itd. Možemo reći da se promjena varijance niza u vremenu očituje povećanim udjelom prve polovice modova (smanjenim udjelom druge polovice modova). Nizom različitih eksperimenata, za različito varirajuće varijance, može se pokazati da, kako bi EOF analiza zamijetila nestacionarnost varijance (zamijetila = udio prvog moda narastao za minimalno 1% s obzirom na udio prvog moda u nizu  $Q$ ), dio niza s minimalnom i maksimalnom standardnom devijacijom mora zadovoljavati  $\sigma_{min}/\sigma_{max} < 0.7$  (nije prikazano). Za  $EOF_v^1$  je vidljivo da posjeduje skok u onoj  $x$  koordinati koja odgovara rednom broju podniza s najvećom varijancom.  $EOF_v^1$  "locira" podniz sa najvećom varijancom,  $EOF_v^2$  sa drugom po redu najvećom varijancom itd. Za nizove s hodom srednjaka te konstantnom varijancom, udio prvog moda je odskočio od svih ostalih, no razlike između ostalih modova nisu bile značajne (udjeli preostalih modova su padali gotovo linearno). Za nizove s konstantnim hodom te promjenjivom varijancom, iako je prva polovica modova porasla, njihovo opadanje bliže prati eksponencijalnu funkciju i to tako da na zadnji (ili zadnjih nekoliko) mod otpada zanemariv udio od  $< 0.01\%$ .

Preostalo je za razmotriti još samo slučaj kada niz posjeduje hod i promjenjivu varijancu (d) slučaj na slici 2.1). Kao primjer takvog niza konstruirat ćemo niz koji podsjeća na čestu stvarnu, neporemećenu situaciju brzine vjetera pred zorom. Nakon stabilne atmosfere tijekom noći, brzine vjetera su relativno male, često s malom varijancom, no u trenutku kada površinu ranom zorom obasja sunce, ona se zagrijava te se povećava nestabilnost, što dovodi do većeg miješanja zraka te brzine vjetera u prosjeku rastu, no raste im i varijanca (npr. Stull, 1988). Niz  $Q_9$  (3.9) skicira takvu situaciju.



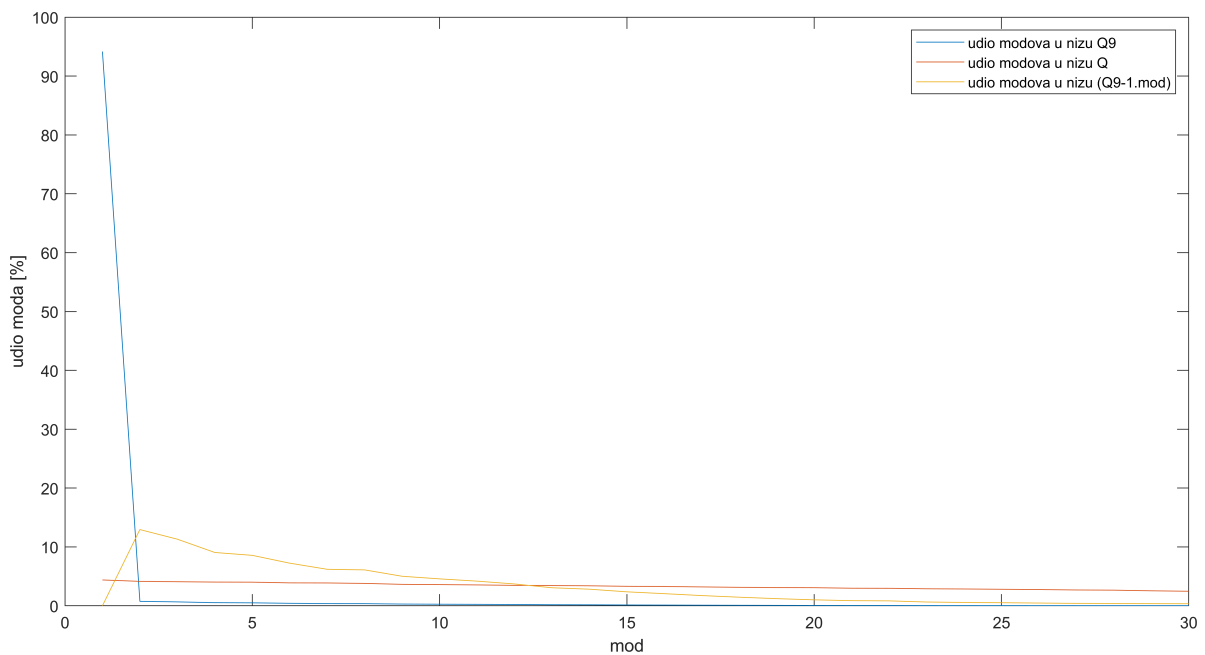
Slika 4.16. Gore - niz  $Q_9$  dan izrazom 3.9. Dolje - udjeli pojedinih modova EOF dekompozicije u vremenskom nizu  $Q_9$  (plava linija) i udjeli pojedinih modova EOF dekompozicije u vremenskom nizu  $Q$  (crvena linija). Prikazani su parametri  $A_s$ ,  $R_s$  i  $NR_s$  dobiveni primjenom RUT, RAT i M98 metoda na statističke momente prvog reda (niz srednjaka) te parametri  $A_v$ ,  $R_v$  i  $NR_v$  dobiveni primjenom RUT, RAT i M98 metoda na statističke momente drugog reda (niz varijanci).

Napomena: prije EOF analize, od niza  $Q_9$  oduzeta je srednja vrijednost. Na slici 4.16 prikazan je niz  $Q_9$  s udjelima pojedinih modova i parametrima RAT, RUT i M98 metoda primijenjenih na hod srednjaka i hod varijance. Sve tri metode su jednoglasne i pokazuju da su momenti prvog i drugog reda nestacionarni. Primjećujemo da je prvi mod u EOF dekompoziciji odskočio na  $\sim 94.16\%$  dok je drugi tek  $\sim 0.76\%$  a trideseti  $\sim 0.002\%$  udjela u nizu  $Q_9$ . Iz odskakanja prvog moda može se zaključiti da je niz



Slika 4.17. Kao na slici 4.6, no za niz  $Q_9$  dan izrazom 3.9.

nestacionaran zbog postojanja hoda, što znamo da je doista istina. Sa slike 4.17 je vidljivo kako EOF 1 opisuje upravo funkciju tangens hiperbolni koji i jest hod niza  $Q_9$ . Ukoliko iz razmatranja izbacimo prvi mod (što zbog nezavisnosti modova možemo učiniti), za udjele modova u preostalom nizu dobivamo raspodjelu na slici 4.18. Vidimo da nova raspodjela udjela modova, nakon što smo uklonili prvi mod, ne izgleda poput one od niza  $Q$  (slika 4.4). Nova raspodjela udjela nema približno linearno opadanje udjela, te je udio zadnjih nekoliko modova zanemariv. To ukazuje da niz ( $Q_9 - 1.mod$ ) posjeduje nestacionarnu varijancu za što znamo da je točno od početka.



Slika 4.18. Plavo - udio modova u nizu  $Q_9$ , crveno - udio modova u nizu  $Q$ , narančasto - udio modova u nizu ( $Q_9 - 1.mod$ ).

Sumirano, može se reći da EOF analiza nestacionarnost niza koji posjeduje nestacionarni hod srednjaka a stacionarni hod varijance, prepoznaje po:

1. značajno izraženom prvom modu,
2. približno linearnom opadanju udjela drugog, pa sve do posljednjeg moda i to tako da je udio zadnjeg moda različit od 0% ,
3.  $PC_v$  prvog moda posjeduje srednjak različit od 0,
4.  $EOF_v$  prvog moda ocrtava oblik hoda,
5. prvi mod opisuje hod vremenskog niza.

EOF analiza nestacionarnost niza koji posjeduje stacionarni hod srednjak te nestacionarni hod varijance, prepoznaje po:

1. značajno izraženoj prvoj polovici modova (udio prvog i drugog moda se ne razlikuje drastično),
2. nelinearnom opadanju udjela modova, tako da je udio zadnjeg moda gotovo 0%,
3.  $EOF_v$  prvog moda locira podniz/podnizove s najvećom varijancom tako da je u većini točaka  $\sim 0$ , osim u njih nekoliko, gdje je gotovo 1,
4. prvi mod opisuje podniz/podnizove koji posjeduju najveću varijancu.

EOF analiza nestacionarnost niza koji posjeduje nestacionarni hod srednjak te nestacionarni hod varijance, prepoznaje po:

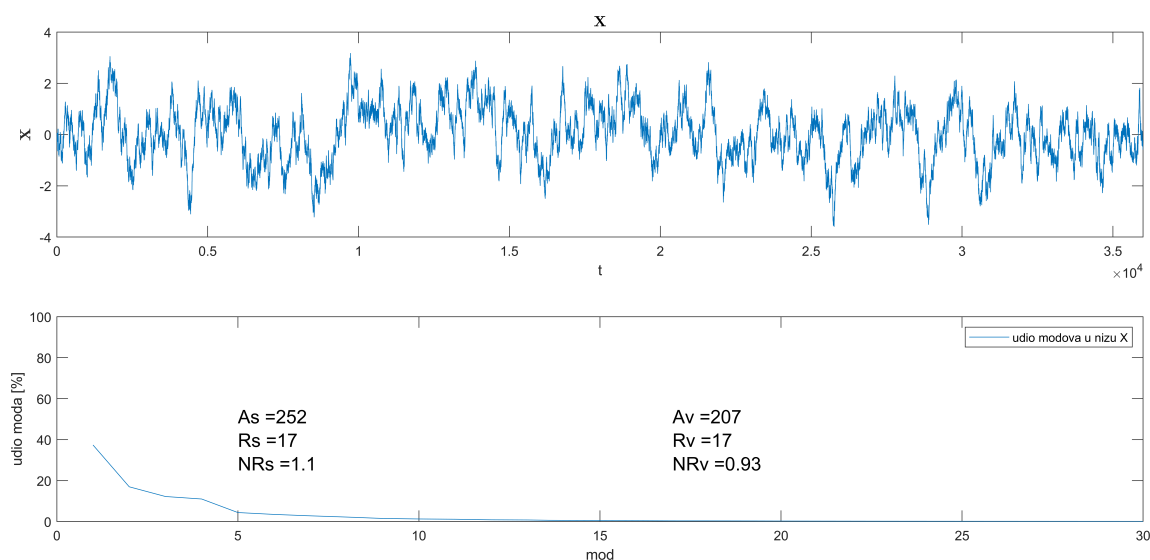
1. značajno izraženom prvom modu,
2. nelinearnom opadanju udjela drugog, pa sve do posljednjeg moda, tako da je udio zadnjeg moda gotovo 0%,
3.  $EOF_v$  prvog moda ocrtava oblik hoda,
4.  $EOF_v$  drugog moda locira podniz/podnizove s najvećom varijancom tako da je u većini točaka  $\sim 0$ , osim u njih nekoliko, gdje je gotovo 1,
5.  $PC_v$  prvog moda posjeduje srednjak različit od 0,
6. prvi mod opisuje hod vremenskog niza.
7. drugi mod opisuje podniz/podnizove koji posjeduju najveću varijancu.



## 4.2. Primjena EOF analize na nizove izvedene iz AR(1) procesa

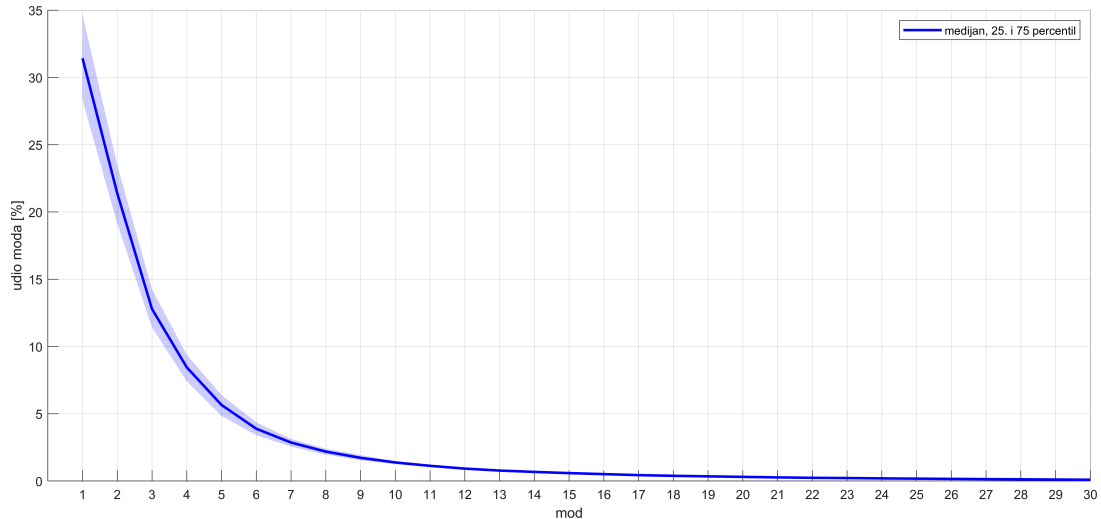
Vremenski nizovi neke meteorološke veličine, nisu posljedica vremenski nezavisnih procesa konstantne, normalne raspodjele, ili procesa izvedenih iz njih. Prirodni procesi su gotovo uvijek autokorelirani u vremenu. To uvodi kompliciraniju strukturu u vremenski niz od one koju smo razmatrali u prošlom poglavlju. Kako smo do sada EOF analizu razmatrali samo na nizovima, izvedenima iz normalno raspodjeljenih slučajnih varijabli, ne možemo biti sigurni da su nam zaključci s kraja prošlog poglavlja doista primjenjivi u realnim vremenskim nizovima.

Kako bi se približili jedan korak bliže realnim procesima uvodimo autokoreliranost. Iako postoji cijeli niz stohastičkih procesa kojima se mogu simulirati takvi vremenski nizovi, ovdje će biti korišten proces autoregresije opisan u poglavlju 2.6. Osim zbog svoje jednostavnosti, procesi autoregresije su pogodni jer se njima mogu simulirati nizovi koji imaju dugo pamćenje. Večenaj i De Wekker (2015) zaključili su da unutar 30 minutnog vremenskog niza, podnizove od 1 minute možemo smatrati međusobno nezavisnima, što drugim riječima znači: podatci udaljeni više od 1 minute međusobno su nekorelirani. Dakle, u tom slučaju, može se govoriti da niz ima pamćenje dugo do 1 min, što upućuje na sporo padajuću autokorelacijsku funkciju. Prema Hyndman i Athanasopoulos (2018), str. 290 - 329, sporo padajuća autokorelacijska funkcija upućuje na pogodnost modeliranja nizova AR modelom. Konkretno, koristit će se AR(1) model dan jednadžbom 3.11. Za koeficijentom  $\phi$  odabiremo  $\phi = 0.995$ . To je pogodan izbor jer autokorelacijska funkcija AR(1) procesa opada s korakom  $\tau$  kao  $\rho(\tau) = \phi^\tau$  što znači da za  $\phi < 1$ ,  $\phi \approx 1$  opada vrlo sporo.



Slika 4.19. Gore - niz X zadan relacijom 3.10, dolje - udjeli pojedinih modova EOF dekompozicije u vremenskom nizu X. Prikazani su parametri  $A_s$ ,  $R_s$  i  $NR_s$  dobiveni primjenom RUT, RAT i M98 metoda na niz srednjaka te parametri  $A_v$ ,  $R_v$  i  $NR_v$  dobiveni primjenom RUT, RAT i M98 metoda na niz varijanci.

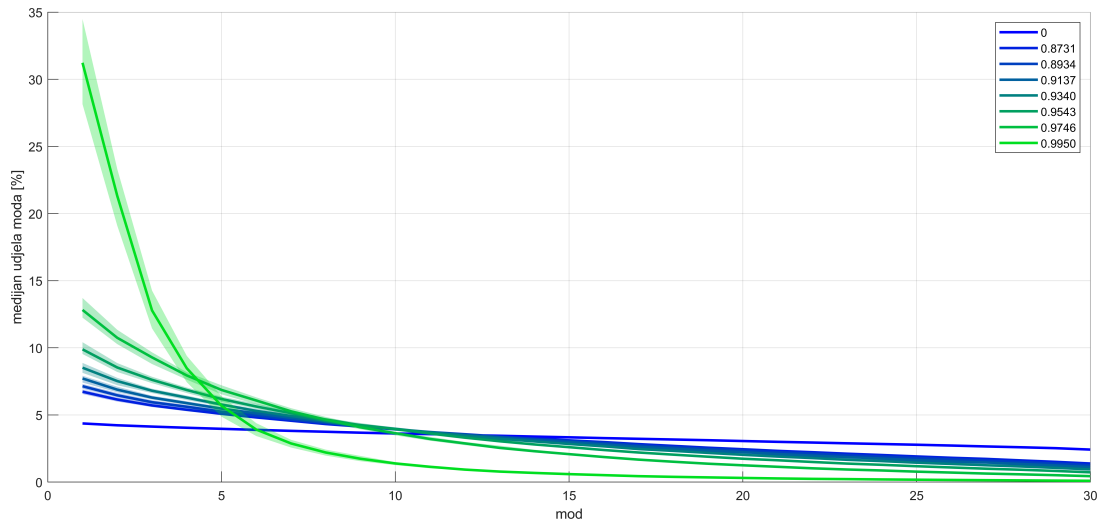
Niz  $X$  (izraz 3.10), s udjelom modova EOF dekompozicije te parametrima RUT, RAT i M98 metoda za hod srednjaka i varijance prikazan je na slici 4.19. Metode RUT, RAT i M98 jednoglasno tvrde da su hod srednjaka i varijance niza stacionarni. Vidljiva je značajna zastupljenost udjela prvih nekoliko modova EOF dekompozicije 1. mod  $\approx 37,4\%$ , 2. mod  $\approx 17\%$ , 3. mod  $\approx 12,2\%$  te nelinearno opadanje udjela sve do 30. mod  $\approx 0,085\%$ . Raspodjela udjela modova podsjeća na sliku 4.12 iz prošlog poglavlja gdje je niz imao nestacionarnu varijancu. Zbog toga bi se moglo, slijedeći zaključke s kraja prošlog poglavlja, pogrešno pripisati nizu  $X$  nestacionarnost



Slika 4.20. Medijan ansambla od 100 realizacija niza  $X$  zadanog jednačbom 3.10. Osjenčano je područje omeđeno 75. i 25. percentilom.

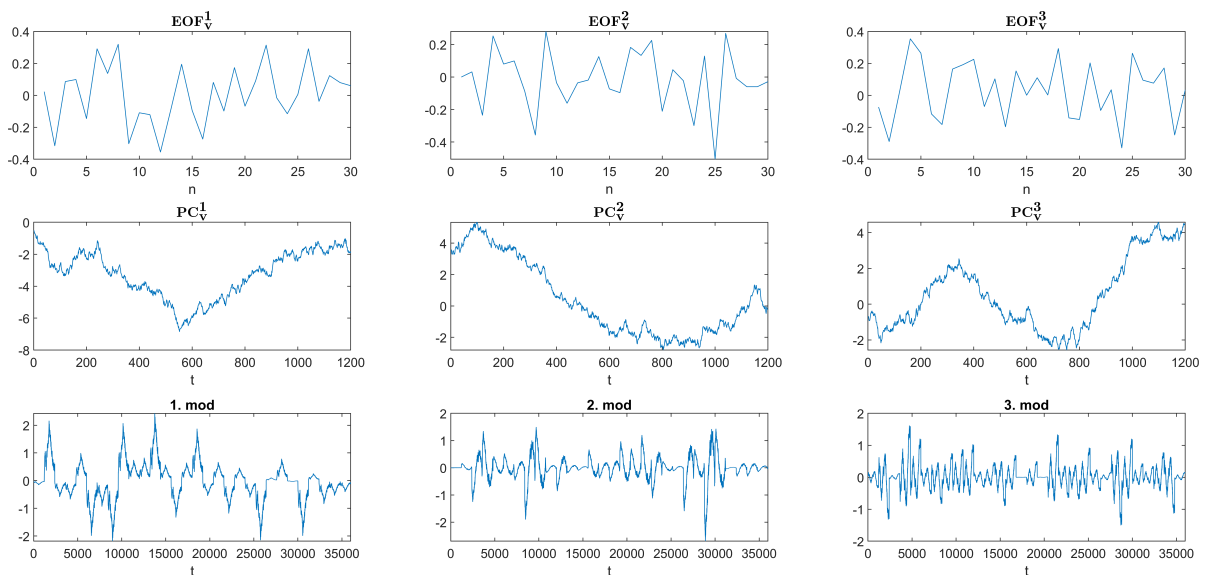
u varijanci. Moramo imati na umu da su zaključci s kraja prošlog poglavlja doneseni na temelju usporedbe EOF dekompozicije nizova  $Q_{2,\dots,9}$  s nizom  $Q$ , koji je služio kao svojevrsna null hipoteza stacionarnog niza. Ovdje se udaljavamo od toga tako što nam niz  $X$  (koji skicira stvarniju sliku nizova kakvi se uistinu pojavljuju prilikom mjerenja meteoroloških varijabli) predstavlja novu null hipotezu stacionarnog niza. Nizovi dugog pamćenja posjeduju određenu strukturu (npr. relativno duge vremenske intervale raste ili pada) za razliku od nizova bijelog šuma. Izgled te strukture snažno ovisi o  $\epsilon_t$  i to na način da dva generirana niza  $X$  na temelju vrlo sličnih  $\epsilon_t$  nemaju sličnosti. Sama struktura niza utječe, kako na parametre RUT, RAT i M98 metode, tako i na rezultate EOF analize. Stoga, umjesto da gledamo raspodjelu udjela modova jedne realizacije takvog procesa, gledamo razdiobu medijana udjela modova ansambla realizacija procesa  $X$ . Upravo to je prikazano na slici 4.20 zajedno s osjenčanim područjem u koje upada 50% vrijednosti udjela pojedinog moda, tj. 25. i 75. percentil.

Sa slike 4.20 je vidljivo da je medijan udjela pojedinih modova glatka eksponencijalno padajuća funkcija s padajućim raspršenjem koje je najveće za prvi mod. Jasno je da konkretni oblik razdiobe udjela modova mora ovisiti o odabiru parametra  $\phi$  u jednačbi 3.11, jer za  $\phi = 0$  niz  $X$  se svodi na niz  $Q$  iz prošlog poglavlja. Ovisnost



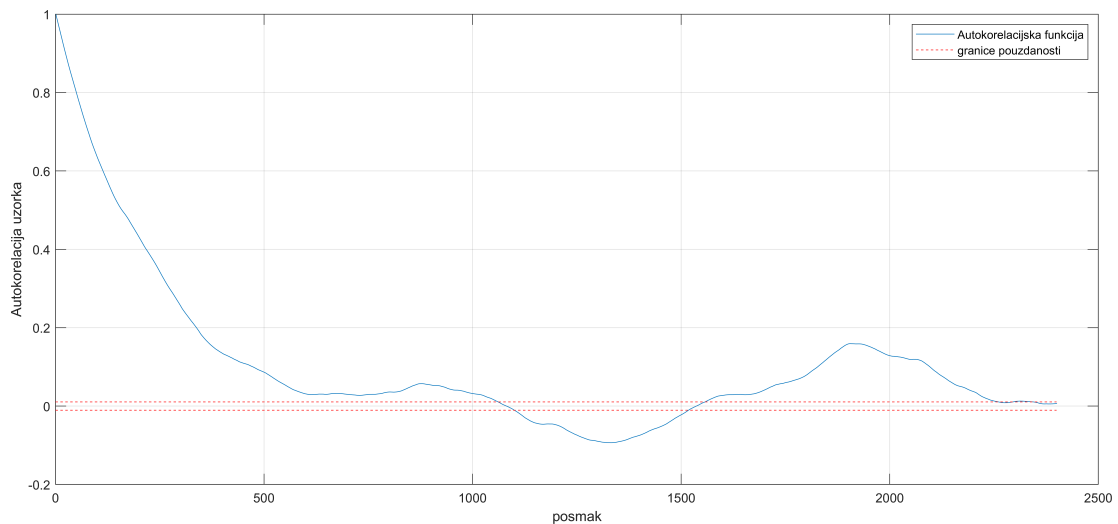
Slika 4.21. Medijan ansambla od 100 realizacija niza  $X$  zadanog jednačbom 3.10 za različite parametre  $\phi$ . Osjenčana su područja omeđena 75. i 25. percentilom.

medijana udjela modova ansambla nizova  $X$ , za različite parametre  $\phi$  prikazana je na slici 4.21. Može se primijetiti pravilno variranje razdiobe udjela modova u ovisnosti o  $\phi$ . U eksplicitnu formulaciju te ovisnosti nećemo ulaziti. Vidljivo je da razdioba raste u području nižih modova sve brže kako se parametar  $\phi$  približava 1. Primijetimo da nizovi s većim  $\phi$  također imaju i veće raspršenje. Dakle, izgled raspodjele udjela modova stacionarnog niza dobivenog AR metodom (null hipoteza) ovisi o odabiru  $\phi$ .



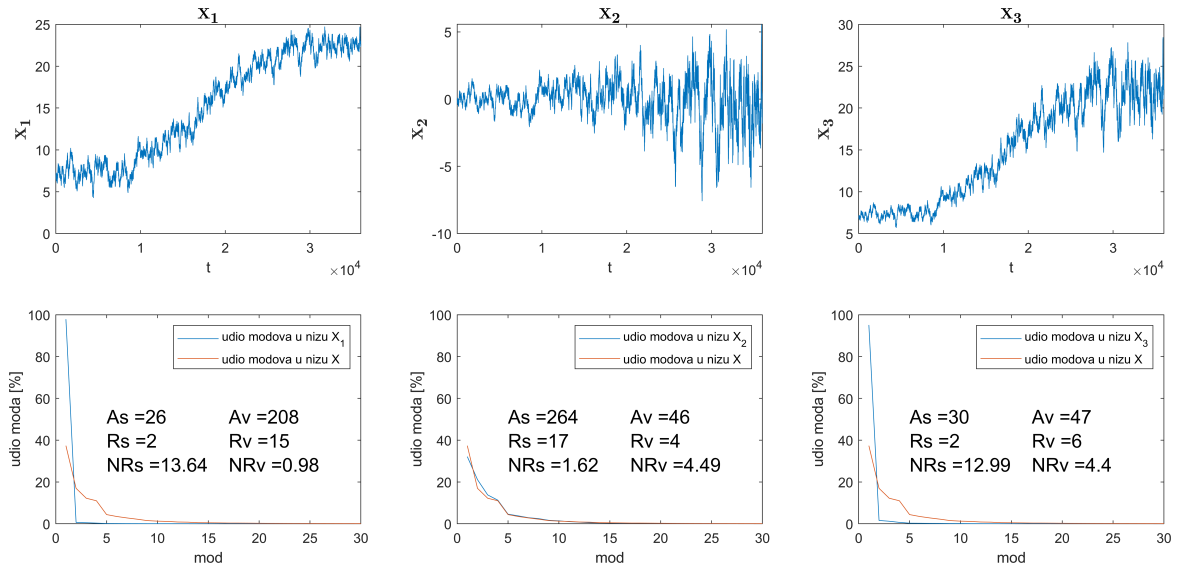
Slika 4.22. Prva 3  $EOF_v$ -a,  $PC_v$ -a i prva 3 moda EOF dekompozicije niza  $X$  zadanog jednačbom 3.10.

Vratimo se na daljnje razmatranje EOF dekompozicije niza  $X$  sa  $\phi = 0.995$ . Na slici 4.22 prikazana su prva tri  $EOF_v$ -a,  $PC_v$ -a te prva tri moda koja oni čine. Prvo što se primjećuje je da, za razliku od niza  $Q$ ,  $PC_v$ -ovi niza  $X$  posjeduju motiv. EOF metoda je prepoznala neku pravilnost u vremenskom razvoju svakog od 30 podnizova. Za taj novitet zaslužno je upravo dugo pamćenje prisutno u  $X$ , koje za posljedicu ima koreliranost podataka međusobno udaljenih više od 1200 vremenskih koraka (što je dužina jednog podniza). Autokorelacijska funkcija niza  $X$  sa  $\phi = 0.995$  prikazana je na slici 4.23.  $EOF_v$  prvog moda i u ovom slučaju, oblikom ocrta niz srednjaka podnizova, dok sam prvi mod pokušava opisati hod srednjaka za kojeg u ovom nizu nije trivijalno reći da li postoji ili ne. Za dovoljno mala vremena usrednjavanja hod srednjaka postoji, no zbog svoje stacionarne i velike varijabilnosti, u praktične svrhe za koje mi razmatramo stacionarnost nizova, možemo reći da je niz stacionarnog srednjaka. Mogli bismo reći da iz perspektive EOF analize, hod u nizu postoji čim se srednjaci podnizova razlikuju. Kako niz  $X$  posjeduje dugo pamćenje, tada se u njemu pojavljuju duža razdoblja u kojem postoji porast/pad koji na skali od 1200 uzastopnih podataka (jer toliko traje svaki od 30 podnizova) uzrokuje značajnije razlike u srednjacima tih podnizova.



Slika 4.23. Autokorelacija niza  $X$  zadanog jednadžbom 3.10.

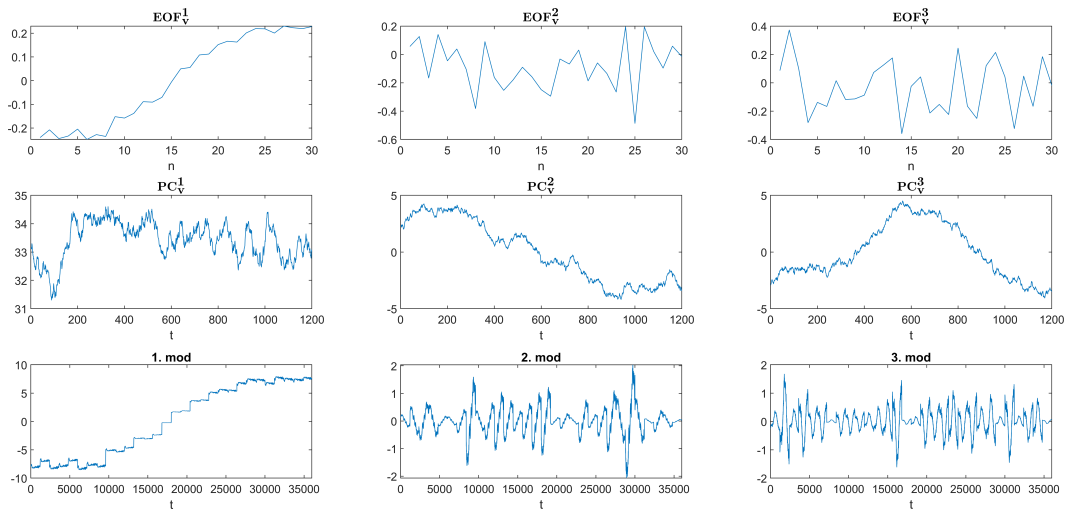
U nastavku promatramo tri niza nastala modificiranjem niza  $X$ :  $X_1$  - sa superponiranim jasnim hodom srednjaka i konstantne varijance,  $X_2$  - s jasnim hodom varijance i stacionarnog srednjaka,  $X_3$  - sa superponiranim hodom i promjenjivom varijancom. Motivacija za konstruiranje upravo takvih nizova je ista motivaciji za konstruiranje niza  $Q_9$ . Jednadžbe nizova su dane sa 3.12, 3.13, 3.14. Slika 4.24 prikazuje spomenute nizove te razdiobu udjela pojedinih modova u kontrastu s razdiobom udjela modova u nizu  $X$ . Prikazani su i parametri RUT,RAT i M98 metoda za hod srednjaka i hod varijance. Za niz  $X_1$  sve tri metode se slažu da posjeduje nestacionarni hod srednjaka te stacionarnu varijancu. Kao i u prošlom poglavlju, za takav tip vremenskog niza udio prvog moda je vrlo velik,  $\approx 97.97\%$  dok je udio drugog moda  $\approx 0.63\%$ . Za niz  $X_2$  sve



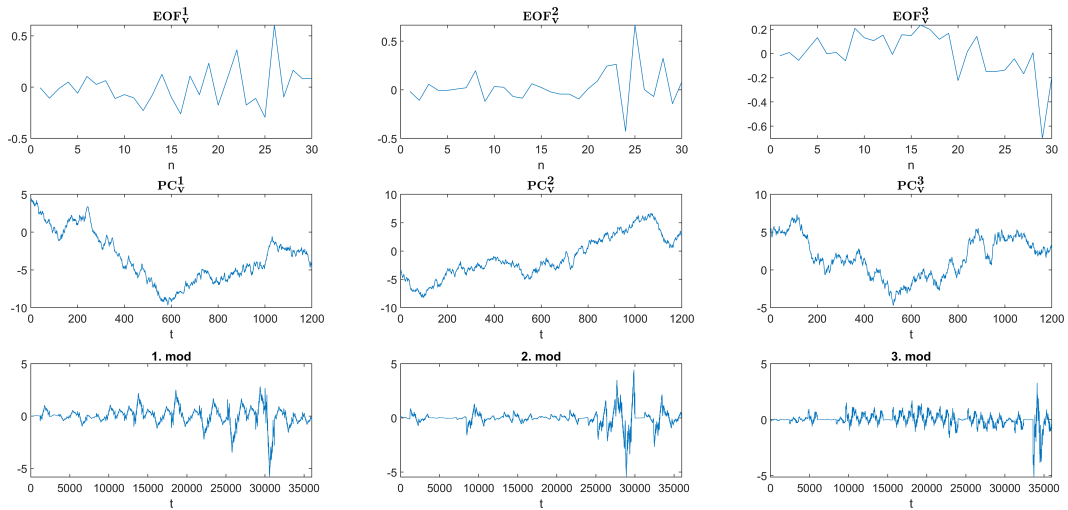
Slika 4.24. Gornji red - nizovi  $X_1$ ,  $X_2$  i  $X_3$  dani izrazima 3.12, 3.13, 3.14 respektivno. Donji red - udjeli pojedinih modova EOF dekompozicije pripadnih nizova (plava linija) u usporedbi sa udjelom pojedinih modova u nizu  $X$  (crvena linija). Prikazani su parametri  $A_s$ ,  $R_s$  i  $NR_s$  dobiveni primjenom RUT, RAT i M98 metoda na niz srednjaka te parametri  $A_v$ ,  $R_v$  i  $NR_v$  dobiveni primjenom RUT, RAT i M98 metoda na niz varijanci pripadnih nizova.

tri metode se slažu da posjeduje stacionaran hod i nestacionarnu varijancu. U razdiobi modova takvog niza gotovo da se i ne primjećuje promjena u odnosu na razdiobu niza  $X$  (koja od početka podsjeća na situaciju nestacionarne varijance iz prošlog poglavlja), iako je promjena varijance vrlo značajna. Za niz  $X_3$  sve tri metode se slažu da posjeduje nestacionaran hod i nestacionarnu varijancu. Udio prvog moda dominira nad ostalima, no s nešto manjim postotkom od  $\approx 95.10\%$ . Udio drugog moda narastao je na  $\approx 1.58\%$ . Dakle efekt je vrlo sličan kao i kod razmatranja nizova izvedenih iz bijelog šuma: nestacionarnost hoda uzrokuje prvenstveno iskakanje prvog moda, dok nestacionarnost varijance uzrokuje povećanje udjela prve polovice modova bez iskakanja prvog moda u odnosu na drugi.

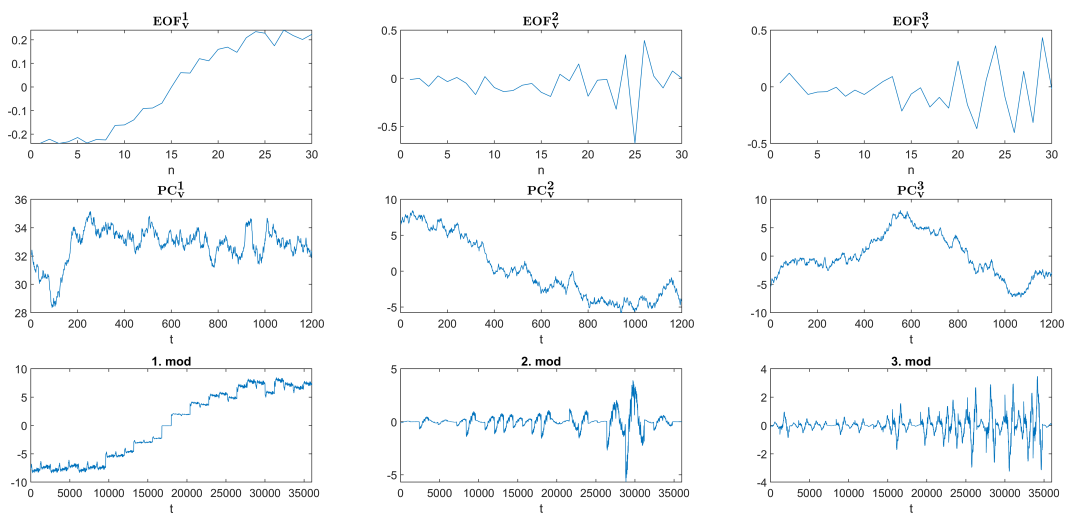
Na slici 4.25 prikazana su prva tri moda s pripadnim  $EOF_v$  i  $PC_v$  niza  $X_1$ . Uočava se očekivano nastojanje prvog moda da opiše oblik hoda srednjaka, za kojeg znamo da je funkcija tangens hiperbolni. Opis hoda srednjaka prvim modom je lošiji od slučaja kada je on bio superponiran na neautokorelirani niz (slika 4.17). Razlog tomu je što početni niz  $X$  posjeduje mali, vlastiti hod srednjaka podnizova, koji je superponiran na stvarni hod. EOF analiza u konačnici, ne razlikuje od kuda potječe hod srednjaka te sve svrstava u prvi mod. Drugi i treći mod podsjećaju upravo na drugi i treći mod niza  $X$ . Na slici 4.26 prikazana su prva tri moda s pripadnim  $EOF_v$  i  $PC_v$  niza  $X_2$ . Iako su razdiobe udjela modova između niza  $X_2$  i niza  $X$  vrlo male, vidi se značajna razlika u tome što ti modovi opisuju. Kako varijanca niza raste tako raste i amplituda hoda srednjaka podnizova, i to je ono što prvi mod opisuje. Drugi i treći mod nastoje



Slika 4.25. Prva 3  $EOF_v$ -a,  $PC_v$ -a i moda niza  $X_1$  zadanog izrazom 3.12.



Slika 4.26. Prva 3  $EOF_v$ -a,  $PC_v$ -a i moda niza  $X_2$  zadanog izrazom 3.13.



Slika 4.27. Prva 3  $EOF_v$ -a,  $PC_v$ -a i moda niza  $X_3$  zadanog izrazom 3.14.

opisati područja najveće varijance. Na slici 4.27 prikazana su prva tri moda s pripadnim  $EOF_v$  i  $PC_v$  niza  $X_3$ . Prvi mod je, kao i kod niza  $X_1$ , opisao hod dok su drugi i teći mod opisali varijancu niza.

Generalno, zaključci izvedeni iz primjene EOF analize na nizove dobivene iz AR(1) procesa se u suštini ne razlikuju od onih iz prethodnog poglavlja. Kako niz izveden iz AR(1) posjeduje mali lokalni hod srednjaka i varijance, glavna razlika je ta da te pojave uzrokuju rast u prvoj polovici modova za inače stacionaran niz (što se prije dogodilo za niz s nestacionarnom varijancom). Razlika je, dakle, da se mijenja pojam stacionarnog niza dok su reakcije na uvedene nestacionarnosti ne mijenjaju bitno. Uzevši to u obzir, svi zaključci s kraja prethodnog poglavlja i dalje vrijede.

## 5. Realni podaci

Do sada smo opisali kako nestacionarnost momenata prvog i drugog reda utječe na rezultate EOF analize. U sljedećem koraku EOF analizu primjenjujemo na realne vremenske nizove. Podaci s kojima radimo su vremenski nizovi brzine vjetra i sonične temperature u trajanju od 30 min za vrijeme epizoda bure. Podaci su prikupljeni u periodu od 09.10.2015. do 09.10.2016. na meteorološkom tornju postavljenom oko 200 m sjeveroistočno od Masleničkog mosta na visinama 2, 5 i 10 m s frekvencijom uzorkovanja 20 Hz (npr. Šoljan et al., 2018). Ukupno je zabilježeno 48 epizoda bure čije se vrijeme trajanja moglo podijeliti u 2363, 30 min intervala. U svakom intervalu je 36000 podataka. Na podacima je provedena transformacija svakog 30 minutnog intervala podataka u koordinatni sustav srednjeg horizontalnog vjetra. Kako nas zanima samo stacionarnost/nestacionarnost pojedinih vremenskih intervala, u analizi neće biti pridavana važnost da li se radi o  $u, v, w$  brzinama vjetra ili soničnoj temperaturi  $t$ . Zbog te činjenice, ukupan broj vremenskih nizova s kojima raspolažemo je 28356.

Tablica 5.1. Broj vremenskih nizova u pojedinoj kategoriji.

Stac. srednjak i stac. varijanca	Nestac. srednjak i stac. varijanca	Stac. srednjak i nestac. varijanca	Nestac. srednjak i nestac. varijanca
11911	1825	267	383

Kako bi mogli uočiti razlike u utjecaju nestacionarnosti momenta prvog reda od nestacionarnosti momenta drugog reda, iz ukupnog bazena podataka izdvojit ćemo 4 različite kategorije vremenskih nizova. Primijenit ćemo metode RUT, RAT i M98 na hod srednjaka i na hod varijanci svih 28356 nizova. Na temelju rezultata tih metoda vremenske nizove dijelimo u kategorije:

1. vremenski nizovi koji istovremeno prema RUT, RAT i M98 metodama posjeduju stacionarnost srednjaka i stacionarnost varijance
2. vremenski nizovi koji istovremeno prema RUT, RAT i M98 metodama posjeduju nestacionarnost srednjaka i stacionarnost varijance
3. vremenski nizovi koji istovremeno prema RUT, RAT i M98 metodama posjeduju stacionarnost srednjaka i nestacionarnost varijance
4. vremenski nizovi koji istovremeno prema RUT, RAT i M98 metodama posjeduju nestacionarnost srednjaka i nestacionarnost varijance

Broj nizova u svakoj od kategorija dan je u tablici 5.1.

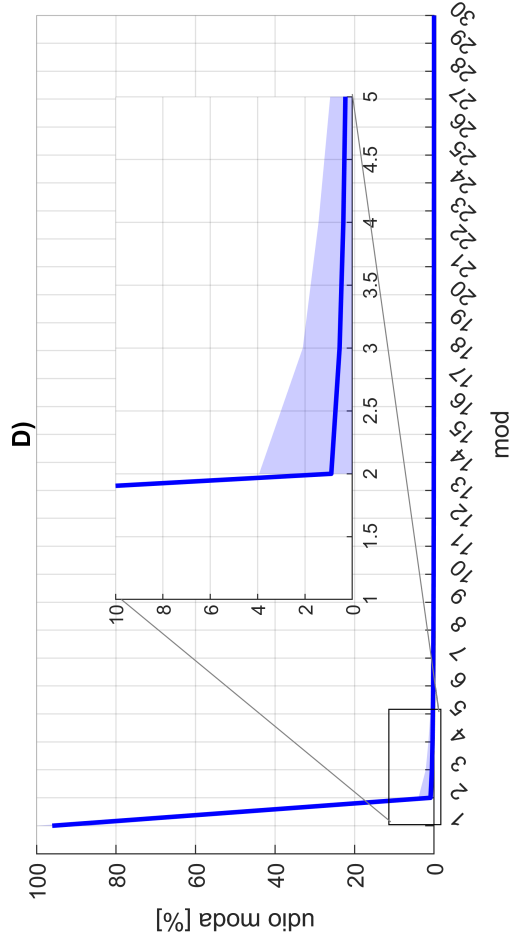
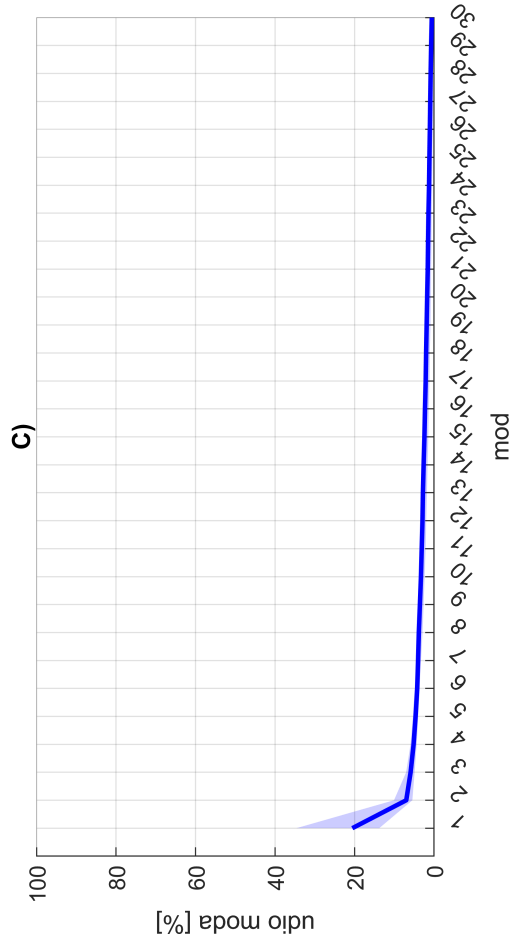
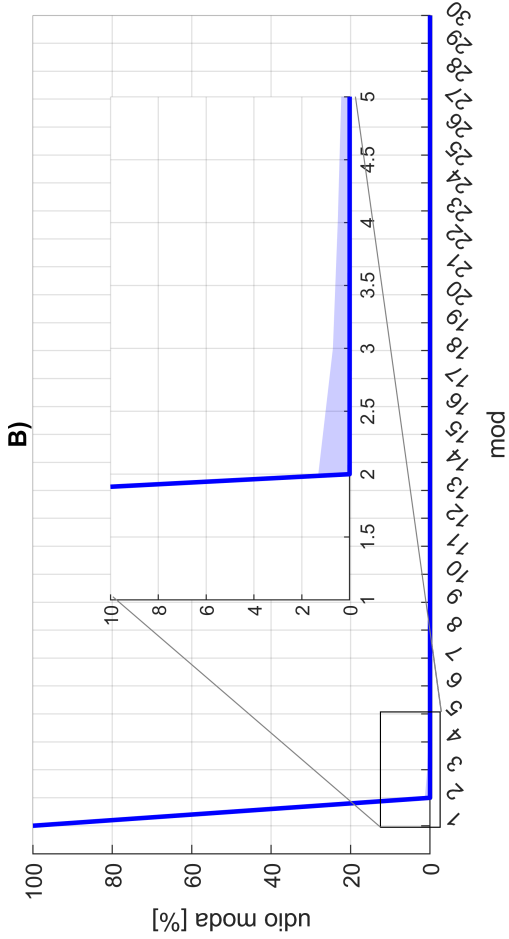
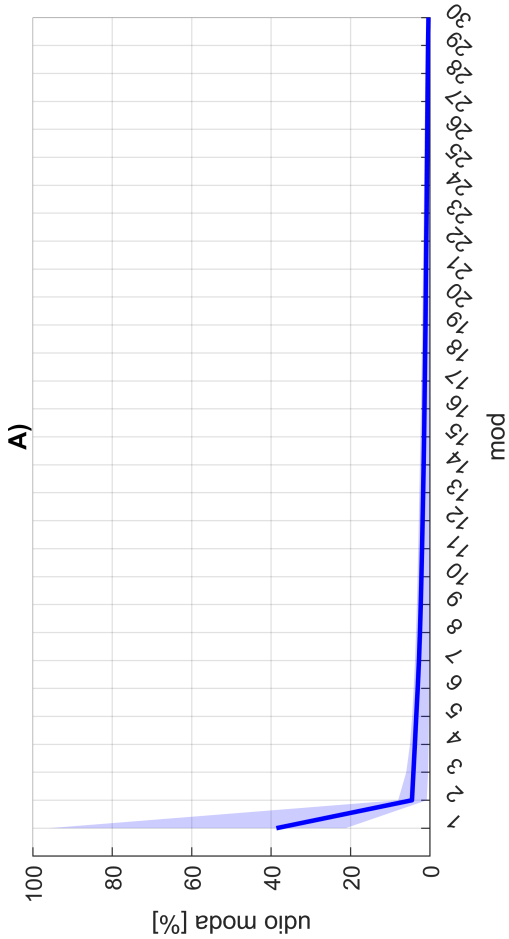


## 6. Primjena EOF analize na realne vremenske nizove

Imajući u vidu zaključke iz prethodnih poglavlja, gdje je EOF analiza primijenjena na različite umjetne nizove, sada se EOF analiza primijenjuje na realne nizove brzine vjetrova i sonične temperature (pobliže opisano u poglavlju 5). EOF analizu primjenjujemo na 4 skupine podataka opisanih u poglavlju 5. Među rezultatima primjene EOF analize na vremenske nizove, najveći indikator nestacionarnosti je raspodjela udjela modova. Medijani udjela svake pojedine kategorije te 75. i 25. percentil koji omeđuju područje koje sadrži 50% udjela prikazani su na slici 6.1 na sljedećoj stranici. Za razliku od prethodnih poglavlja, ovdje nema raspodjele udjela modova potpuno stacionarnog niza s kojom bi uspoređivali rezultate. Iz tog razloga, sva opažanja su izvedena na temelju prethodno stečenih saznanja.

Možemo primijetiti da za skupinu nizova stacionarnog hoda i stacionarne varijance (slika 6.1 A), medijan pokazuje relativno visok udio u prvom modu  $\approx 40\%$ , te primjetnu zastupljenost nižih modova (1-5 %). Prema zaključcima iz prethodnog poglavlja, to je upravo ono što bi i očekivali od stacionarnog niza s dugom autokorelacijom. Osim toga, uočavamo vrlo visoko rasipanje u udjelu prvog moda, unutar 75. i 25. percentila. 25. percentil je na 20% udjela prvog moda dok je 75. na 93% udjela prvog moda. Iz perspektive EOF analize i onog što smo zaključili, veliko rasipanje iznad medijana bi mogli objasniti kao posljedicu toga što su neki nestacionarni nizovi, proglašeni stacionarnim (njihov hod i varijanca istovremeno) od strane sve tri metode. Za veliko rasipanje ispod medijana odgovorni su nizovi s manjom autokorelacijom tj. nizovi koji više podsjećaju na niz  $Q$  iz poglavlja 4.1 (to zaključujemo prema slici 4.21).

Slika 6.1 B odgovara skupini nizova nestacionarnog hoda srednjaka i stacionarne varijance. Vidljivo je vrlo malo rasipanje (gotovo nezamjetno) udjela svih modova, pa tako i prvog, te gotovo 100% udio prvog moda. Takav rezultat se slaže s dosadašnjim zaključcima - vrlo visok udio prvog moda te nezamjetni udjeli svih ostalih sugerira postojanje hoda srednjaka te stacionarnost varijance. Vrlo malo rasipanje mogli bi objasniti time da je EOF analiza stroža (osjetljivija) na nestacionarnost prvog momenta od kombinacije RUT, RAT i M98 metoda. Malo rasipanje upućuje na to da su gotovo svi nizovi, koji prema RUT, RAT i M98 metodama posjeduju nestacionarnost hoda srednjaka, izazvali velik udio prvog moda što je marker nestacionarnosti srednjaka. Imajući u vidu objašnjenje za veliko rasipanje udjela prvog moda u slučaju A, može se zaključiti da je EOF analiza osjetljivija na nestacionarnost prvog momenta od kombinacije RUT, RAT i M98 metoda.



Slika 6.1. A) - medijan (plava linija) te 75. i 25. percentil (granice osjenčanog područja) udjela modova u nizovima stacionarnog hoda i stacionarne varijance. B) - medijan (plava linija) te 75. i 25. percentil (granice osjenčanog područja) udjela modova u nizovima nestacionarnog hoda i stacionarne varijance. C) - medijan (plava linija) te 75. i 25. percentil (granice osjenčanog područja) udjela modova u nizovima stacionarnog hoda i nestacionarne varijance. D) - medijan (plava linija) te 75. i 25. percentil (granice osjenčanog područja) udjela modova u nizovima nestacionarnog hoda i nestacionarne varijance.

Slika 6.1 C odgovara skupini nizova stacionarnog hoda srednjaka i nestacionarne varijance. Vidimo očekivanu situaciju za vremenske nizove s nestacionarnim momentom drugog reda - značajan udio prve polovice modova, udjeli prvog i drugog moda nisu drastično različiti, udjeli modova opadaju približno eksponencijalno. Rasipanje prvog i drugog moda postoji, no značajno je manje nego u slučaju A.

Slika 6.1 D odgovara skupini nizova nestacionarnog hoda srednjaka i nestacionarne varijance. Vidimo odskakanje prvog moda na  $\approx 97\%$  što je manje nego u slučaju B. Također vidimo veći udio drugog i trećeg moda nego u slučaju B. Upravo to malo povećanje u zastupljenosti drugog (i svih sljedećih) moda sugerira nestacionarnost hod srednjaka i varijance istovremeno. Također vidimo i izraženije rasipanje vrijednosti drugog moda od one u slučaju B kada je varijanica bila stacionarna. Iz tog rasipanja je vidljiv utjecaj nestacionarnosti hoda varijance na porast udjela prve polovice modova.

## 7. Zaključak

Cilj rada je bio provjeriti da li se primjenom EOF analize na vremenske nizove, može odrediti njihova stacionarnost/nestacionarnost. Da bi razumjeli na koji način EOF analiza reagira na nestacionarnost, u prvom koraku je primjenjujemo na niz bijelog šuma za kojeg sa sigurnošću možemo reći da je stacionaran. Uočavamo da je EOF analiza osjetljiva na postojanje srednje vrijednosti niza. Ukoliko ona postoji, udio prvog moda EOF dekompozicije niza zasjenit će sve ostale. Kako srednja vrijednost niza nikako ne utječe na stacionarnost, a utječe znatno na EOF analizu, iz svih sljedećih nizova uklonjena je njihova srednja vrijednost. Zatim modifikacijama niza bijelog šuma konstruiramo nestacionarne nizove, nestacionarnog hoda srednjaka i/ili varijance. Na njih je primijenjena analiza te su razmatrane različitosti i sličnosti u dobivenim rezultatima s obzirom na bijeli šum. Zaključujemo da EOF analiza doista reagira na nestacionarnosti u vremenskom nizu. Na nestacionarnost niza koji posjeduje nestacionarni hod srednjaka a stacionarni hod varijance, reagira sa:

1. značajno izraženim prvim modom,
2. približno linearnim opadanjem udjela drugog, pa sve do posljednjeg moda i to tako da je udio zadnjeg moda različit od 0% ,
3.  $PC_v$  prvog moda posjeduje srednjak različit od 0,
4.  $EOF_v$  prvog moda ocrtava oblik hoda,
5. prvi mod opisuje hod vremenskog niza.

EOF analiza na nestacionarnost niza koji posjeduje stacionarni hod srednjaka a nestacionarni hod varijance, reagira sa:

1. značajno izraženom prvom polovicom modova (udio prvog i drugog moda se ne razlikuje drastično),
2. nelinearnim opadanjem udjela modova, tako da je udio zadnjeg moda gotovo 0%,
3.  $EOF_v$  prvog moda locira podniz/podnizove s najvećom varijancom tako da je u većini točaka  $\sim 0$ , osim u njih nekoliko, gdje je gotovo 1,
4. prvi mod opisuje podniz/podnizove koji posjeduju najveću varijancu.

EOF analiza nestacionarnost niza koji posjeduje nestacionarni hod srednjaka i nestacionarni hod varijance, prepoznaje po:

1. značajno izraženom prvom modu,
2. nelinearnom opadanju udjela drugog, pa sve do posljednjeg moda, tako da je udio zadnjeg moda gotovo 0%,
3.  $EOF_v$  prvog moda ocrtava oblik hoda,
4.  $EOF_v$  drugog moda locira podniz/podnizove s najvećom varijancom tako da je u većini točaka  $\sim 0$ , osim u njih nekoliko, gdje je gotovo 1,
5.  $PC_v$  prvog moda posjeduje srednjak različit od 0,
6. prvi mod opisuje hod vremenskog niza.
7. drugi mod opisuje podniz/podnizove koji posjeduju najveću varijancu.

Kako su realni vremenski nizovi, za razliku od bijelog šuma, posljedica autokoreliranog procesa, konstruiramo umjetni vremenski niz pomoću AR(1) metode. U sljedećem koraku na njemu, te na njegovim modifikacijama (autokorelirani niz sa stacionarnim/nestacionarnim srednjakom i stacionarnom/nestacionarnom varijancom) ispituje se EOF analiza. Na temelju dobivenih rezultata dolazimo do istih zaključaka kao i u prethodnom slučaju.

Na kraju, EOF analizu provodimo na 4 grupe realnih vremenskih nizova dobivenih mjerenjem brzine vjetera te sonične temperature u epizodama bure na meteorološkom tornju nedaleko Masleničkog mosta. Te 4 grupe su vremenski nizovi za koje su metode RUT, RAT i M98 istodobno ustvrdile da posjeduju: 1) stacionaran hod i stacionarnu varijancu, 2) nestacionaran hod i stacionarnu varijancu, 3) stacionaran hod i nestacionarnu varijancu i 4) nestacionaran hod i nestacionarnu varijancu. Medijani raspodjela udjela modova EOF dekompozicije nizova iz spomenutih grupa potvrđuju da je EOF analizom doista moguće prepoznati nestacionarnost. Štoviše, raspodjela udjela modova EOF dekompozicije poprma drugačiji oblik s obzirom na to da li je niz nestacionaran zbog nestacionarnosti hoda srednjaka ili nestacionarnosti varijance.

Sve navedeno upućuje na to da je moguće konstruirati EOF metodu kojom bi bilo moguće detektirati nestacionarnost vremenskih nizova. Kako bi definirali konkretnu metodu na temelju EOF analize, nedostaju konkretni kriteriji. Sljedeći koraci u istraživanju mogućnosti uporabe EOF analize za određivanje nestacionarnosti trebali bi biti upravo definiranje kriterija te usporedba nove EOF metode s postojećim.

## 8. Daljnja istraživanja

Zaključili smo da EOF analiza doista može prepoznati nestacionarnost vremenskih nizova i da je idući korak u uspostavi nove, EOF, metode upravo određivanje kriterija na temelju kojih bi ona mogla donijeti ocjenu stacionarnosti. U ovom poglavlju sugeriramo nekoliko smjernica za koje smatramo da bi mogle biti od važnosti na tom putu.

Kako je pojava nestacionarnosti najuočljivija u raspodjeli udjela EOF modova, jedan od mogućih kriterija za određivanje nestacionarnosti nekog niza mogla bi biti upravo razdioba udjela pojedinih modova EOF dekompozicije u tom nizu. Nestacionarnost hoda srednjaka niza bilo bi vrlo lako detektirati samo na temelju udjela prvog moda. Jedna od mogućnosti bila bi postavljanje gornje granice udjela prvog moda  $GG_1$ , za koju bi nizove smatrali još uvijek stacionarnim. Nameće se, dakle, problem odabira  $GG_1$ , koji nas vraća na temu s početka poglavlja 4.1. - teorija slučajnih matrica. Rezultati nedavnih istraživanja (Bejan, 2005; Yin et al., 1988) u tom području ukazuju da postoji mogućnost određivanja najveće svojstvene vrijednosti matrica kovarijanci dobivenih iz matrica čiji su elementi dobiveni iz proizvoljnog ARMA(p,q) stohastičkog procesa (npr. Jin et al. (2009); Wang et al. (2011)). U tom slučaju, postoji mogućnost da se odredi neki pogodni ARMA(p,q) proces, kojeg bi proglasili dovoljno dobrim u opisu vremenskih nizova koje promatramo. Na temelju tog ARMA(p,q) procesa odredili bi najveću teoretsku svojstvenu vrijednost. Iz koje bi bilo moguće dobiti udio  $GG_1$  koji bi bio proglašen gornjom granicom stacionarnosti. No problem postavljanja kriterija je u tom slučaju samo odgođen. Umjesto odabira  $GG_1$ , potrebno je odabrati ARMA(p,q) proces na temelju kojeg bi bilo moguće izračunati  $GG_1$ . Smatramo da je odabir ARMA procesa lakše opravdati od jednostavnog odabira granice  $GG_1$ . Nadalje, ako bi odredili raspodjelu svojstvenih vrijednosti tog ARMA(p,q) procesa, tada bi jedan od mogućih kriterija bio i kakvoća poklapanje raspodjele svojstvenih vrijednosti dobivenih iz EOF analize primijenjene na danom nizu s teorijskom raspodjelom. Nizovi čija raspodjela svojstvenih vrijednosti previše odskaka od teorijske bili bi proglašeni nestacionarnim.

Detektiranje nestacionarnosti hoda varijance je teži zadatak. Ono što su rezultati u ovom radu dali naslutiti je da bi za detektiranje hoda varijance bilo potrebno proučavati sve modove osim prvog. Razlog tomu je što i najmanja pojava hoda srednjaka uzrokuje značajan udio prvog moda što zasjenjuje utjecaj nestacionarnosti varijance. Uz sve rečeno, evidentno je da prostora za nastavak istraživanja definitivno ne manjka.

## Literatura

- Bejan, A. (2005). Largest eigenvalues and sample covariance matrices. *Diplomski rad, University of Warwick*, 1–88.
- Bendat, J. S. i A. G. Piersol (1986). Random data: Analysis and measurements procedures. *John Wiley & Sons*, 1–621.
- Björnsson, H. i S. Venegas (1997). A manual for eof and svd analyses of climatic data. *Department of Atmospheric and Oceanic Sciences and Centre for Climate and Global Change Research, McGill University*, 1–50.
- Dommenget, D. (2007). Evaluating eof modes against a stochastic null hypothesis. *Climate dynamics* 28, 517–531.
- Dommenget, D. i M. Latif (2002). A cautionary note on the interpretation of eofs. *Journal of climate* 15, 216–225.
- Hamilton, J. D. (1994). Time series analysis. *Princeton New Jersey* 2, 1–792.
- Hannachi, A. (2004). A primer for eof analysis of climate data. *Department of Meteorology, University of Reading*, 1–29.
- Hyndman, R. J. i G. Athanasopoulos (2018). Forecasting: principles and practice. *OTexts*, 1–504.
- Jin, B., C. Wang, B. Miao, i M.-N. L. Huang (2009). Limiting spectral distribution of large-dimensional sample covariance matrices generated by varma. *Journal of multivariate analysis* 100, 2112–2125.
- Mahrt, L. (1998). Flux sampling errors for aircraft and towers. *Journal of Atmospheric and Oceanic Technology* 15, 416–429.
- Myers, D. E. (1989). To be or not to be... stationary? that is the question. *Mathematical Geology* 21(3), 347–362.
- Preisendorfer, R. W. i C. D. Mobley (1988). Principal component analysis in meteorology and oceanography. *Developments in atmospheric science* 17, 1–425.
- Šoljan, V., A. Belušić, K. Šarović, I. Nimac, S. Brzaj, J. Suhin, M. Belavić, Ž. Večenaj, i B. Grisogono (2018). Micro-scale properties of different bora types. *Atmosphere*, 1–116.
- Stuart, A., M. G. Kendall, et al. (1963). The advanced theory of statistics. *Griffin*, 1–684.

- Stull, R. B. (1988). An Introduction to Boundary Layer Meteorology. *Kluwer Academic Publishers: Dordrecht, The Netherlands*, 1–666.
- Večenaj, Ž. i S. F. De Wekker (2015). Determination of non-stationarity in the surface layer during the t-rex experiment. *Quarterly Journal of the Royal Meteorological Society* 141(690), 1560–1571.
- Wang, C., B. Jin, i B. Miao (2011). On limiting spectral distribution of large sample covariance matrices by varma (p, q). *Journal of Time Series Analysis* 32, 539–546.
- Yin, Y.-Q., Z.-D. Bai, i P. R. Krishnaiah (1988). On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability theory and related fields* 78, 509–521.