

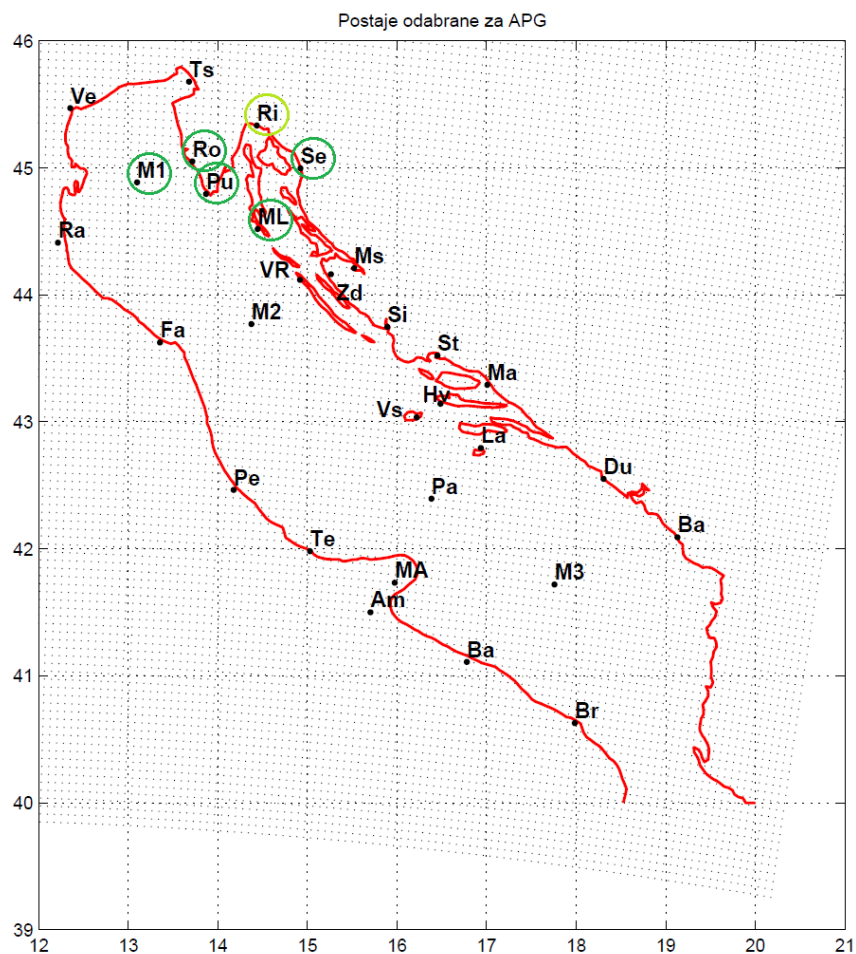
Linearna regresija

HELENA LATEČKI, II-134-19

Geofizički odsjek, PMF, Horvátovac ul. 95, 10 000 Zagreb

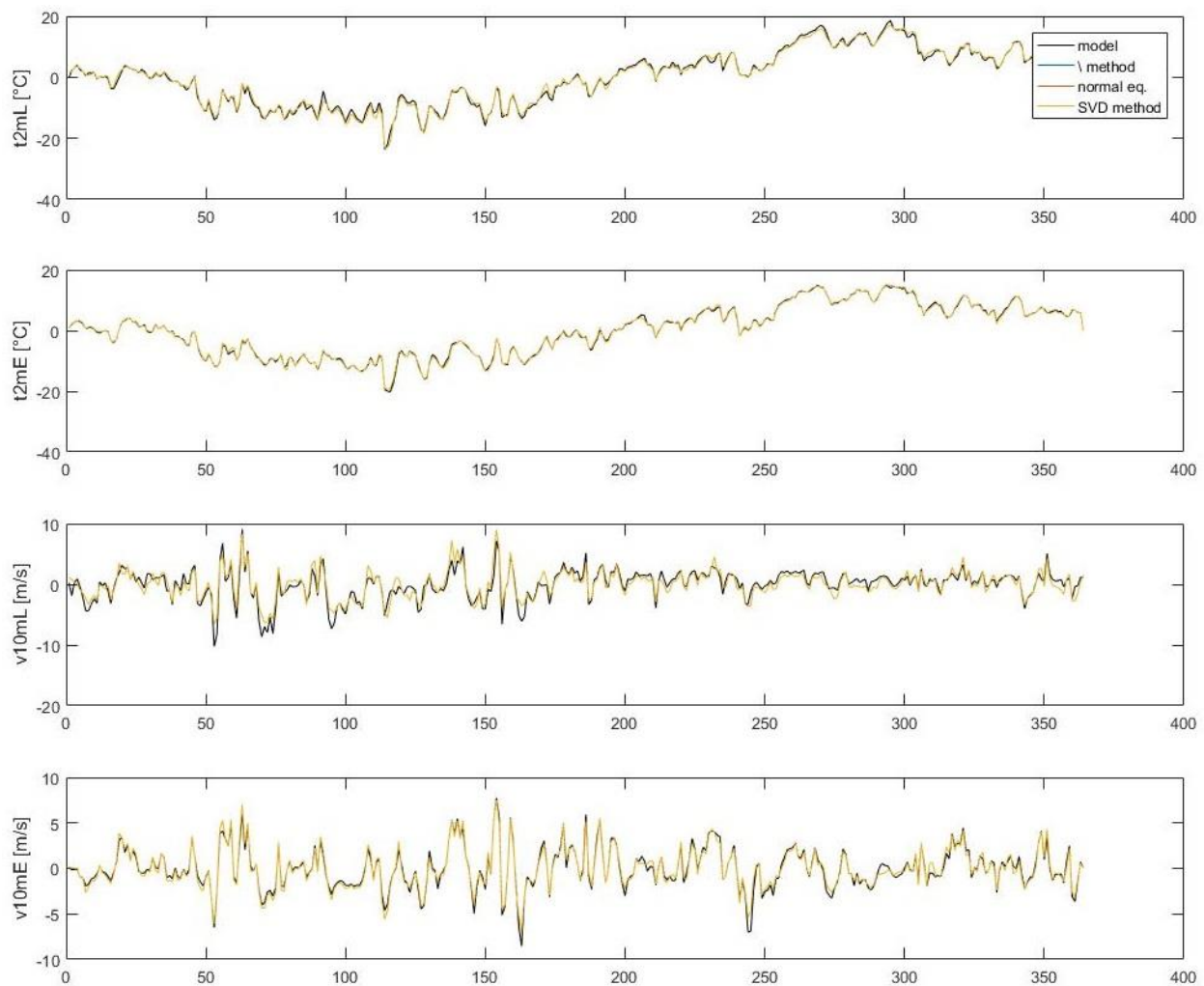
1.10.2021.

Za odabrane postaje prikazane na Slici 1. (fiksirana je postaja Ri (Rijeka), te je odabrano još pet postaja: M1 , Ro - Rovinj, Pu - Pula, Se - Senj i ML - Mali Lošinj), analizirani su vremenski nizovi srednjih dnevnih temperatura na 2 m iznad tla i sjeverne komponente vjetra na 10 m visine. Nizovi potječu iz meteorološkog numeričkog modela LAMI (Italija) te modela Europskog centra za srednjoročnu prognozu vremena (ECMWF) te se odnose na razdoblje od 2.11.2002. do 30.9.2003.



Slika 1. Karta postaja odabranih za APG - svijetlo zelenom bojom naznačena je fiksirana postaja, a tamno zelenom bojom preostale odabrane postaje.

Korištenjem Matlab funkcije *detrend*(., 'constant'), za nizove koji potječu iz modela LAMI, uklonjena je njihova srednja vrijednost. U slučaju nizova ECMWF modela, zbog prisutnosti *NaN* vrijednosti, srednja vrijednost uklonjena je direktnim oduzimanjem zadanog niza i njegove srednje vrijednosti dobivene pomoću Matlab funkcije *mean*(., 'omitnan'). Radi provedbe daljnjeg računa, *NaN* vrijednosti su zamijenjene s nulama. Potom je provedena linearna regresija fiksirane postaje Rijeka pomoću preostalih pet odabranih postaja i to na tri načina: korištenjem Matlab funkcije "\", putem normalnih jednadžbi te singularne dekompozicije (SVD-rastav). Ulazni vremenski nizovi i nizovi dobiveni regresijom prikazani su na Slici 2.



Slika 2. Izvorni vremenski niz (u legendi model) i vremenski nizovi dobiveni linearnom regresijom na tri načina.

Iz Slike 2. može se zaključiti da sve tri metode daju identične rezultate do na grešku zaokruživanja, čemu ujedno svjedoče i same razlike regresijskih koeficijenata prikazanih u Tablici 1. Iznosi samih koeficijenata regresije dani su u Tablici 2.

Tablica 1. Razlike regresijskih koeficijenata računatih na tri načina.

	<i>koef(\) - koef(norm.)</i>	<i>koef(\) - koef(SVD)</i>	<i>koef(norm.) - koef(SVD)</i>
<i>t2DnL</i>	4.5103e-14	8.8818e-16	8.3877e-14
<i>t2DnE</i>	2.9443e-12	6.2172e-15	2.1865e-12
<i>v10DnL</i>	2.1372e-15	4.9960e-16	3.3307e-16
<i>v10DnE</i>	1.2074e-15	3.6082e-16	9.7977e-15

Tablica 2. Koeficijenti regresije za temperaturu i sjevernu komponentu vjetra te modele LAMI i ECMWF.

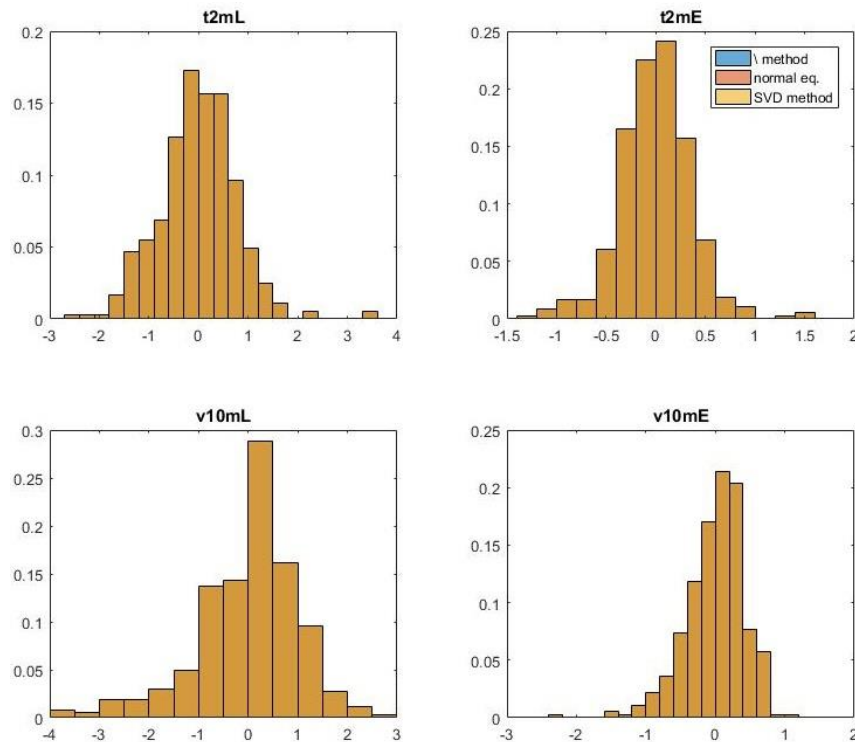
	<i>koef. regresije</i>			
<i>postaja</i>	temp. (LAMI)	temp. (ECMWF)	v (LAMI)	v (ECMWF)
<i>Rovinj</i>	0.6287	1.6015	-0.2387	0.3587
<i>Senj</i>	0.3280	0.6147	0.6599	0.9967
<i>M1</i>	-0.0812	-0.3247	0.1182	-0.1213
<i>Pula</i>	0.2331	-0.8817	0.6520	0.1960
<i>Mali Lošinj</i>	-0.1541	-0.0180	-0.3035	-0.1861

Slaganje između ulaznog vremenskog niza i vremenskih nizova dobivenih linearnom regresijom za model LAMI bolje je u slučaju sjeverne komponente vjetra ($\max(t-t_i) = 9.72$ °C; $\max(v-v_i) = 7.45$ m/s), dok je za model ECMWF slaganje bolje u slučaju temperature ($\max(t-t_i) = 6.27$ °C; $\max(v-v_i) = 7.69$ m/s). Za oba modela iznosi koeficijenata korelacije (Tablica 3) su veći u slučaju temperature nego vjetra, što je i za očekivati zbog veće prostorne varijabilnosti vjetra. Isto tako, postaje za koje su vrijednosti koeficijenata korelacije veće više utječu na fiksiranu postaju Rijeka. S obzirom na to, u slučaju temperature na postaju Rijeka najviše utječe Rovinj, a u slučaju vjetra, najveći utjecaj ima postaja Senj.

Tablica 3. Kvadrati koeficijenta korelacije između promatranih nizova na zadanim postajama i postaje Rijeka za temperaturu i sjevernu komponentu vjetra te modele LAMI i ECMWF.

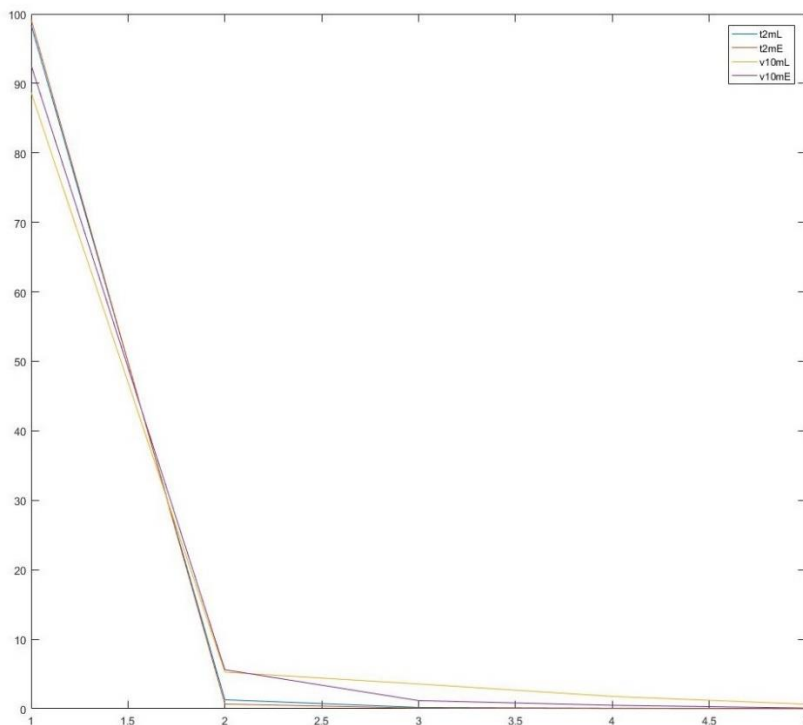
	<i>kvadrat koef. višestruke korelacije</i>			
<i>postaja</i>	temp. (LAMI)	temp. (ECMWF)	v (LAMI)	v (ECMWF)
<i>Rovinj</i>	0.9887	0.9820	0.5409	0.8567
<i>Senj</i>	0.9840	0.9933	0.6767	0.9414
<i>M1</i>	0.9438	0.9606	0.4895	0.7188
<i>Pula</i>	0.9596	0.9783	0.7092	0.8635
<i>Mali Lošinj</i>	0.9384	0.9750	0.3585	0.7881

Na Slici 3. prikazani su histogrami pogrešaka regresije za tri korištene metode. Histogrami triju metoda razlikuju se do na grešku zaokruživanja i grupirani su oko nule s oblikom nalik Gaussovoj raspodjeli. Rasap pogreške manji je u slučaju modela ECMWF, tj. linearna regresija u ovom slučaju bolje reproducira rezultate.



Slika 3. Histogrami pogrešaka linearne regresije provedene na tri načina.

Na Slici 4. prikazani su kvadrati singularnih vrijednosti normiranih na 100%. Kvadrati singularnih vrijednosti ukazuju na varijabilnost u podatcima - zbroj kvadrata singularnih vrijednosti jednak je zbroju kvadrata dijagonalnih elemenata matrice, a samim time i varijanci. Za temperaturu, prva singularna vrijednost nosi veći dio varijance, a sve ostale su približno jednake nuli. Drugim riječima, za opis ulaznog vremenskog niza temperature, pri regresiji je dovoljno uzeti samo prvi singularnu vrijednost. Za vjetar, prve tri (za ECMWF model), tj. četiri (za LAMI model) singularne vrijednosti nisu približne nulama, pa se samim time trebaju uzeti u obzir prilikom analize.



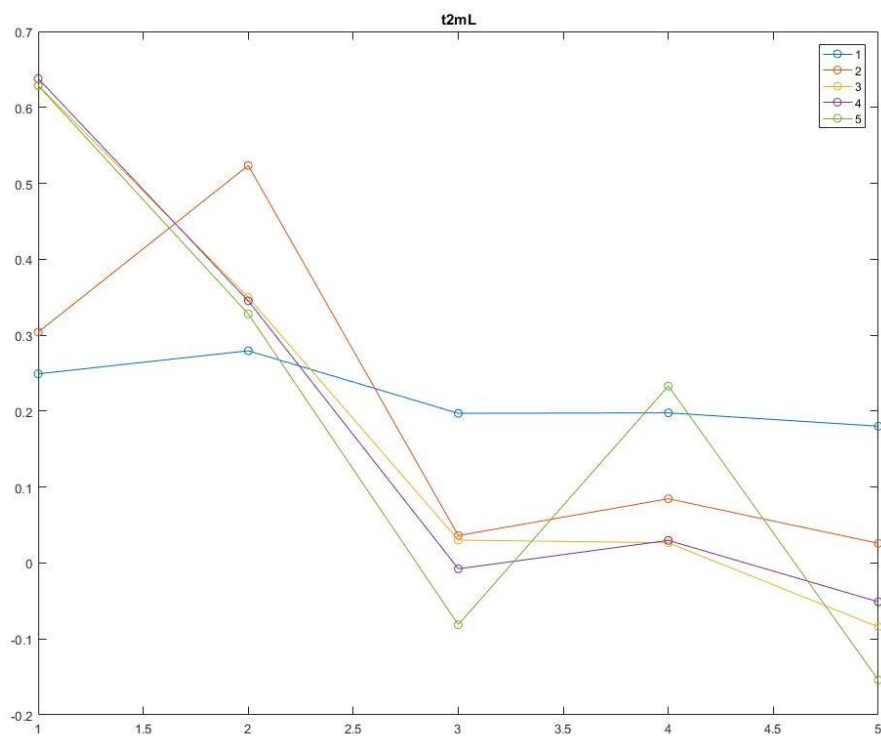
Slika 4. Kvadrati singularnih vrijednosti normirani na 100%.

U daljnjoj analizi, odabrani su vremenski nizovi koji potječu iz LAMI modela te je proveden račun linearne regresije preko SVD tako da su redom uključene po jedna, dvije, ..., pet komponenti SVD-rastava. Pripadni koeficijenti regresije popisani su u Tablici 4. te prikazani na Slikama 5. i 6. Koeficijenti korelacije na kvadrat popisani su u Tablici 5. i prikazani na Slici 7. Varijabilnost koeficijenta regresije između odabranih postaja za temperaturu i sjevernu komponentu vjetra raste s povećanjem uključenih komponenti SVD-rastava. S druge strane, kvadrati koeficijenata korelacije dobiveni kumulativnim uključivanjem komponenti SVD-rastava, prate trend koji se uočava i kod kvadrata singularnih vrijednosti (Slika 4). Uzimanjem samo jedne komponente SVD-rastava za temperaturu se uspjeva objasniti više od 98% varijance ulaznog vremenskog niza. U slučaju sjeverne komponente vjetra, čak i nakon uključivanja svih pet komponenti rastava, objašnjava se tek nešto više od 82% varijance.

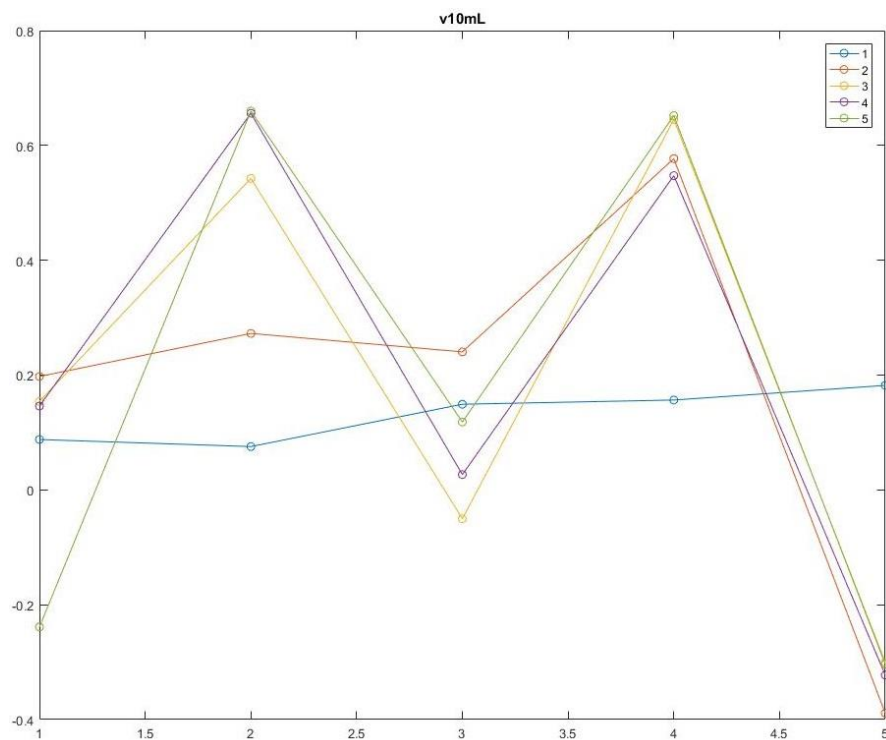
Tablica 4. Koeficijenti regresije s redom uključenom jednom, dvije, ..., pet komponenti SVD-rastava za temperaturu i sjevernu komponentu vjetra.

postaja	broj komponente SVD-rastava za temperaturu				
	1	2	3	4	5
Rovinj	0.2491	0.3041	0.6293	0.6379	0.6287
Senj	0.2795	0.5235	0.3498	0.3453	0.3280
M1	0.1971	0.0362	0.0302	-0.0076	-0.0812
Pula	0.1977	0.0846	0.0267	0.0298	0.2331
Mali Lošinj	0.1802	0.0260	-0.0842	-0.0511	-0.1541

	<i>broj komponente SVD-rastava za vjetar</i>				
<i>postaja</i>	1	2	3	4	5
<i>Rovinj</i>	0.0878	0.1976	0.1542	0.1460	-0.2387
<i>Senj</i>	0.0755	0.2726	0.5425	0.6560	0.6599
<i>M1</i>	0.1492	0.2404	-0.0502	0.0267	0.1182
<i>Pula</i>	0.1566	0.5770	0.6451	0.5474	0.6520
<i>Mali Lošinj</i>	0.1820	-0.3890	-0.3004	-0.3225	-0.3035



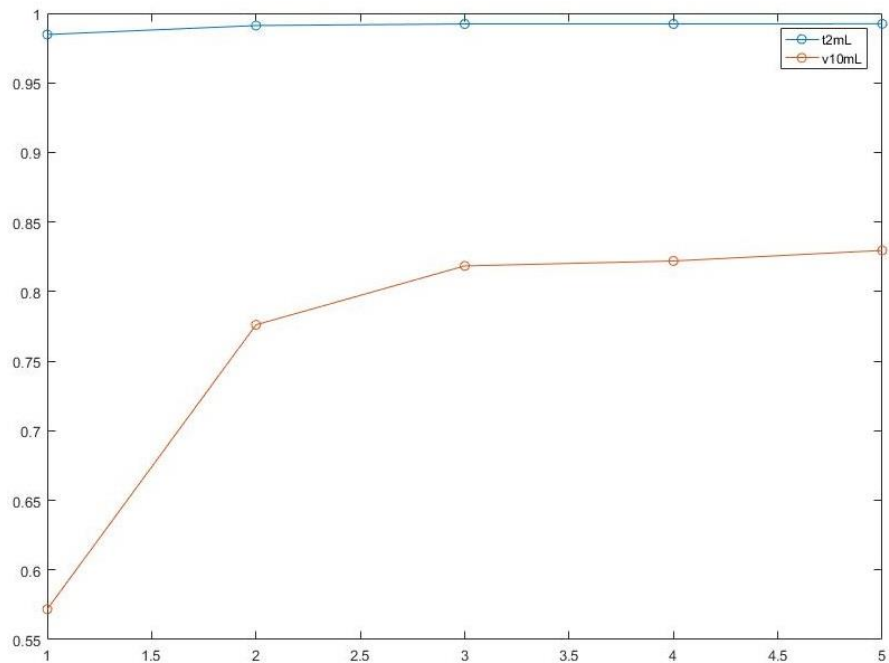
Slika 5. Koeficijenti regresije s redom uključenom jednom, dvije, ..., pet komponenti SVD-rastava za temperaturu.



Slika 6. Koeficijenti regresije s redom uključenom jednom, dvije, ..., pet komponenti SVD-rastava za sjevernu komponentu vjetra.

Tablica 5. Koeficijenti korelacije na kvadrat dobiveni kumulativnim uključivanjem komponenti SVD-rastava za temperaturu $corr^2(t)$ i sjevernu komponentu vjetra $corr^2(v)$.

	broj komponente SVD-rastava				
$corr(t)^2$	0.9848	0.9912	0.9923	0.9924	0.9924
$corr(v)^2$	0.5717	0.7761	0.8185	0.8220	0.8296



Slika 7. Koeficijenti korelacije na kvadrat dobiveni kumulativnim uključivanjem komponenti SVD-rastava za temperaturu i sjevernu komponentu vjetra.