Analiza podataka o kvalitetu vazduha

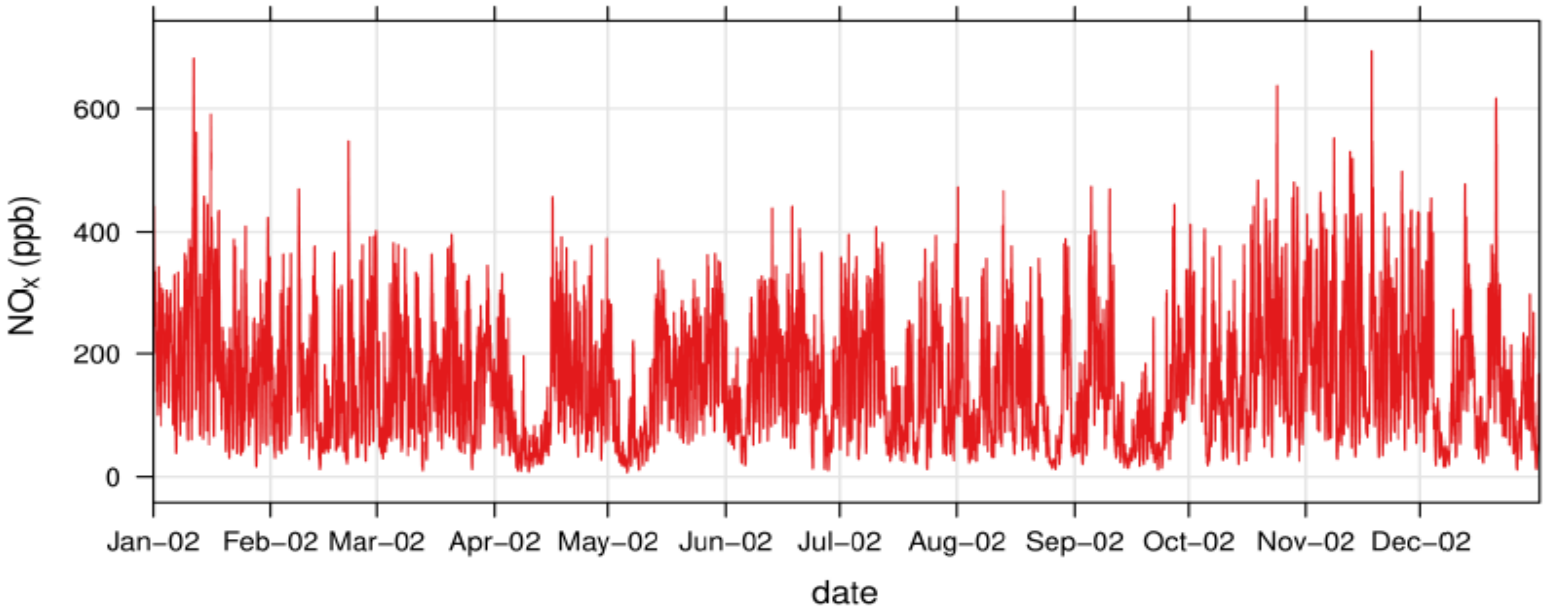
dr Uroš Pešović

Fakultet tehničkih nauka u Čačku,

6. april 2019. godine

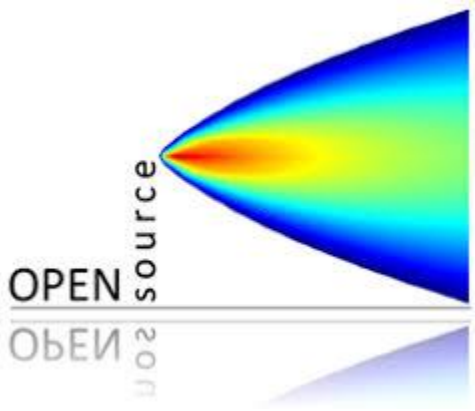
Problem analize podataka

- Kako izdvojiti korisne podatke iz ovog grafikona?



OPENAIR

- Nastao kao rezultat projekta čiji je cilj kreiranje besplatnog open-source alata za analizu podataka o kvalitetu vazuha
- Alat je implementiran u programskom jeziku R
- Osnovni zahtev
  - Bez potrebe za poznavanjem programskih jezika
- Učitavanje biblioteke
  - library(openair)



Učitavanje podataka

- Učitavamo godišnje podatke za sve stanice

```
air.17 = read.csv("D:/Air/air 17.csv", sep=";")
```

air.17

Filter

	date_time	station_id	component_id	value
1	2017-01-01 00:00:00	1	7	6.5242026
2	2017-01-01 00:00:00	1	10	3.5565497
3	2017-01-01 00:00:00	1	10051	0.5683699
4	2017-01-01 00:00:00	1	10052	134.4015223
5	2017-01-01 00:00:00	1	10053	1022.9877333
6	2017-01-01 00:00:00	1	10054	-4.0506784
7	2017-01-01 00:00:00	1	10058	87.2249548
8	2017-01-01 00:00:00	2	5	74.2401382
9	2017-01-01 00:00:00	2	8	22.5384779
10	2017-01-01 00:00:00	2	9	101.6185639
11	2017-01-01 00:00:00	2	38	51.5926663
12	2017-01-01 00:00:00	2	6001	64.5539485
13	2017-01-01 00:00:00	2	6002	67.7450678
14	2017-01-01 00:00:00	2	9001	70.0383315
15	2017-01-01 00:00:00	2	10051	0.7212432
16	2017-01-01 00:00:00	2	10052	203.6020568
17	2017-01-01 00:00:00	2	10053	1020.5804167
18	2017-01-01 00:00:00	2	10054	-0.5521223

Izdvajanje podataka stanice

- Potrebno je izdvojiti podatke za odgovarajuću stanicu prema njenom ID – u
- Spisak stanica sa njihovim podacima je dostupan u csv fajlu stations.csv

- Učitaćemo podatke za mernu stanicu Čačak

```
Cacak = air.17 %>% filter(station_id == 31)
```

Podaci sa merne stanice Čačak

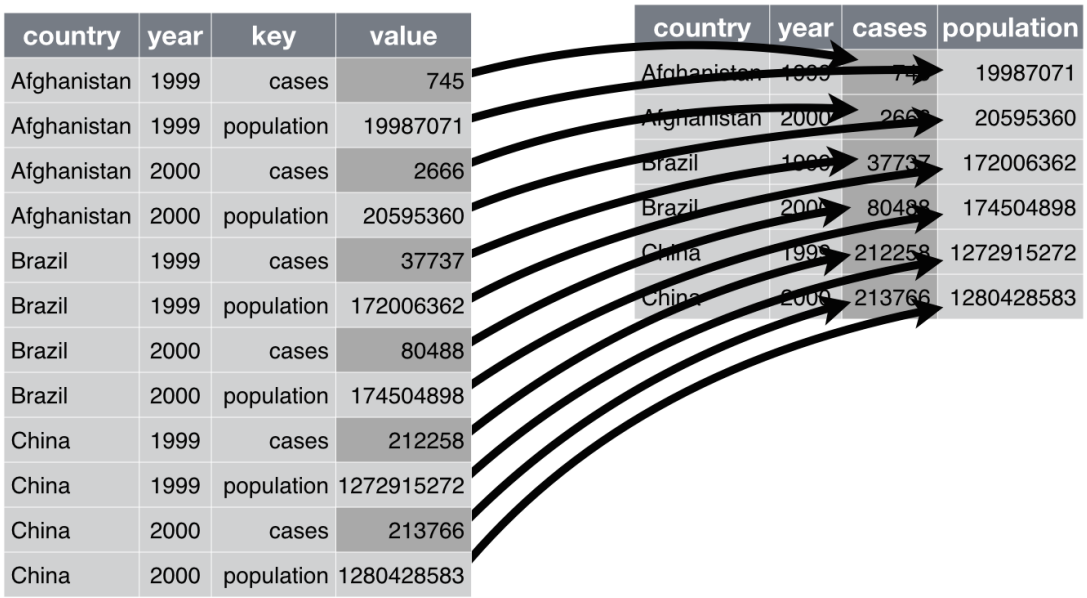
Cacak

Filter

	date_time	station_id	component_id	value
1	2017-01-01 00:00:00	31	1	17.8617716
2	2017-01-01 00:00:00	31	8	4.9690235
3	2017-01-01 00:00:00	31	9	6.7509427
4	2017-01-01 00:00:00	31	10	0.9850734
5	2017-01-01 00:00:00	31	38	1.1570000
6	2017-01-01 01:00:00	31	1	16.5961728
7	2017-01-01 01:00:00	31	8	11.8815001
8	2017-01-01 01:00:00	31	9	35.5754233
9	2017-01-01 01:00:00	31	10	2.8601467
10	2017-01-01 01:00:00	31	38	15.4571735
11	2017-01-01 02:00:00	31	1	15.5851622
12	2017-01-01 02:00:00	31	8	10.3376301
13	2017-01-01 02:00:00	31	9	23.7170371
14	2017-01-01 02:00:00	31	10	1.8624599
15	2017-01-01 02:00:00	31	38	8.7304532
16	2017-01-01 03:00:00	31	1	13.4180328
17	2017-01-01 03:00:00	31	8	9.6859161
18	2017-01-01 03:00:00	31	9	27.2986755

Transformacija podataka

- Potrebno je izvršiti transformaciju tabele postavljanjem izmerenih vrednosti prikazanih po različitim redovima u jedinstvene kolone



country	year	key	value
Afghanistan	1999	cases	745
Afghanistan	1999	population	19987071
Afghanistan	2000	cases	2666
Afghanistan	2000	population	20595360
Brazil	1999	cases	37737
Brazil	1999	population	172006362
Brazil	2000	cases	80488
Brazil	2000	population	174504898
China	1999	cases	212258
China	1999	population	1272915272
China	2000	cases	213766
China	2000	population	1280428583

table2

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

Transformacija podataka

- Za transformaciju podataka koristimo biblioteku tinyvesre koju je potrebno učitati sledećom naredbom

  library(tidyverse)

- Potrebno je izvršiti transformaciju podataka po nazivu merene komponente component_id

  Cacak.Data = Cacak %>% spread(key = component_id, value = value)

Transformisani podaci za mernu stanicu Čačak

Cacak | Cacak.Data

Filter

	date_time	station_id	1	8	9	10	38	10053	10054	10058
1	2017-01-01 00:00:00	31	17.861772	4.9690235	6.7509427	0.9850734	1.156999998	NA	NA	NA
2	2017-01-01 01:00:00	31	16.596173	11.8815001	35.5754233	2.8601467	15.457173451	NA	NA	NA
3	2017-01-01 02:00:00	31	15.585162	10.3376301	23.7170371	1.8624599	8.730453202	NA	NA	NA
4	2017-01-01 03:00:00	31	13.418033	9.6859161	27.2986755	2.1334124	11.491306570	NA	NA	NA
5	2017-01-01 04:00:00	31	14.437927	6.6008335	14.7620086	1.8636767	5.328786653	NA	NA	NA
6	2017-01-01 05:00:00	31	13.467818	5.2155749	8.3907232	1.0929868	2.067866655	NA	NA	NA
7	2017-01-01 06:00:00	31	14.237269	8.2156398	30.7225025	2.6537199	14.684106486	NA	NA	NA
8	2017-01-01 07:00:00	31	16.430675	9.0102905	35.7576797	2.4680720	17.447907568	NA	NA	NA
9	2017-01-01 08:00:00	31	20.774292	12.4232227	34.0326169	2.2669698	14.093906299	NA	NA	NA
10	2017-01-01 09:00:00	31	36.382119	14.4859667	27.0595633	1.7266553	8.206466382	NA	NA	NA
11	2017-01-01 10:00:00	31	29.731996	8.2685103	12.2640798	1.0515892	2.607800019	NA	NA	NA
12	2017-01-01 11:00:00	31	27.191185	8.2233444	11.4601308	1.0300094	2.112933289	NA	NA	NA
13	2017-01-01 12:00:00	31	26.876086	9.2858011	13.0797196	0.9387266	2.475546632	NA	NA	NA
14	2017-01-01 13:00:00	31	25.947232	8.4438595	11.6391995	0.7910464	2.085546653	NA	NA	NA
15	2017-01-01 14:00:00	31	25.474587	8.2512411	11.7058852	0.7365744	2.251946590	NA	NA	NA
16	2017-01-01 15:00:00	31	43.828333	14.1623680	20.1611687	0.8997914	3.901559979	NA	NA	NA
17	2017-01-01 16:00:00	31	21.961862	20.3155296	27.2269443	1.4643667	4.511866632	NA	NA	NA
18	2017-01-01 17:00:00	31	20.487331	16.3964760	23.8530665	3.1341975	4.873093202	NA	NA	NA

Promena naziva kolona

- Potrebno je dodeliti nazive kolonama koje ćemo koristiti za obraćanje kolonama

colnames(Cacak.Data) = c("date","station_id","SO2","NO2","NOx","CO","NO","P","t","RH")

Podaci za mernu stanicu Čačak

Cacak | Cacak.Data

Filter

	date	station_id	SO2	NO2	NOx	CO	NO	P	t	RH
1	2017-01-01 00:00:00	31	17.861772	4.9690235	6.7509427	0.9850734	1.156999998	NA	NA	NA
2	2017-01-01 01:00:00	31	16.596173	11.8815001	35.5754233	2.8601467	15.457173451	NA	NA	NA
3	2017-01-01 02:00:00	31	15.585162	10.3376301	23.7170371	1.8624599	8.730453202	NA	NA	NA
4	2017-01-01 03:00:00	31	13.418033	9.6859161	27.2986755	2.1334124	11.491306570	NA	NA	NA
5	2017-01-01 04:00:00	31	14.437927	6.6008335	14.7620086	1.8636767	5.328786653	NA	NA	NA
6	2017-01-01 05:00:00	31	13.467818	5.2155749	8.3907232	1.0929868	2.067866655	NA	NA	NA
7	2017-01-01 06:00:00	31	14.237269	8.2156398	30.7225025	2.6537199	14.684106486	NA	NA	NA
8	2017-01-01 07:00:00	31	16.430675	9.0102905	35.7576797	2.4680720	17.447907568	NA	NA	NA
9	2017-01-01 08:00:00	31	20.774292	12.4232227	34.0326169	2.2669698	14.093906299	NA	NA	NA
10	2017-01-01 09:00:00	31	36.382119	14.4859667	27.0595633	1.7266553	8.206466382	NA	NA	NA
11	2017-01-01 10:00:00	31	29.731996	8.2685103	12.2640798	1.0515892	2.607800019	NA	NA	NA
12	2017-01-01 11:00:00	31	27.191185	8.2233444	11.4601308	1.0300094	2.112933289	NA	NA	NA
13	2017-01-01 12:00:00	31	26.876086	9.2858011	13.0797196	0.9387266	2.475546632	NA	NA	NA
14	2017-01-01 13:00:00	31	25.947232	8.4438595	11.6391995	0.7910464	2.085546653	NA	NA	NA
15	2017-01-01 14:00:00	31	25.474587	8.2512411	11.7058852	0.7365744	2.251946590	NA	NA	NA
16	2017-01-01 15:00:00	31	43.828333	14.1623680	20.1611687	0.8997914	3.901559979	NA	NA	NA
17	2017-01-01 16:00:00	31	21.961862	20.3155296	27.2269443	1.4643667	4.511866632	NA	NA	NA
18	2017-01-01 17:00:00	31	20.487331	16.3964760	23.8530665	3.1341975	4.873093202	NA	NA	NA

Konverzija datuma i vremena

- Datum i vreme su prikazani u formatu kao faktor
- Potrebno je konvertovati u format za datum i vreme prema POSIXct formatu

  Cacak.Data$date = as.POSIXct(Cacak.Data$date, format = "%Y-%m-%d %H:%M:%S")
  - gde je Y četvorocifrena godina
  - m – mesec
  - d – dan u mesecu
  - H – čas
  - M – minut
  - S - sekund

Pregled parametara tabele

- Statistički pregled parametara

  summary(Cacak.Data)

```
> summary(Cacak.Data)
      date                       station_id      SO2              NO2              NOx
 Min.   :2017-01-01 00:00:00   Min.   :31   Min.   :  6.213   Min.   : 0.00   Min.   :  0.000
 1st Qu.:2017-03-26 00:45:00   1st Qu.:31   1st Qu.: 15.100   1st Qu.: 3.53   1st Qu.:  5.801
 Median :2017-07-14 02:30:00   Median :31   Median : 19.207   Median : 7.15   Median : 10.805
 Mean   :2017-07-05 09:06:04   Mean   :31   Mean   : 24.685   Mean   :10.31   Mean   : 18.942
 3rd Qu.:2017-10-06 03:15:00   3rd Qu.:31   3rd Qu.: 28.584   3rd Qu.:13.70   3rd Qu.: 21.734
 Max.   :2017-12-31 23:00:00   Max.   :31   Max.   :144.941   Max.   :91.09   Max.   :550.206
                                            NA's   :7431      NA's   :2471    NA's   :2471
       CO                NO                 P                t                RH
 Min.   :0.0000   Min.   : -2.337   Min.   :0      Min.   :0      Min.   :0
 1st Qu.:0.2232   1st Qu.:  1.483   1st Qu.:0      1st Qu.:0      1st Qu.:0
 Median :0.4225   Median :  2.335   Median :0      Median :0      Median :0
 Mean   :0.6658   Mean   :  5.636   Mean   :0      Mean   :0      Mean   :0
 3rd Qu.:0.8015   3rd Qu.:  4.577   3rd Qu.:0      3rd Qu.:0      3rd Qu.:0
 Max.   :8.9061   Max.   :299.491   Max.   :0      Max.   :0      Max.   :0
                  NA's   :2471      NA's   :7742   NA's   :7742   NA's   :7742
>
```

Numerička analiza podataka

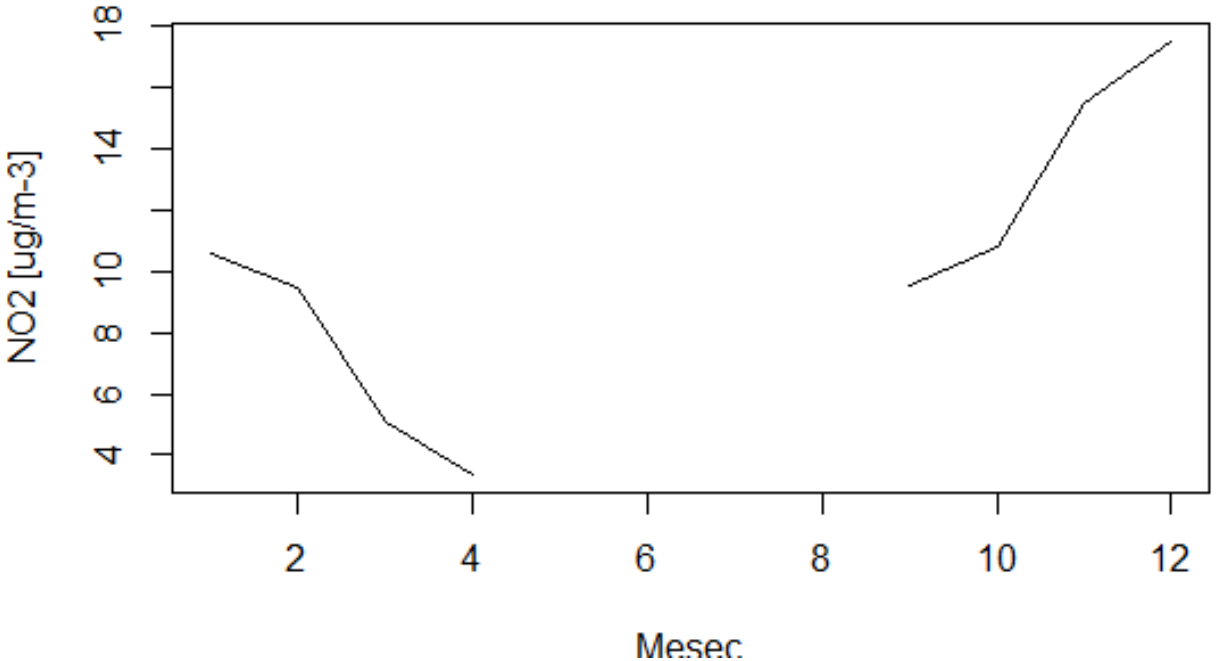
- Za numeričku analizu koristimo funkciju

  aggregate(x, by, FUN, ...,)

- Funkcija **aggregate** deli podatke **x** u podskupove definisane kriterijum **by** i nad njima primenjuje određenu funkciju fun
- Mogu se primenti funkcije za
  - Srednju vrednost – mean
  - Minimum – min
  - Maksimum – max
- Grupisani rezltati numeričke obrade se prikazuju u formi tabele

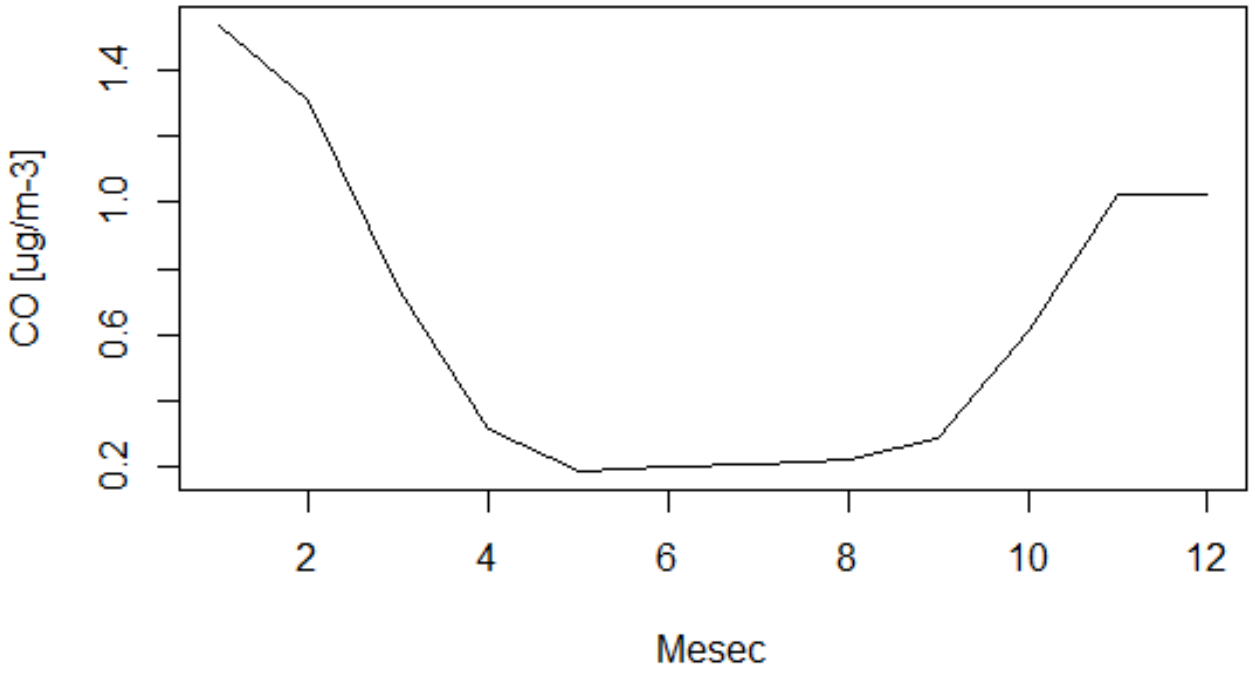
Prosečne vrednosti NO2 po mesecima

- monthly_NO2 = aggregate(Cacak.Data["NO2"], format(Cacak.Data["date"],"%Y-%m"), mean, na.rm = TRUE)
- plot(monthly_NO2 $NO2,type="l", xlab="Mesec", ylab="NO2 [ug/m-3]")



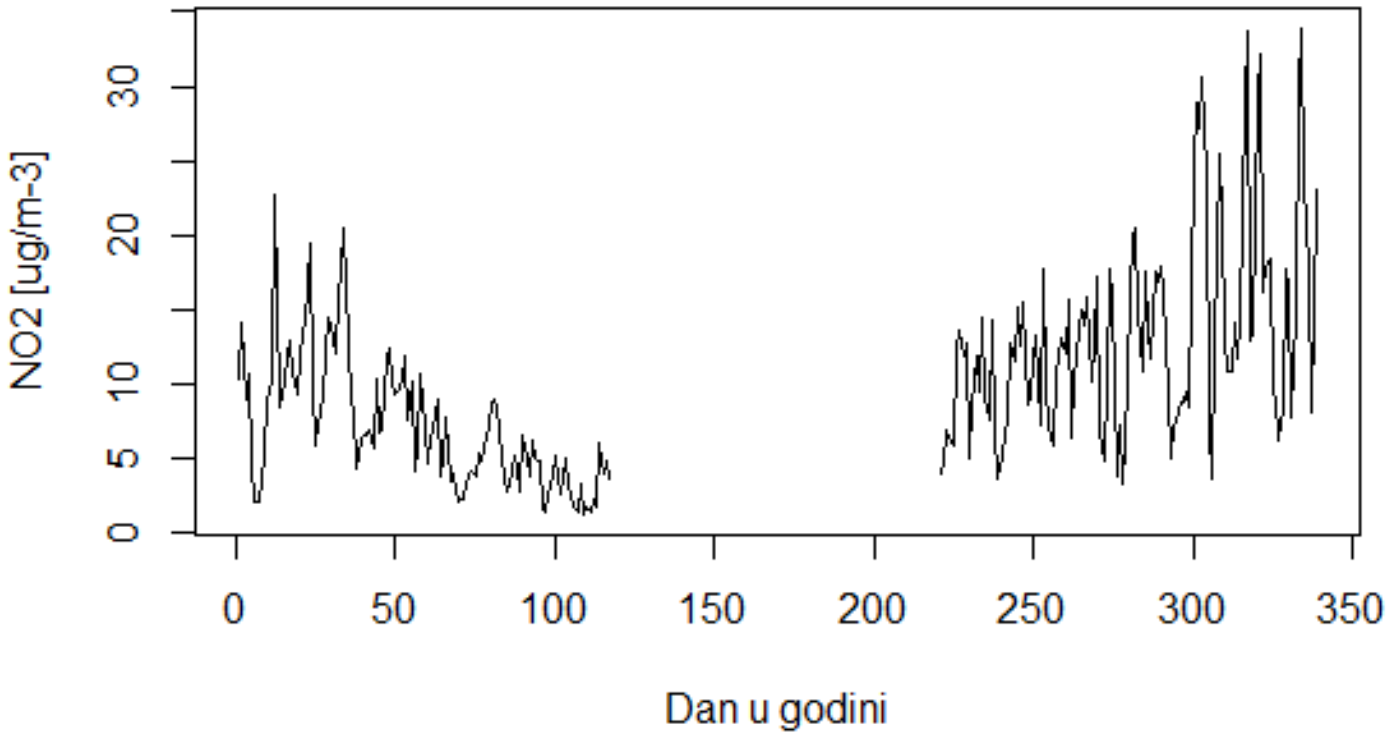
Prosečne vrednosti CO po mesecima

- monthly_CO = aggregate(Cacak.Data["CO"], format(Cacak.Data["date"],"%Y-%m"), mean, na.rm = TRUE)
- plot(monthly_CO $CO,type="l", xlab="Mesec", ylab="CO [ug/m-3]")



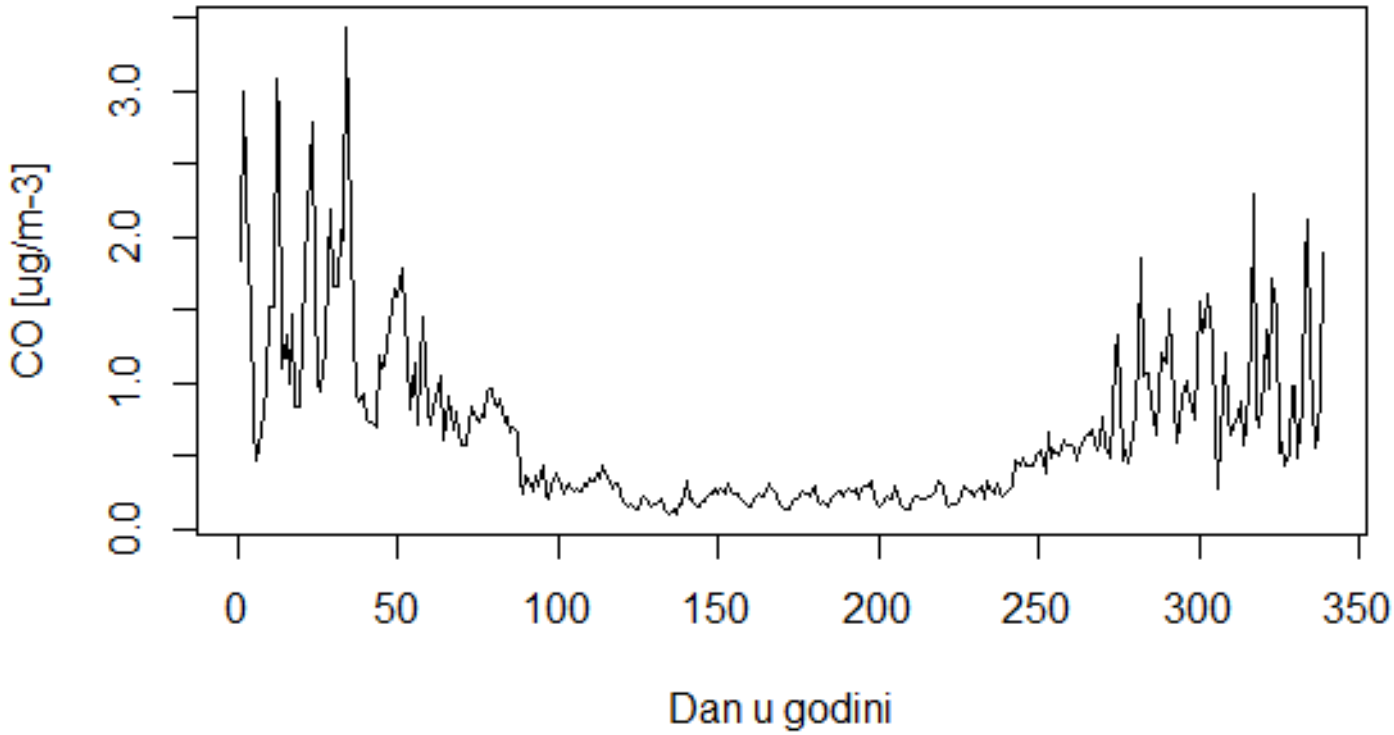
Prosečne vrednosti NO2 po danima

- daily_NO2 = aggregate(Cacak.Data["NO2"], format(Cacak.Data["date"],"%Y-%j"), mean, na.rm = TRUE)
- plot(daily_NO2$NO2,type="l", xlab="Dan u godini", ylab="NO2 [ug/m-3]")



Prosečne vrednosti CO po danima

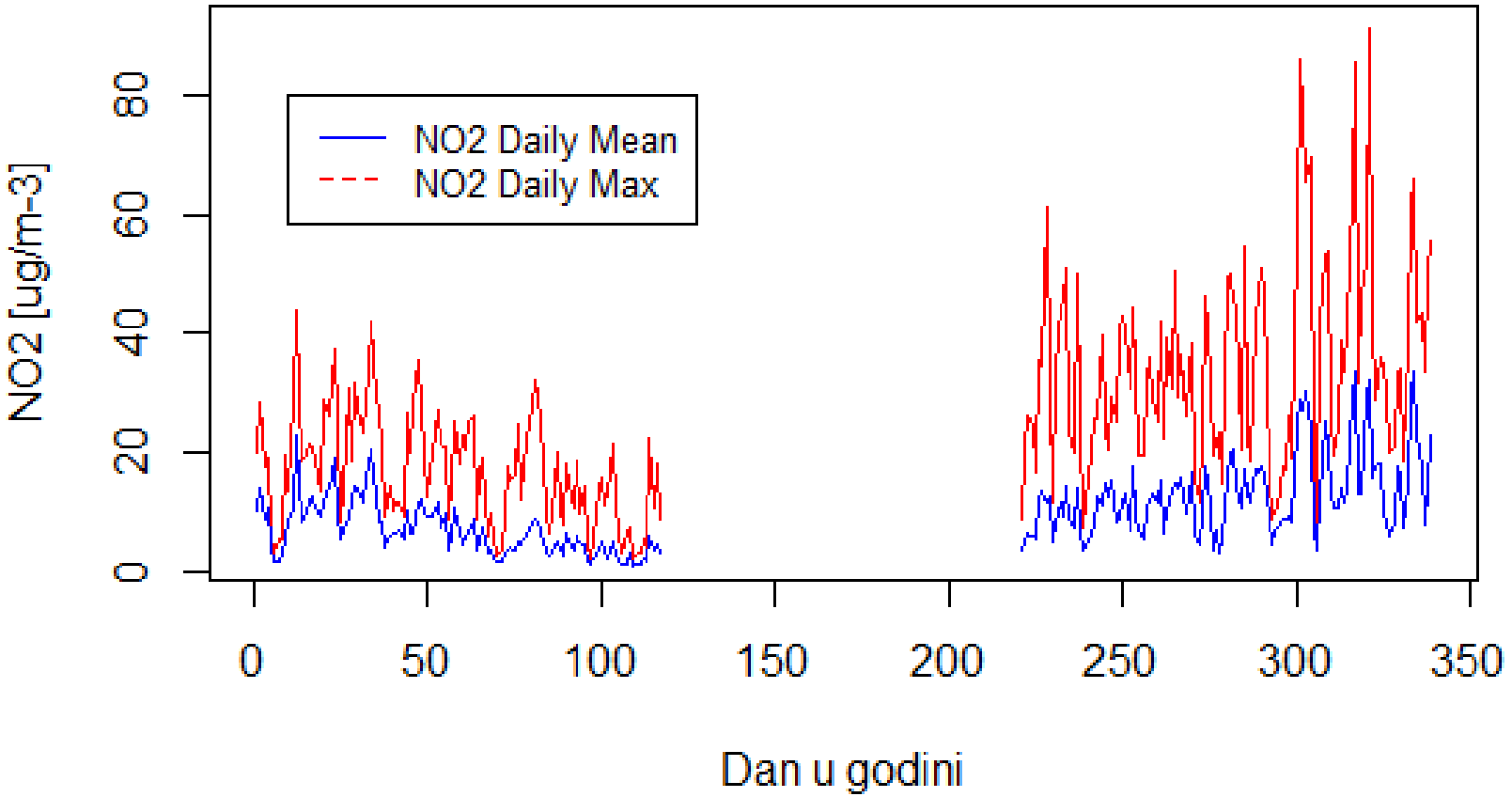
- daily_CO <- aggregate(Cacak.Data["CO"], format(Cacak.Data["date"],"%Y-%j"), mean, na.rm = TRUE)
- plot(daily_CO$CO,type="l", xlab="Dan u godini", ylab="CO [ug/m-3]")



Maksimalne dnevne vrednosti NO2

- daily_max_NO2 = aggregate(Cacak.Data["NO2"], format(Cacak.Data["date"],"%Y-%j"), max, na.rm = TRUE)
- plot(daily_max_NO2$NO2,type="l",col="red", xlab="Dan u godini", ylab="NO2 [ug/m-3]")
- lines(daily_NO2$NO2,type="l",col="blue", xlab="Dan u godini", ylab="NO2 [ug/m-3]")
- legend(10, 80, legend=c("NO2 Daily Mean", "NO2 Daily Max"), + col=c("blue", "red"), lty=1:2, cex=0.8)

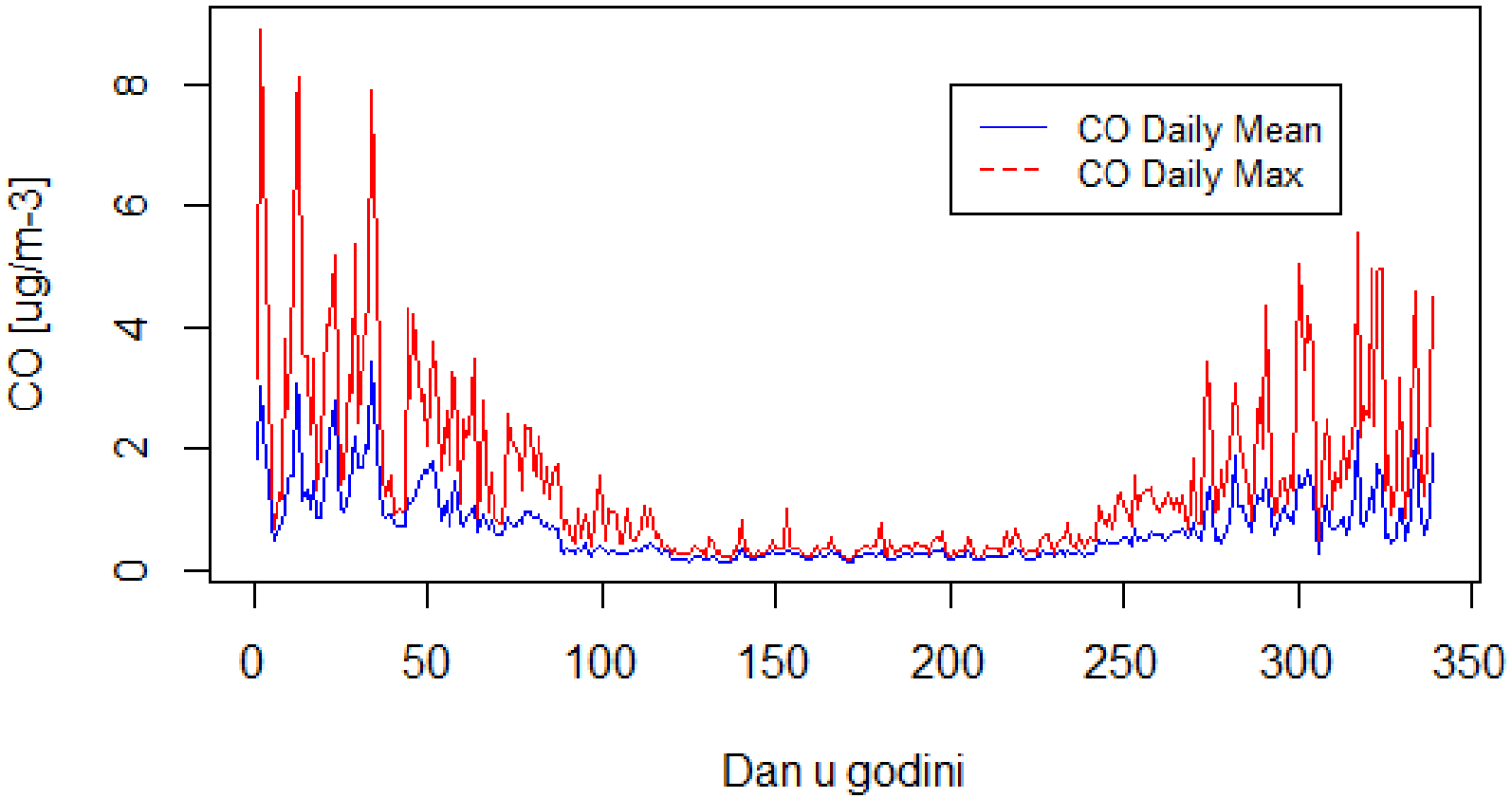
Maksimalne dnevne vrednosti NO2



Maksimalne dnevne vrednosti CO

- daily_max_CO <- aggregate(Cacak.Data["CO"], format(Cacak.Data["date"],"%Y-%j"), max, na.rm = TRUE)
- plot(daily_max_CO$CO,type="l",col="red", xlab="Dan u godini", ylab="CO [ug/m-3]")
- lines(daily_CO$CO,type="l",col="blue", xlab="Dan u godini", ylab="CO [ug/m-3]")
- legend(250, 8, legend=c("Mean", "Max"), col=c("blue", "red"), lty=1:2, cex=0.8)

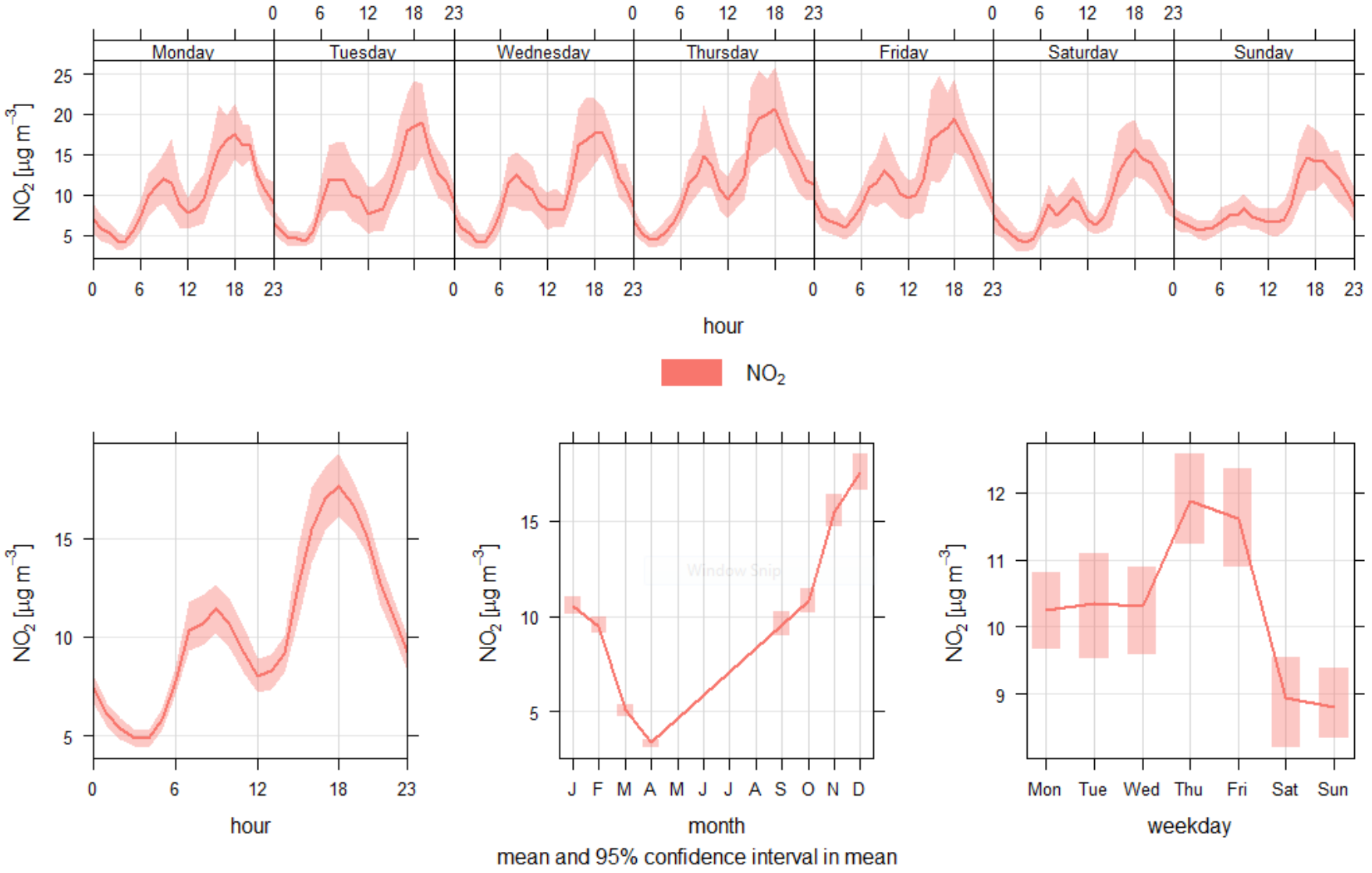
Maksimalne dnevne vrednosti CO



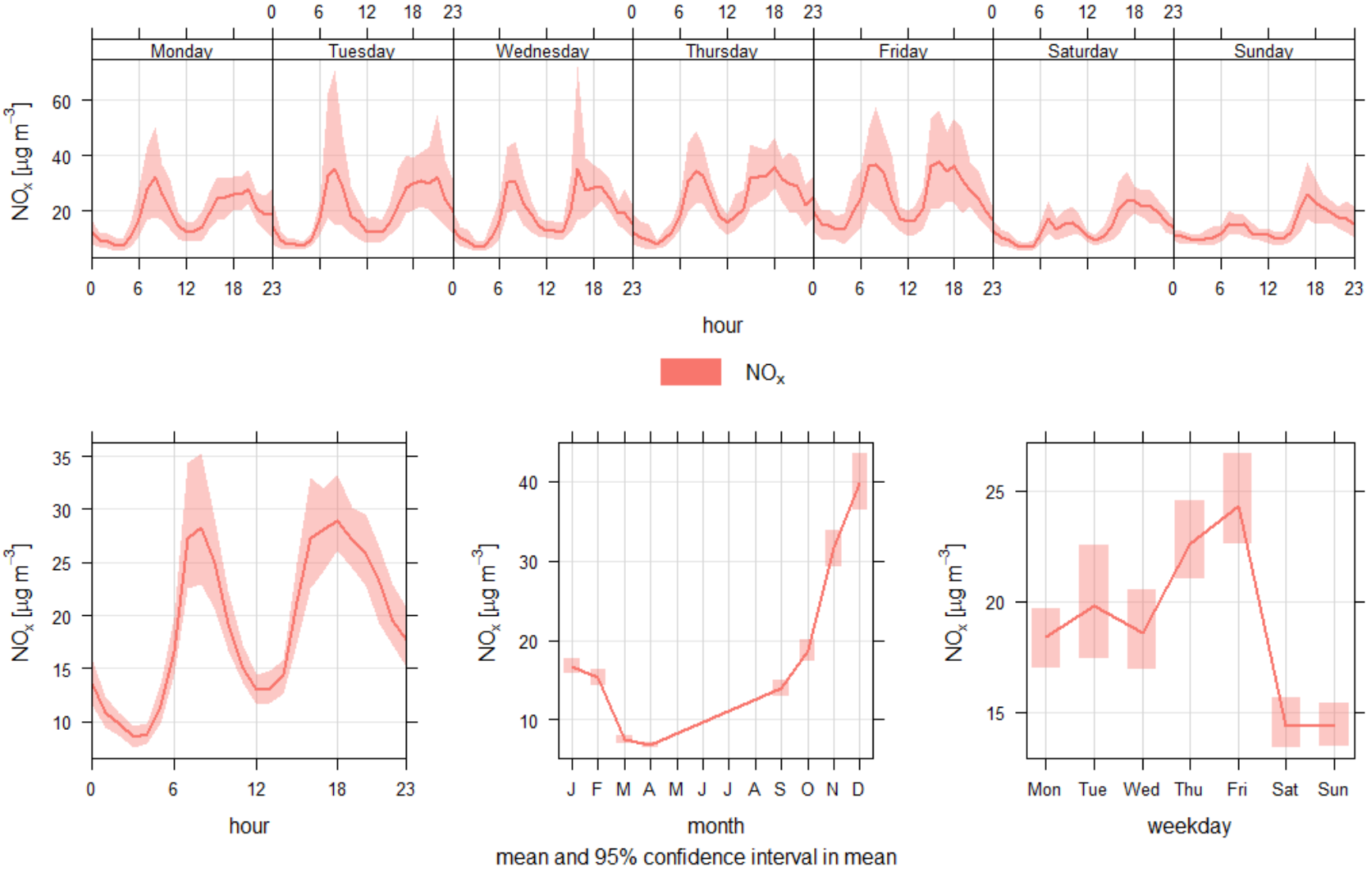
Varijacije parametara u vremenu

- Varijacije parametara u vremenu mogu se prikazati pomoću funkcije timeVariation iz biblioteke openair
- timeVariation(subset(Cacak.Data), pollutant = "NO2", ylab = "NO2 [ug/m-3]")
- timeVariation(subset(Cacak.Data), pollutant = "NOx", ylab = "NOx [ug/m-3]")
- timeVariation(subset(Cacak.Data), pollutant = "CO", ylab = "CO [ug/m-3]")
- timeVariation(Cacak.Data, pollutant = c("NOx", "CO", "NO2"), normalise = TRUE)

VARIJACIJE NO2 PARAMETRA U VREMENU



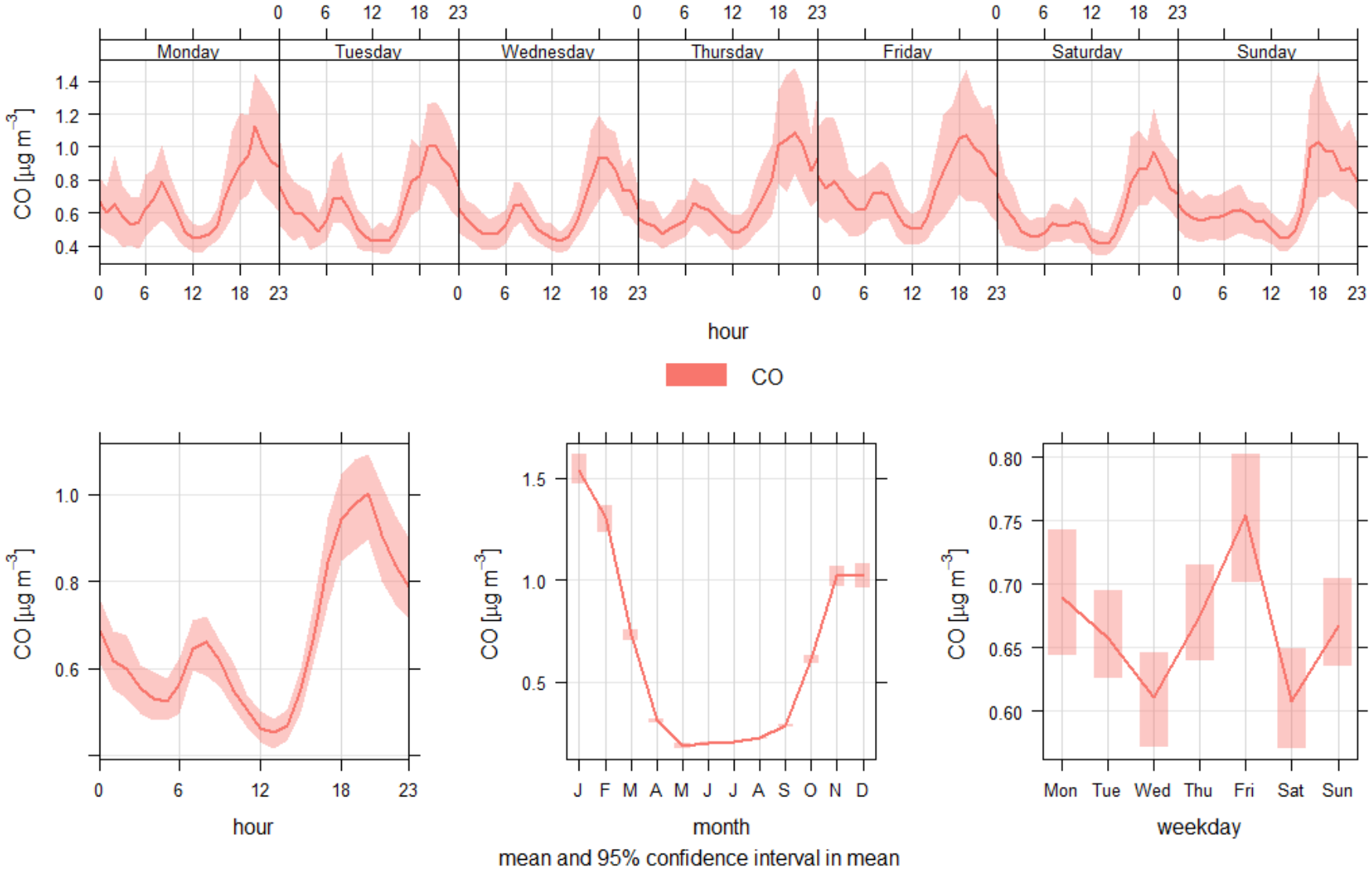
Varijacije NOx parametra u vremenu



Analiza Nox i N02 parametara kvaliteta vazduha

- Azotovi oksidi NOx i NO2 su štetni produkti motora sa unutrašnjim sagorevanjem i podudaraju se sa intenzitetom saobraćaja u gradu
- Najveće dnevne vrednosti su u večernjim časovima
- Koncentracije su jednako izražene radnim danima kao i vikendom
- Na godišnjem nivou najniži su u proleće (april mesec) a najviši u zimskom periodu (decembar)

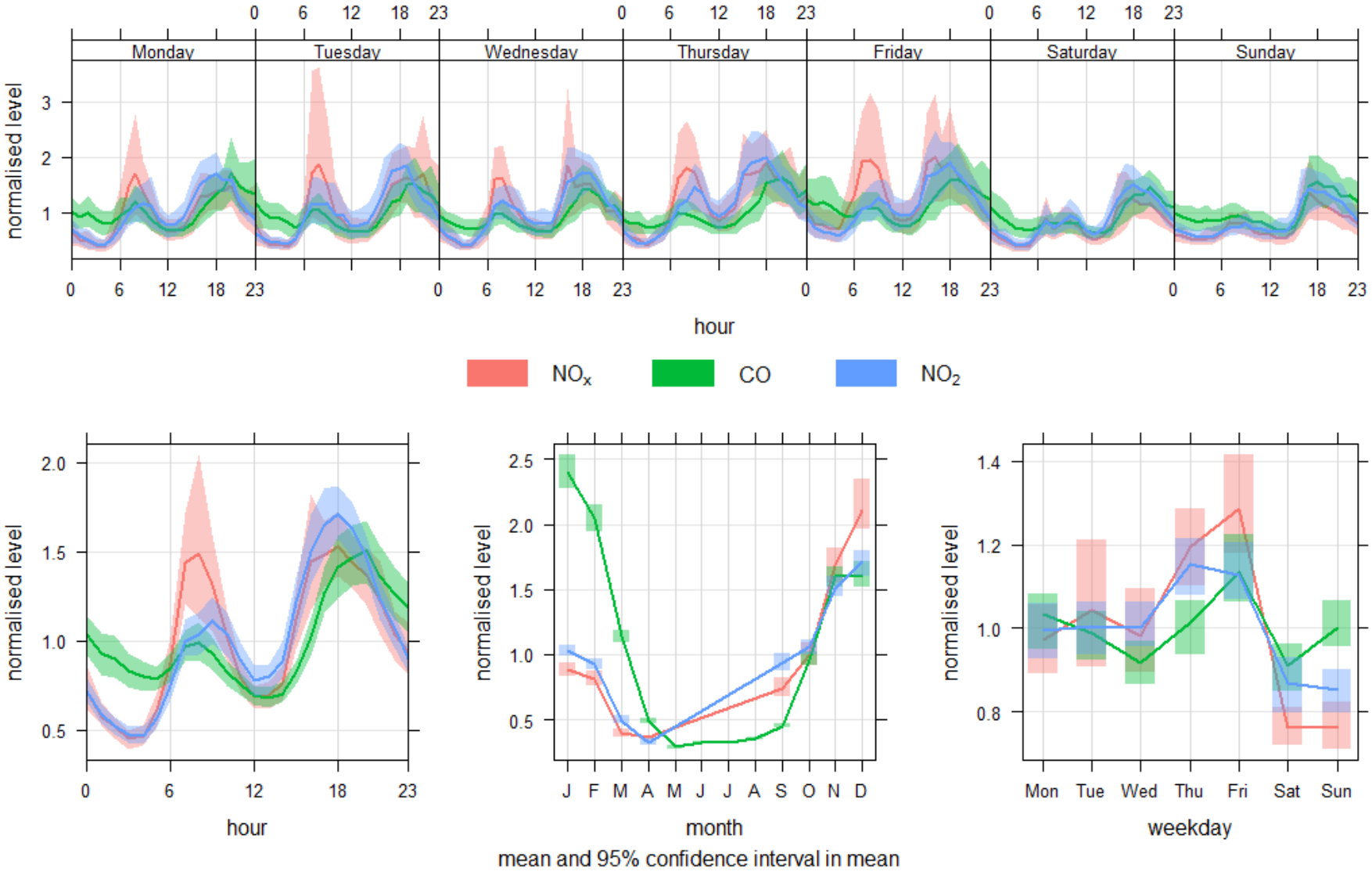
Varijacije CO parametra u vremenu



Analiza CO parametara kvaliteta vazduha

- Ugljenmonoksid CO produkt je nepotpunog sagorevanja čvrstih goriva koji potiču iz domaćinstava i kotlarnica.
- Najveće dnevne vrednosti su u jutarnjim i popodnevnim časovima
- Izraženi su radnim danima dok preko vikenda opadaju približno 25%
- Na godišnjem nivou najniži je tokom letnjih meseci a najviši tokom grejne sezone u zimskom periodu (novembar - februar) kada je u proseku 10 puta veći

Varijacije parametara u vremenu



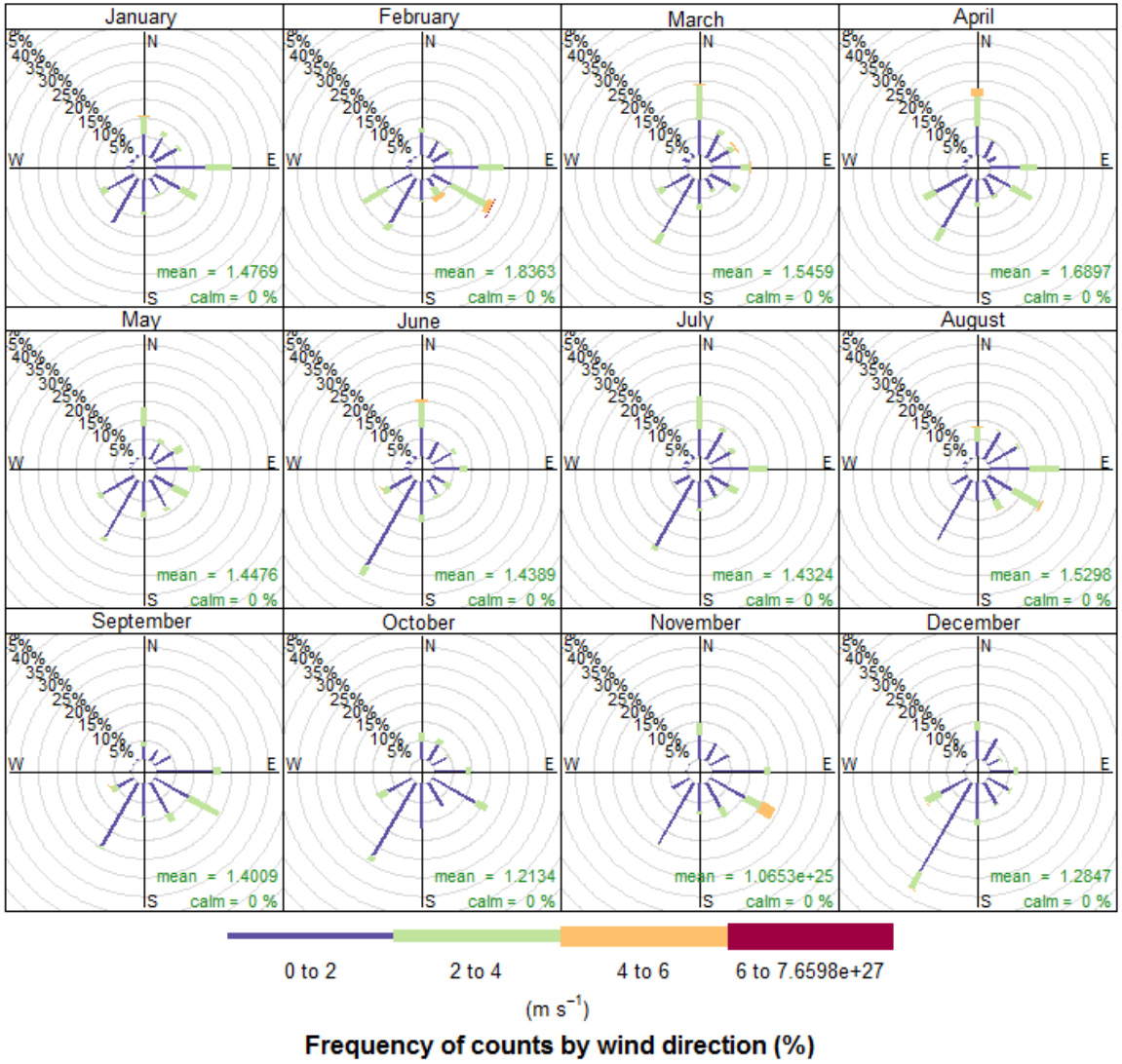
Uticaj smera i brzine vetra

- Potrebno je koristiti podatke sa stanice na kojoj se meri brzina i smer vetra
- Nazvati kolone za brzinu (ws) i smer vetra (wd)
- Učitaćemo podatke za stanicu Novi Sad Spens

```
Spens = air.17 %>% filter(station_id == 2)
```

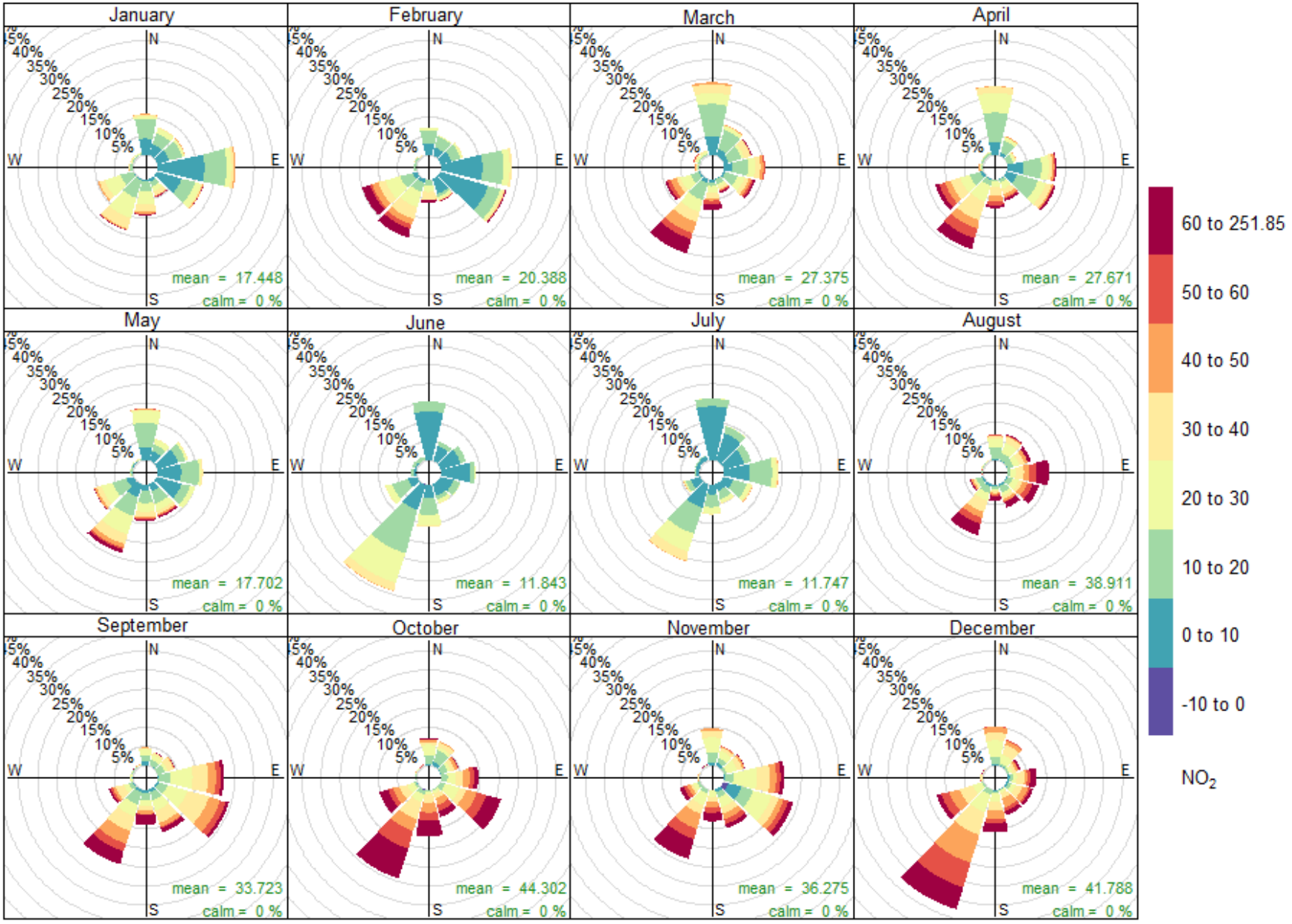
"Ruža" vetrova

- windRose(NSSpens.Data , type = "month")

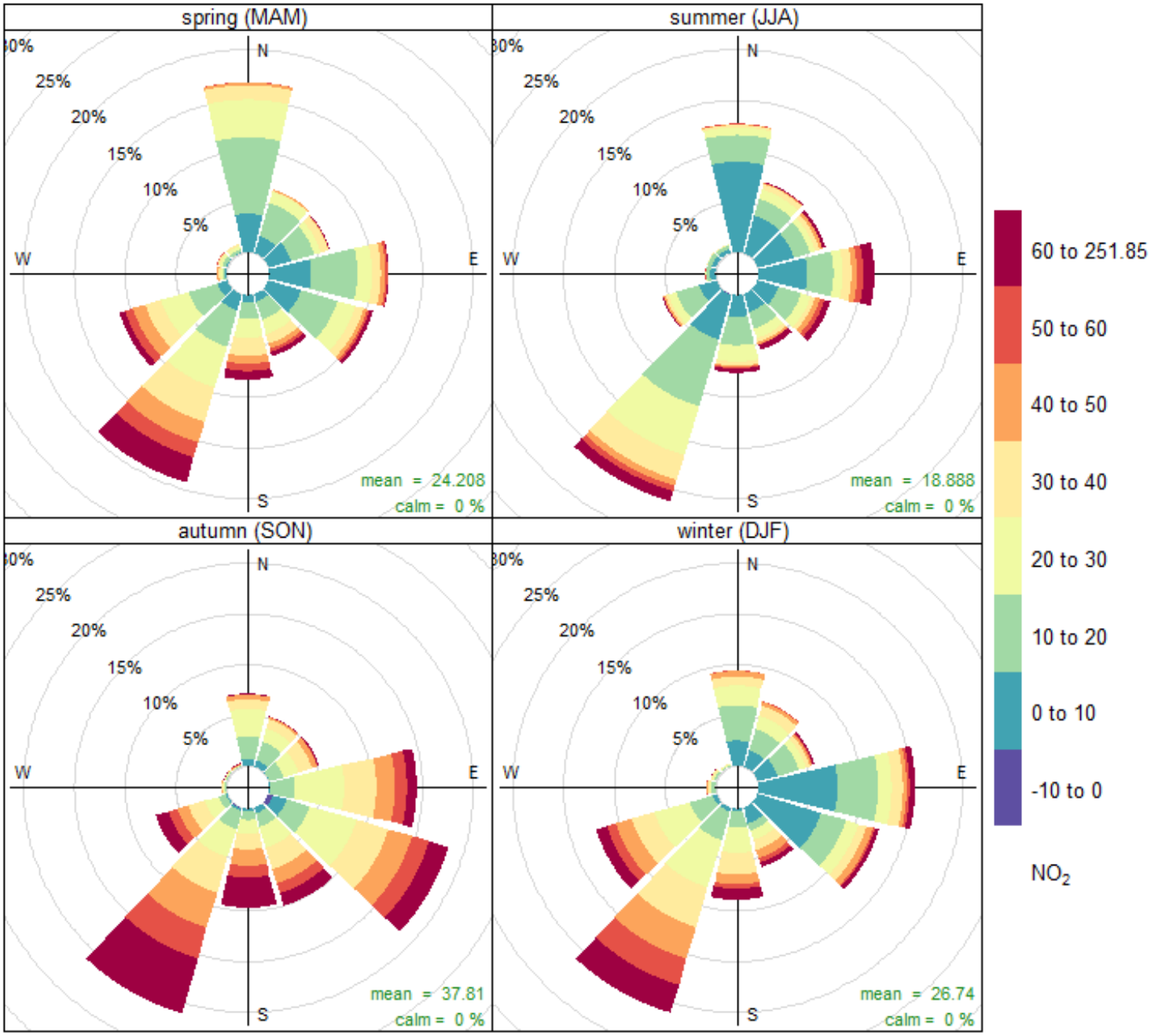


“RUŽA” ZAGAĐENJA

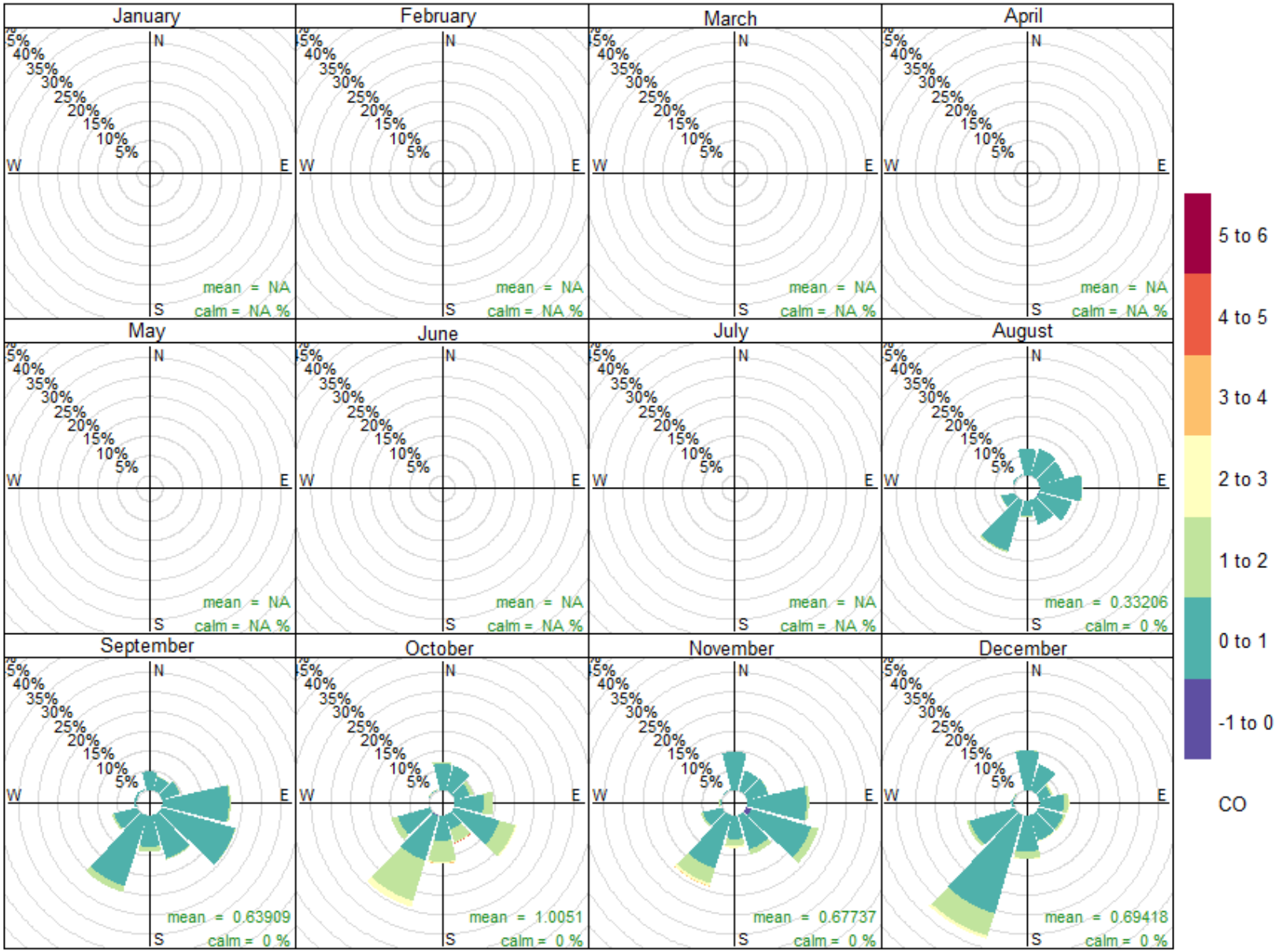
- pollutionRose(NSSpens.Data , pollutant = "NO2", type = "month")
- pollutionRose(NSSpens.Data , pollutant = “NO2", type = "season")
- pollutionRose(NSSpens.Data , pollutant = "CO", type = "month")
- pollutionRose(NSSpens.Data , pollutant = "CO", type = "season")
- pollutionRose(NSSpens.Data , pollutant = "PM10", type = "month")
- pollutionRose(NSSpens.Data , pollutant = "PM10", type = "month")



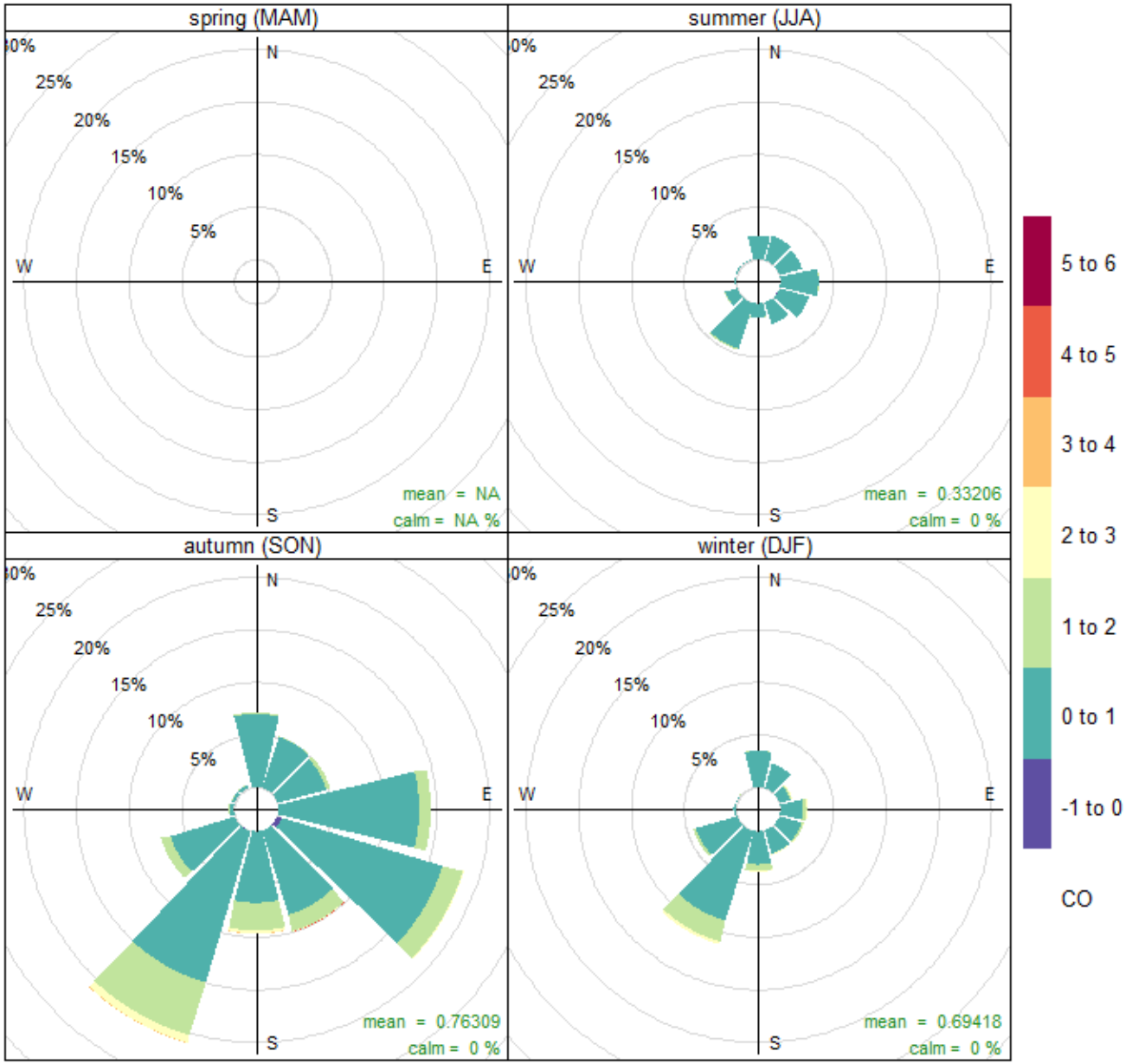
Frequency of counts by wind direction (%)



Frequency of counts by wind direction (%)



Frequency of counts by wind direction (%)



Frequency of counts by wind direction (%)

Zaključak

- Iz priloženih grafičkih prikaza možemo uočiti:
  - Vetar donosi zagađenje iz smerova
    - Jug,
    - Jugoistok,
    - Istok
  - Zagađenje vazduha je najizraženije u zimskim mesecima

HVALA NA PAŽNJI!

VIDIMO SE S NOVIM PODACIMA