Practical exercise with R

Optimal Fingerprint

Qiuzi Han Wen, Francis Zwiers and Xuebin Zhang

July 21-Aug 2nd, 2014

Trieste, Italy

Research Problem

- Background:
  - Changes in high-latitude climate affect the global climate.
  - Warming in the high latitudes is more rapid than in lower latitudes.
  - reductions in snow cover and sea-ice extent, and in thawing permafrost observed
  - strong evidence of anthropogenic contribution to warming in the Arctic land SAT during the past 50 years and in the reduction of Arctic sea ice extent
- Objective: to understand the possible causes of observed changes in high-latitude precipitation.

(Zhang et al., 2014, submitted)

Detection exercise

1. Attempt to detect "ALL" signal in high latitude precipitation

2. Apply the 8-step procedure

3. R codes for different OF algorithms are provided.

Step 1: scale of interest and filtering

- 50°N~90°N
- 1966~2005

- Temporal: 5-year mean
- Spatial: four different spatial configurations (1,2,3, or 6 sub-regions)
  - 1-region: area mean
  - 2-region: southern (50°N~60°N) and northern (60°N~90°N)
  - 3-region: NA (40°W~180°W), WE (30°W~60°E), EE (60°E~180°E)
  - 6-region: SNA, NNA, SWE, NEW, SEE, NEE

Step 2: gather data (OBS)

Observations collected from 3832 stations (450 Canadian stations, 737 Alaskan stations (82 used), 518 Russian stations, 4 Chinese stations, ECA&D dataset)

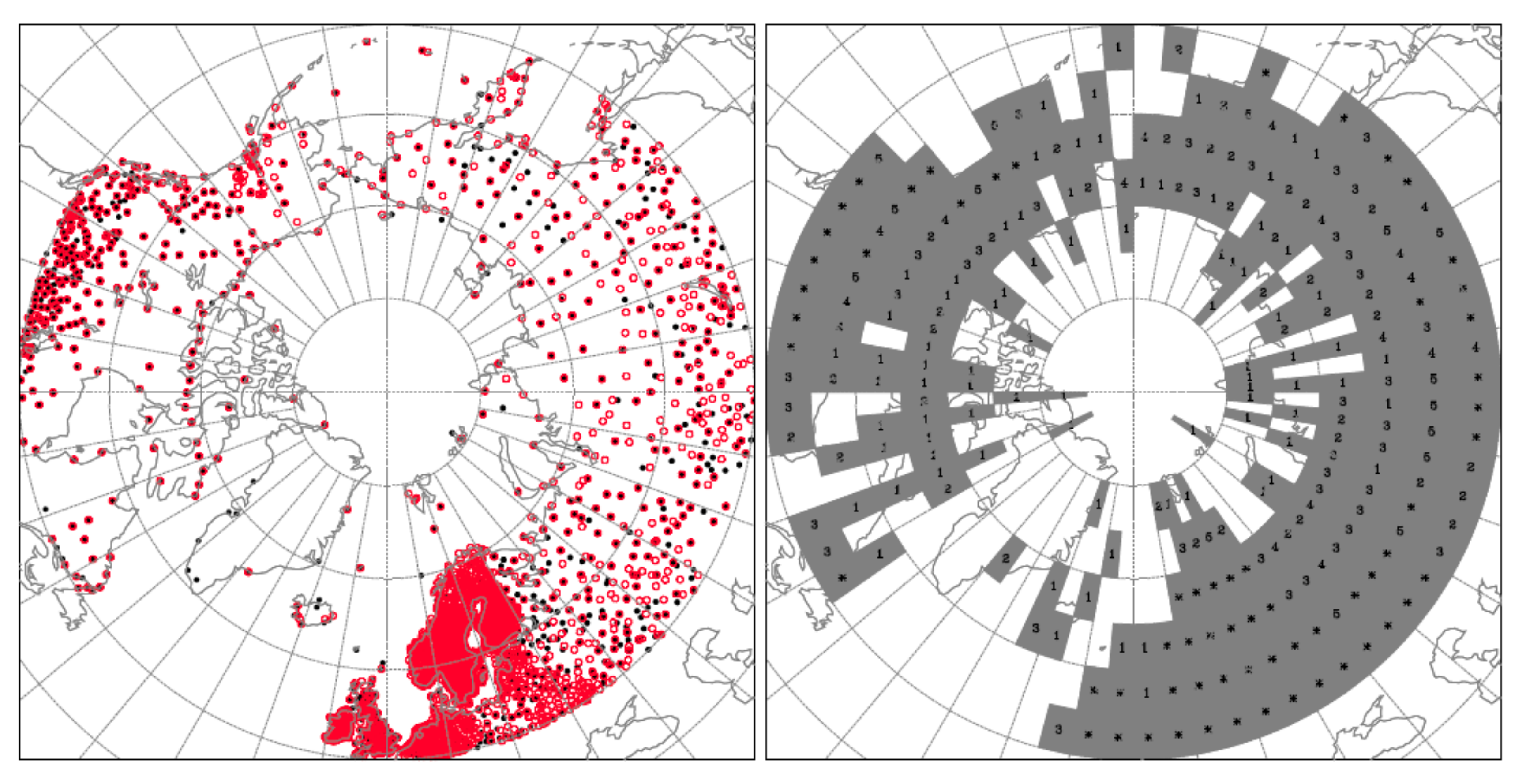


Figure 1, Zhang et al., 2014

Step 2: gather data (model simulations)

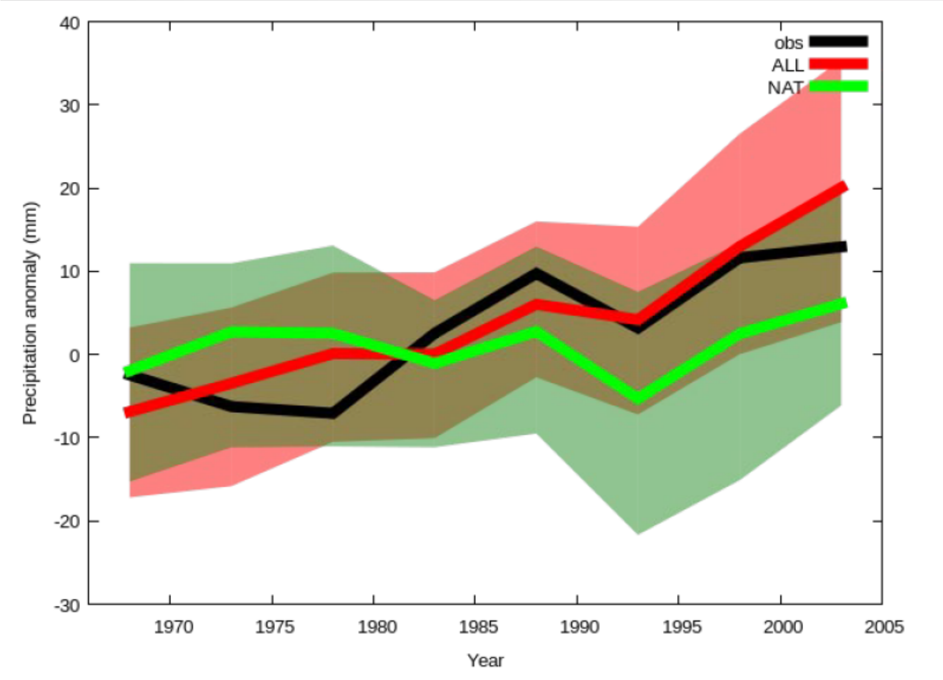
- CMIP5 multi-model ensembles
  - ALL: 32-model ensemble (158 runs)
  - NAT: 14-model ensemble (59 runs)
  - Control runs: 48-model ensemble (over 24,000 year)

Step 3: process data

- Observations
  - Quality control
  - Identify criteria of missing values
  - Calculate anomalies for each station (~66-95)
  - Interpolate to 5°×5° grid boxes

- Model simulations
  - Interpolate to the same spatial resolution, e.g., 5°×5°
  - Extract data of target time period
  - The key point of model data processing: masked by and processed as observations.

Preliminary analysis



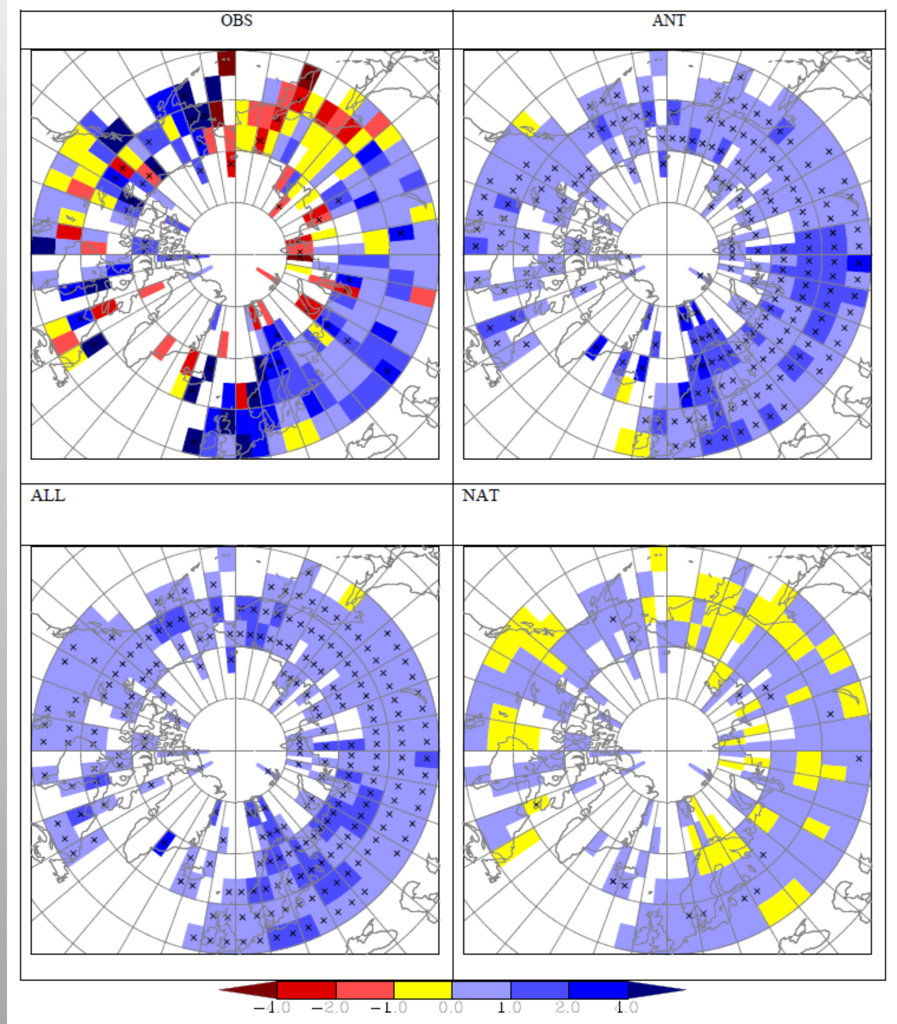


Figure 3, Zhang et al., 2014

Figure 4, Zhang et al., 2014

Steps 1-3 have been completed for you

- **/afs/ictp/public/shared/smr2595/tutorials/Day_1**

- ALL_ann_1area_obs_sig.dat
  - Two rows of 5-yr regional mean anomalies
    - 8 observed anomalies
    - 8 Multi-model ensemble mean anomalies
      - » averaged across 158 ALL-forcings runs
- noise1_05yr_ann_piC_1area.dat
  - Used to estimate variability from internal sources
  - 302 rows, 8 values each
    - 1 row for each 40-yr chunk obtained from control run simulations
      - » centered relative to first 30 years
      - » masked by and processed as observations
- noise2_05yr_ann_piC_1area.dat
  - as above

CMIP5 Model used

Model	ALL
ACCESS1-3	3
Bcc-csm1-1	3
Bcc-csm1-1-m	3
CNRM-CM5	10
CSIRO-Mk3-6-0	10
CanESM2	5
CCSM4	6
CESM1-CAM5-1-FV2	4
CESM1-CAM5	3
CESM1-FASTCHEM	3
EC-EARTH	7
FGOALS-g2	5
FIO-ESM	3
GFDL-CM2p1	10
GFDL-CM3	5
GFDL-ESM2G	3
HadCM3	10
HadGEM2-ES	4
IPSL-CM5A-LR	6
IPSL-CM5A-MR	3
MIROC5	5
MIROC-ESM	3
MPI-ESM-LR	3
MPI-ESM-MR	3
MRI-CGCM3-p1	3
NorESM1-M	3
GISS-E2-R-p1	6
GISS-E2-R-p2	5
GISS-E2-R-p3	6
GISS-E2-H-p1	5
GISS-E2-H-p2	5
GISS-E2-H-p3	5
SUM (models)	158 (32)

Table 1, Zhang et al., 2014

Comparing with CMIP3

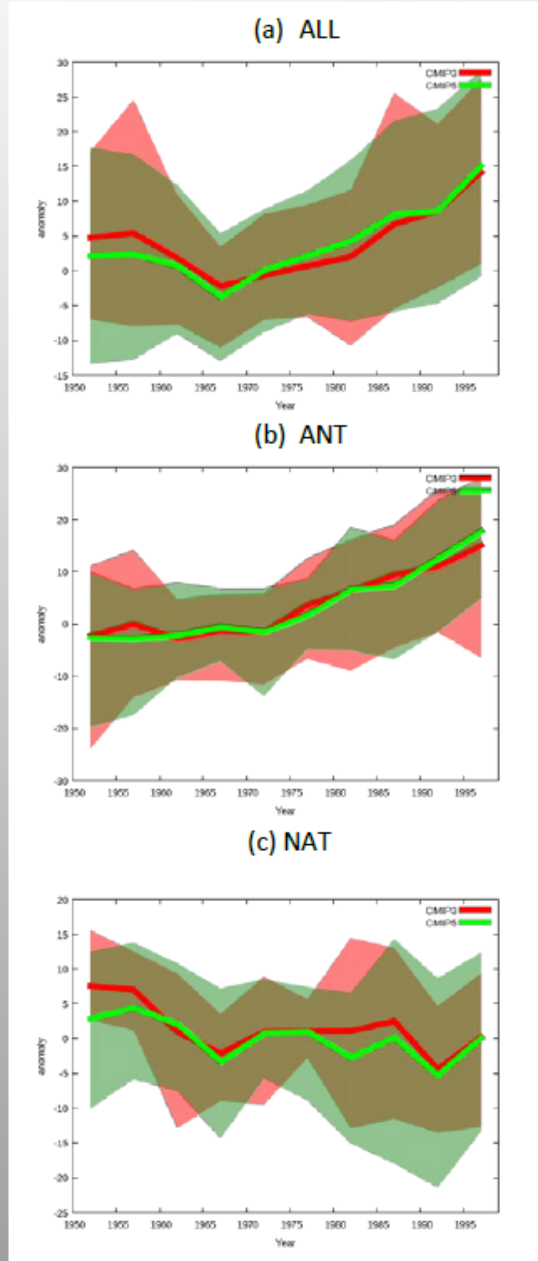


Figure 7, Zhang et al., 2014

Step 4-8

4. Estimate internal variability for Optimization
5. Fit regression model
6. Evaluate the goodness of fit
7. Determine number of EOF truncation
8. Make inferences about scaling factor(s)

These steps have all been coded for you in R

- main functions in ECOF_V1.r
  - readin.r – ingest of data from step 3
  - ols – carries out detection analysis using ordinary least squares
  - tls.A03- carries out detection analysis using total lease squares algorithm
  - tls.ROF-carries out detection analysis using regularized optimal fingerprint
  - plotbetas-visualization of scaling factor estimates
  - plotrstat-visualization of results for residual consistency check

afs/ictp.it/public/shared/smr2595/code/ECOF-package

Suggested activities:

Load functions into R

- Click on “File”
- Click on “Source R Code …”
- Enter the function name to list the function
- source() can also be used to load code

R Console

```
> source('ECOF_V1.r')
> ls()
 [1] "checkOF"    "Creg"       "ols"        "plotbetas" "plotrstat" "readin"
 [7] "redECOF"    "redop"      "redvec"     "tls.A03"   "tls.ROF"
> 
```

Suggested activities:

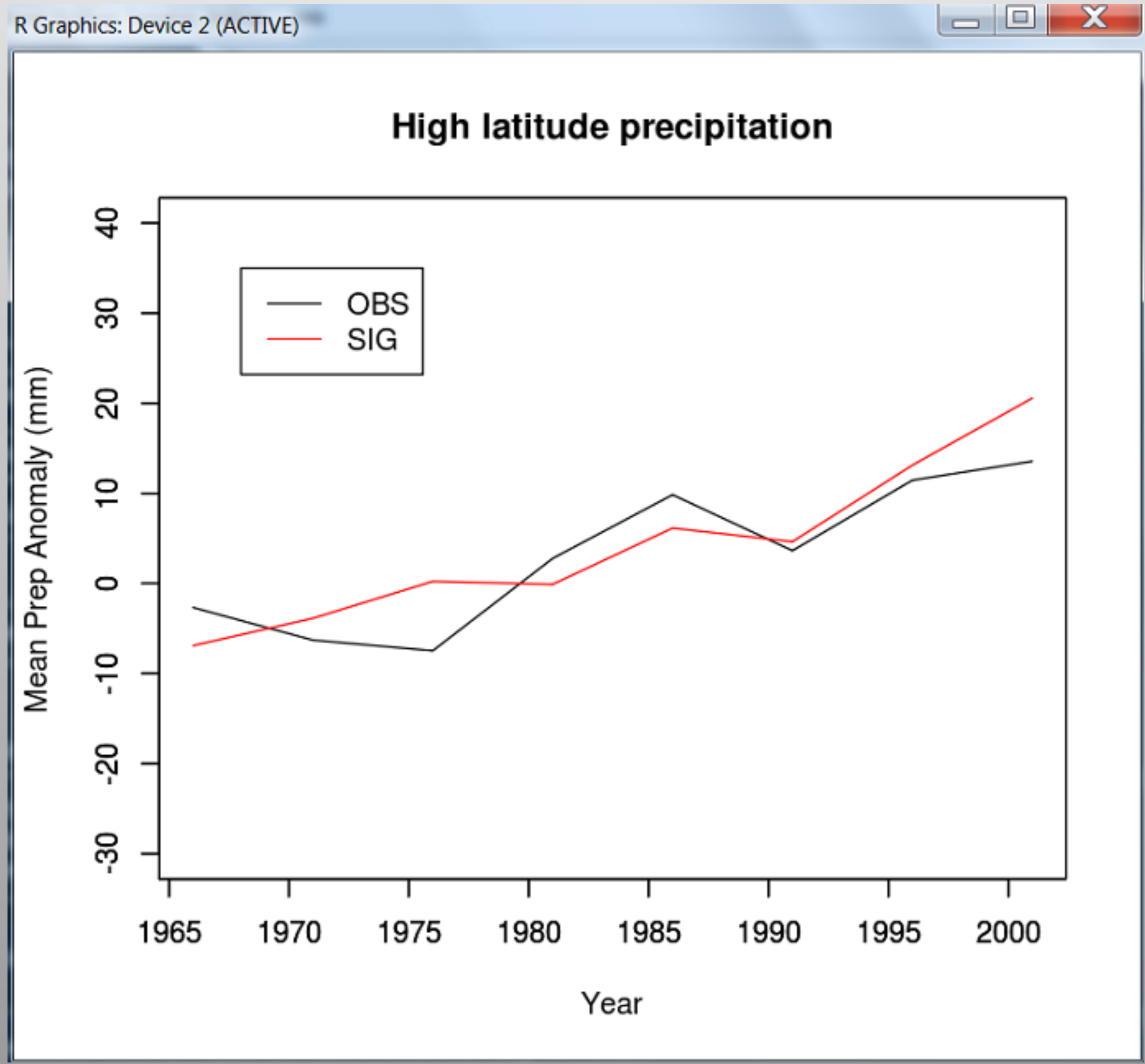
Use “readin” to get the data into R

R Console

```
> source('ECOF_V1.r')
> z=readin('ALL_ann_1area_obs_sig.dat','noise1_05yr_ann_piC_1area.dat', 'noise2_05yr_ann_piC_1area.dat')
>
```

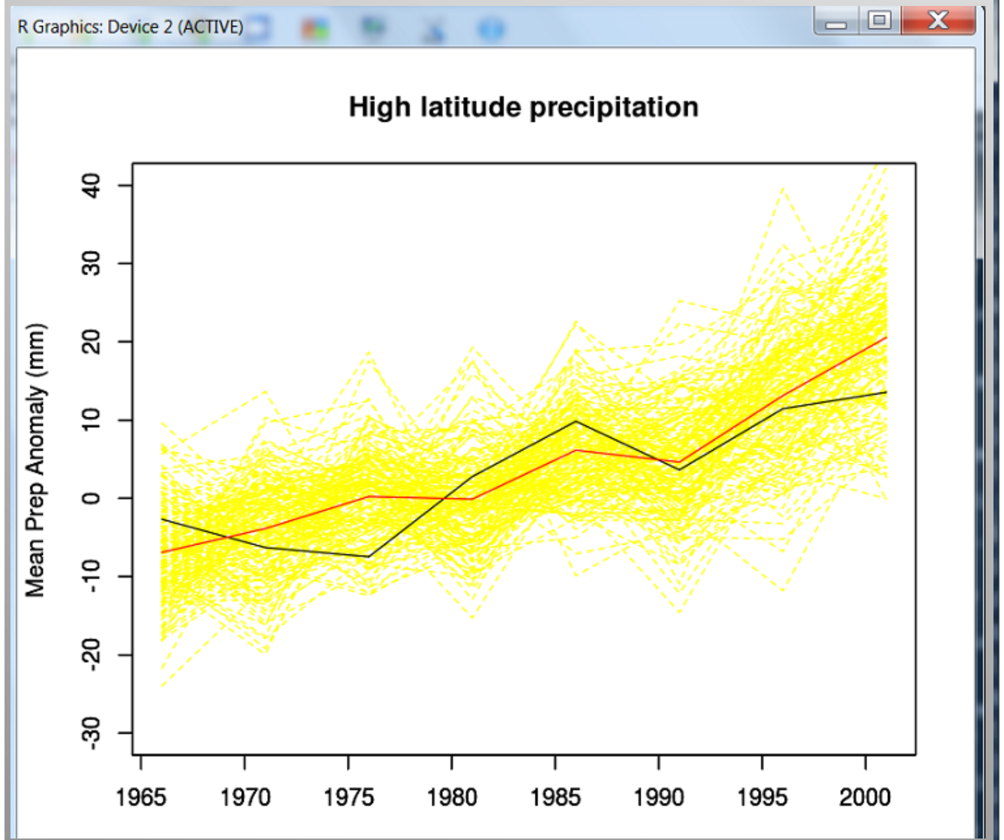
- Z@X (signal)
- Z@Y (observation)
- Z@noise1
- Z@noise2
- Have a look at these variables
- Plot observation and signal versus time

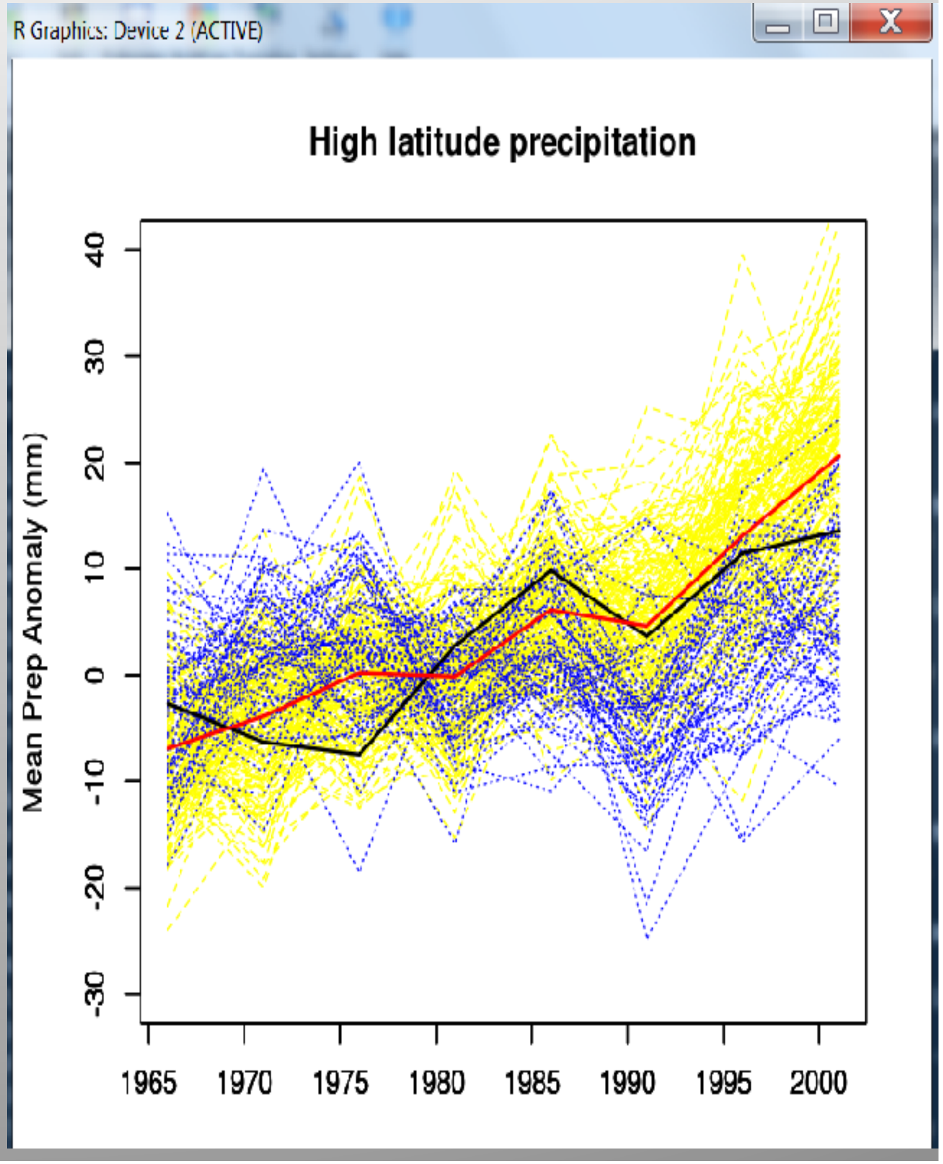
obs and **sig** look like this
… can you reproduce this plot?



Can you plot the individual ALL simulations? What about the NAT simulations?

- Data in ‘ALL_ann_1area.dat’
- Read in as a matrix
- NAT data in ‘NAT_ann_1area.dat’





Do the detection analysis…

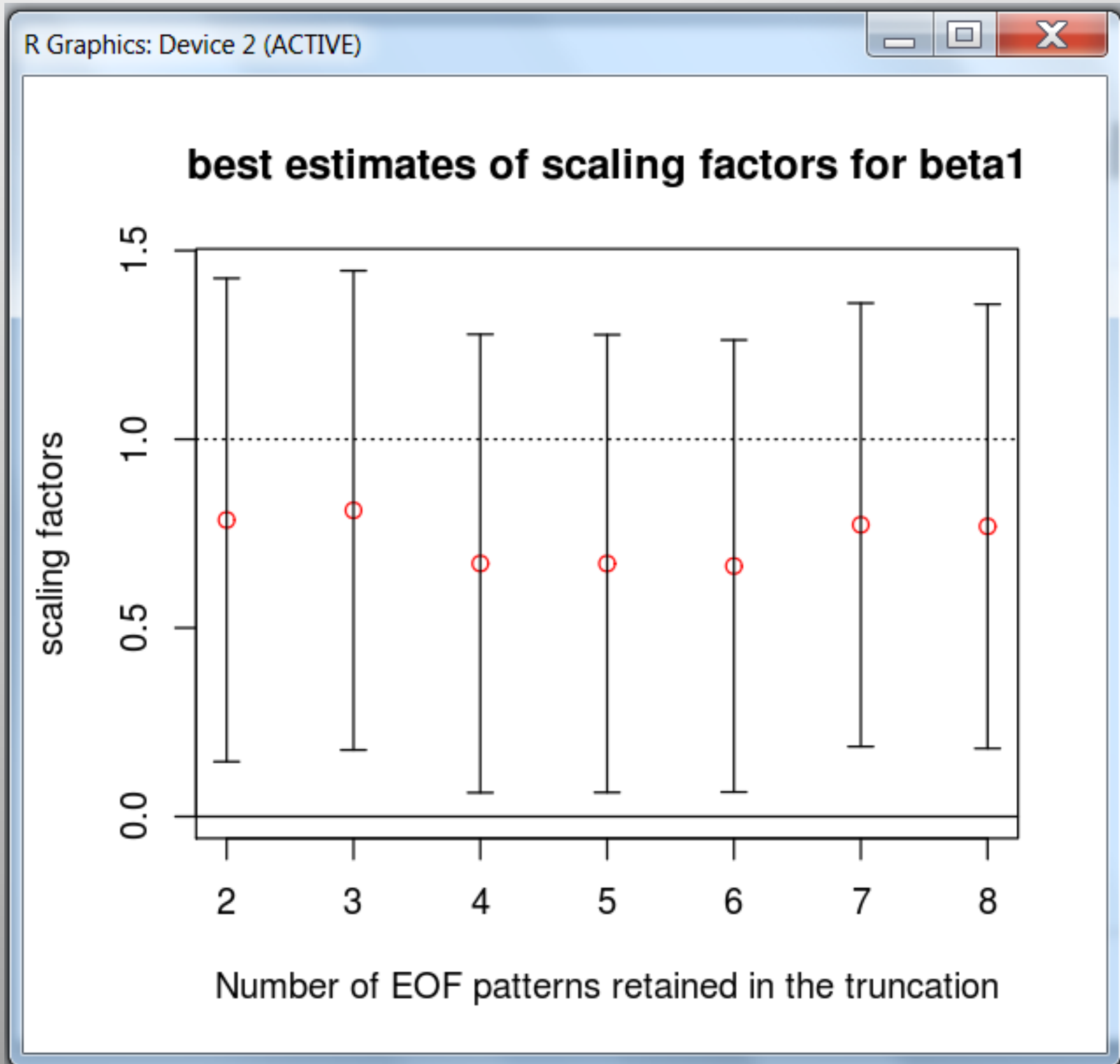
- o1.ols<-ols(Z@Y,Z@X,Z@noise1,Z@noise2,nsig=158)

R Console

```
> o1.tls<-tls.A03(Z@Y,Z@X,Z@noise1,Z@noise2,nsig=158)
> par(mfrow=c(1,2))
> plotbetas(o1.tls)
> plotrstat(o1.tls)
> o1.ols
     #EOF  beta1_low beta1_hat beta1_up    RCstat       RClow      RCup
[1,]    2 0.14546289 0.7859558 1.426449 0.3963152 0.003951788  3.885404
[2,]    3 0.17639898 0.8114696 1.446540 0.7246988 0.103288158  6.092039
[3,]    4 0.06322550 0.6706222 1.278019 1.4629156 0.354992626  7.983712
[4,]    5 0.06378339 0.6703762 1.276969 1.4633652 0.718712369  9.736728
[5,]    6 0.06487550 0.6639281 1.262981 1.5114460 1.161147031 11.410882
[6,]    7 0.18552136 0.7731514 1.360781 2.3540613 1.661920421 13.034545
[7,]    8 0.18026794 0.7690119 1.357756 3.0044387 2.208225327 14.623728
> 
```

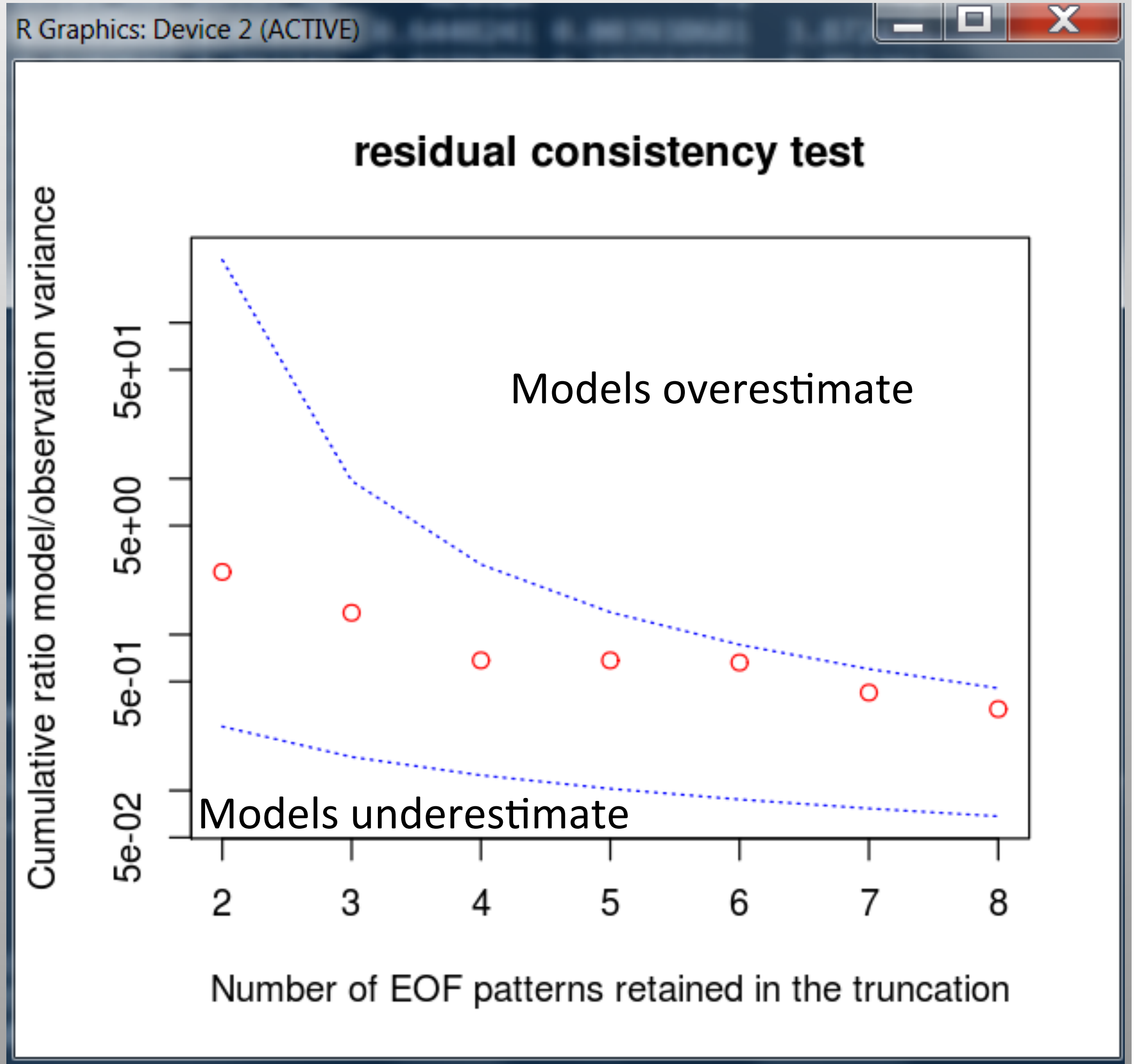
Display and analysis of scaling factors

- plotbetas(o1.ols)
- Is result robust w.r.t. number of EOF truncations?
- What is the indication of SF<1?



Evaluating the fitted model

- Residual consistency test
  - Tests the hypothesis that model simulated internal variability is equal to observed
  - F-test based on ratio of model internal variance to observed residual variance
  - Dashed curves show the critical values for rejecting null hypothesis at the 10% level
  - plotrstat(o1.ols)



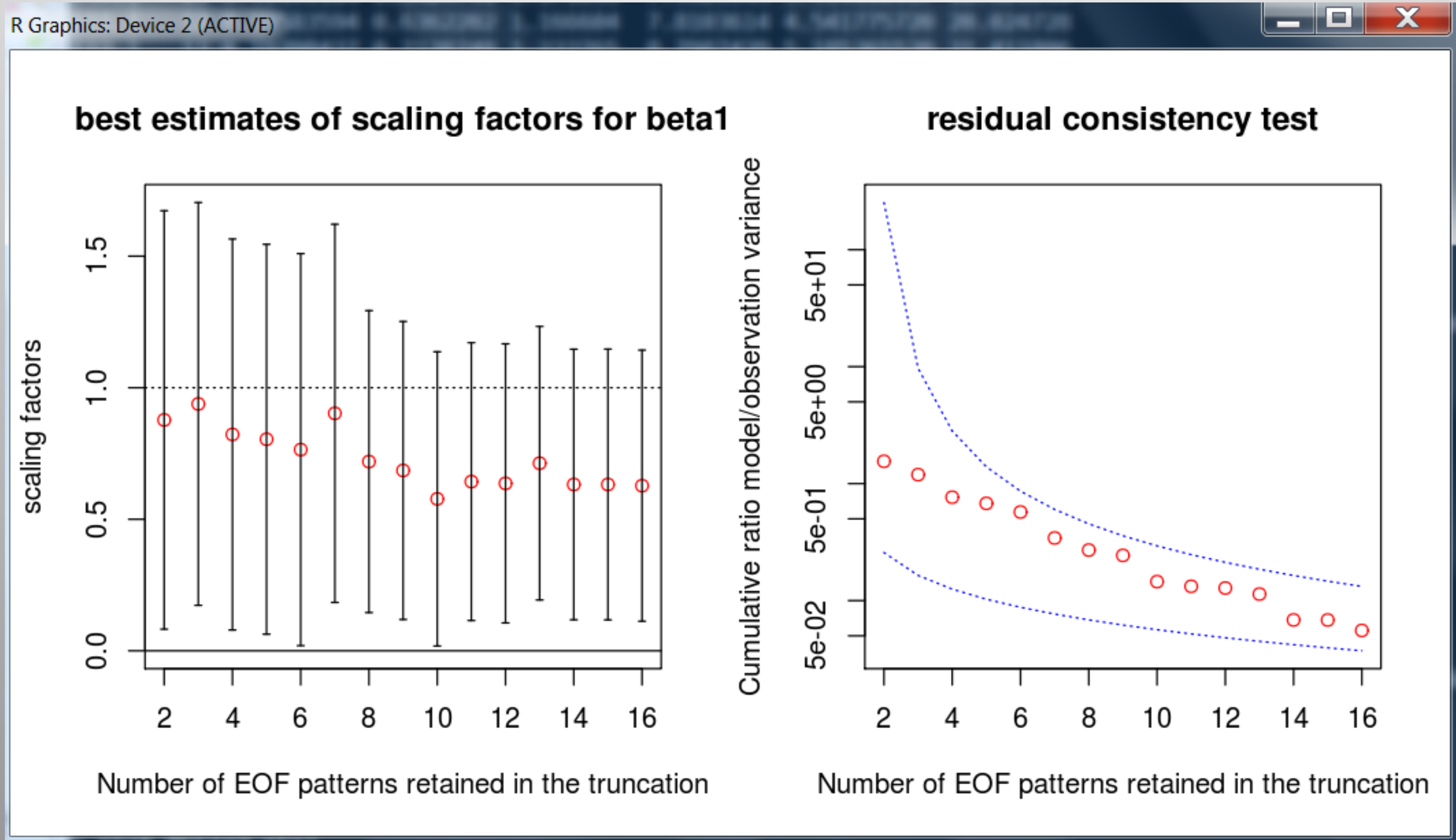
Can you redo the analysis at different spatial scales?

- 2 area
- 3 area
- 6 area
- Do we get similar results and why?

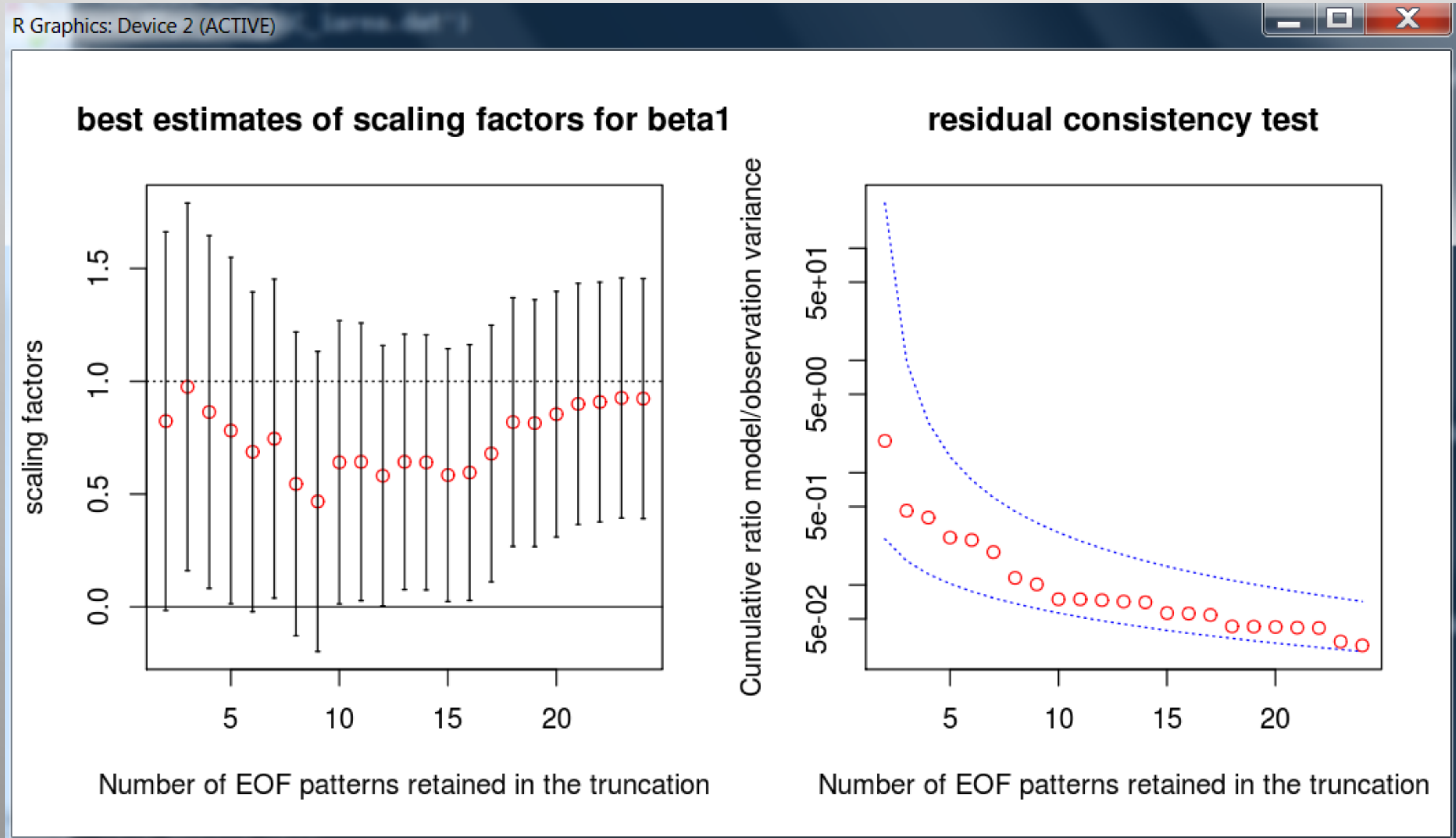
R Console

```
     #EOF  beta1_low beta1_hat beta1_up    RCstat          f1        f2
[1,]    2 0.14546289 0.7859558 1.426449 0.3963152 0.003938681  3.872435
[2,]    3 0.17639898 0.8114696 1.446540 0.7246988 0.102604015  6.051293
[3,]    4 0.06322550 0.6706222 1.278019 1.4629156 0.351469973  7.903506
[4,]    5 0.06378339 0.6703762 1.276969 1.4633652 0.709213168  9.606164
[5,]    6 0.06487550 0.6639281 1.262981 1.5114460 1.141981206 11.219435
[6,]    7 0.18552136 0.7731514 1.360781 2.3540613 1.629029899 12.771900
[7,]    8 0.18026794 0.7690119 1.357756 3.0044387 2.157278064 14.279690
> Z2<-readin('ALL_ann_2area_obs_sig.dat','noise1_05yr_ann_piC_2area.dat','noi$
> o2.ols<-ols(Z2@Y,Z2@X,Z2@noise1,Z2@noise2,nsig=158)
> o2.ols
      #EOF  beta1_low beta1_hat beta1_up     RCstat          f1        f2
 [1,]    2 0.08209706 0.8773951 1.672693  0.6440241 0.003938681  3.872435
 [2,]    3 0.17277660 0.9382592 1.703742  0.8378780 0.102604015  6.051293
 [3,]    4 0.07906522 0.8221289 1.565193  1.3066532 0.351469973  7.903506
 [4,]    5 0.06304409 0.8040054 1.544967  1.4746139 0.709213168  9.606164
 [5,]    6 0.01948073 0.7645583 1.509636  1.7490704 1.141981206 11.219435
 [6,]    7 0.18377884 0.9025019 1.621225  2.9161703 1.629029899 12.771900
 [7,]    8 0.14520507 0.7190926 1.292980  3.6984943 2.157278064 14.279690
 [8,]    9 0.11895781 0.6853389 1.251720  4.1076791 2.718009179 15.752937
 [9,]   10 0.01808104 0.5773076 1.136534  6.8823211 3.305120339 17.198485
[10,]   11 0.11491104 0.6429731 1.171035  7.5523543 3.914160917 18.621210
[11,]   12 0.10583594 0.6362202 1.166604  7.8103614 4.541775720 20.024728
[12,]   13 0.19298437 0.7128748 1.232765  8.7987429 5.185365520 21.411806
[13,]   14 0.11787659 0.6320074 1.146138 14.6323297 5.842870634 22.784618
[14,]   15 0.11747742 0.6321292 1.146781 14.6464140 6.512627551 24.144906
[15,]   16 0.11216501 0.6274524 1.142740 18.0032273 7.193270776 25.494094
>
```

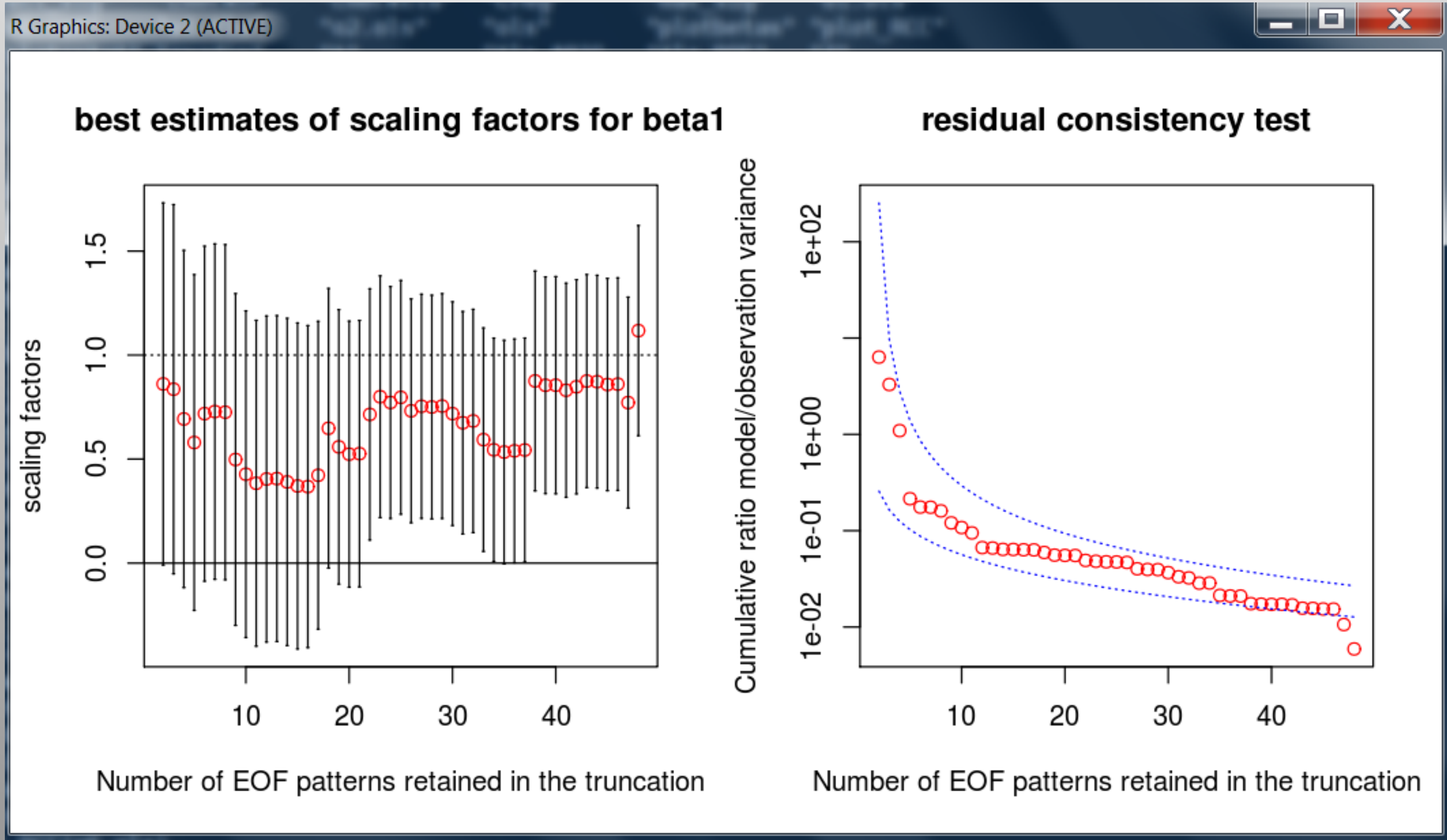
Results at 2-area spatial resolution:



Results at 3-area spatial resolution:



Results at 6-area spatial resolution …



Discussions

- Will results be different when adopting different spatiotemporal filtering?
- Why?
- What insights can we gain from varying the spatiotemporal resolutions?

In-depth exercises

- Try TLS method to redo the analysis
- Try ROF method to redo the analysis

Discussions

- What are the assumptions behind each method?
- Do you get consistent results by using different methods? Why?
- Any other questions?

Thank you!

Photo by F. Zwiers

Results at 3-area spatial resolution:

R Console

```
> Z3<-readin('ALL_ann_3area_obs_sig.dat','noise1_05yr_ann_piC_3area.dat','noi$
> o3.ols<-ols(Z3@Y,Z3@X,Z3@noise1,Z3@noise2,nsig=158)
> o3.ols
      #EOF    beta1_low beta1_hat beta1_up     RCstat          f1        f2
 [1,]    2 -0.014978838 0.8237876 1.662554  0.5190184  0.003938681  3.872435
 [2,]    3  0.161203193 0.9756151 1.790027  2.1776578  0.102604015  6.051293
 [3,]    4  0.082367956 0.8640346 1.645701  2.5081731  0.351469973  7.903506
 [4,]    5  0.014475406 0.7817215 1.548968  3.7781144  0.709213168  9.606164
 [5,]    6 -0.020851915 0.6875829 1.396018  3.9742626  1.141981206 11.219435
 [6,]    7  0.039006618 0.7457442 1.452482  5.0888257  1.629029899 12.771900
 [7,]    8 -0.128028337 0.5453597 1.218748  8.6409429  2.157278064 14.279690
 [8,]    9 -0.197021296 0.4675433 1.132108  9.8565770  2.718009179 15.752937
 [9,]   10  0.013752434 0.6410993 1.268446 13.3971972  3.305120339 17.198485
[10,]   11  0.028281195 0.6429766 1.257672 13.3981855  3.914160917 18.621210
[11,]   12  0.004501736 0.5815024 1.158503 13.6512650  4.541775720 20.024728
[12,]   13  0.077158411 0.6430265 1.208895 14.0422875  5.185365520 21.411806
[13,]   14  0.075521128 0.6408657 1.206210 14.2426336  5.842870634 22.784618
[14,]   15  0.024526817 0.5845863 1.144646 17.7702595  6.512627551 24.144906
[15,]   16  0.028985843 0.5959435 1.162901 17.9161830  7.193270776 25.494094
[16,]   17  0.111299319 0.6799717 1.248644 18.4897199  7.883663693 26.833360
[17,]   18  0.268363034 0.8192868 1.370211 23.3456423  8.582848670 28.163694
[18,]   19  0.267624492 0.8149095 1.362194 23.4387515  9.290010253 29.485934
[19,]   20  0.310587041 0.8544912 1.398395 23.6035989 10.004447516 30.800803
[20,]   21  0.365009437 0.8995950 1.434181 23.9871007 10.725552927 32.108921
[21,]   22  0.377164593 0.9083757 1.439587 24.1337884 11.452795957 33.410833
[22,]   23  0.394680850 0.9262991 1.457917 31.9031331 12.185710195 34.707013
[23,]   24  0.391694058 0.9233058 1.454918 34.5225549 12.923883100 35.997883
>
```

Reference results

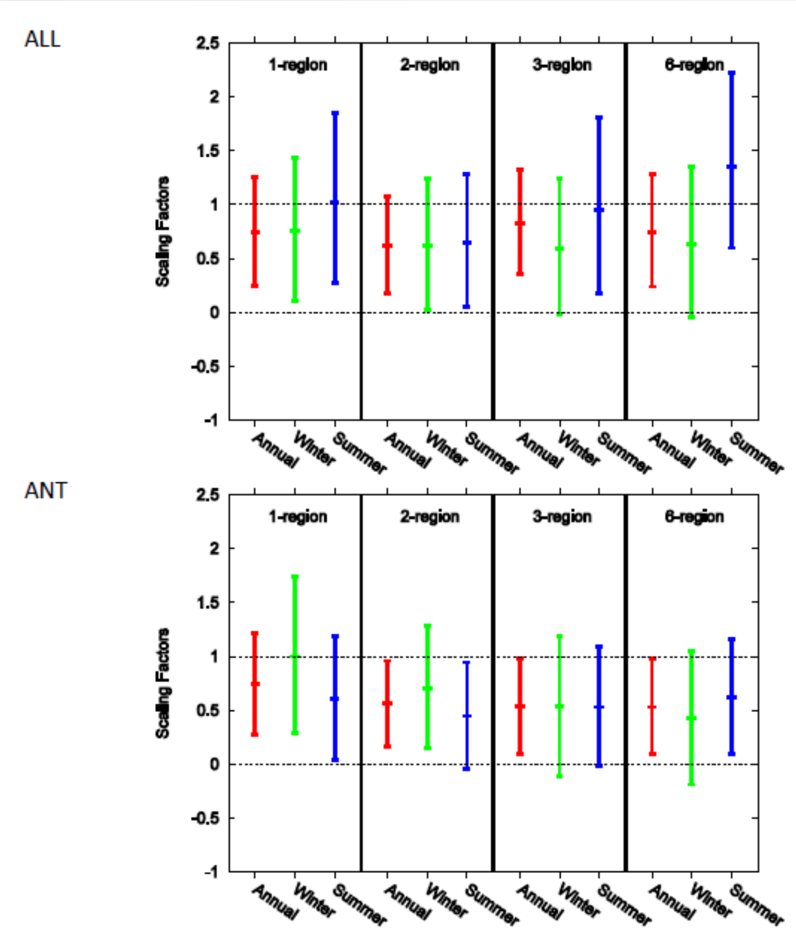


Figure 5, Zhang et al., 2014

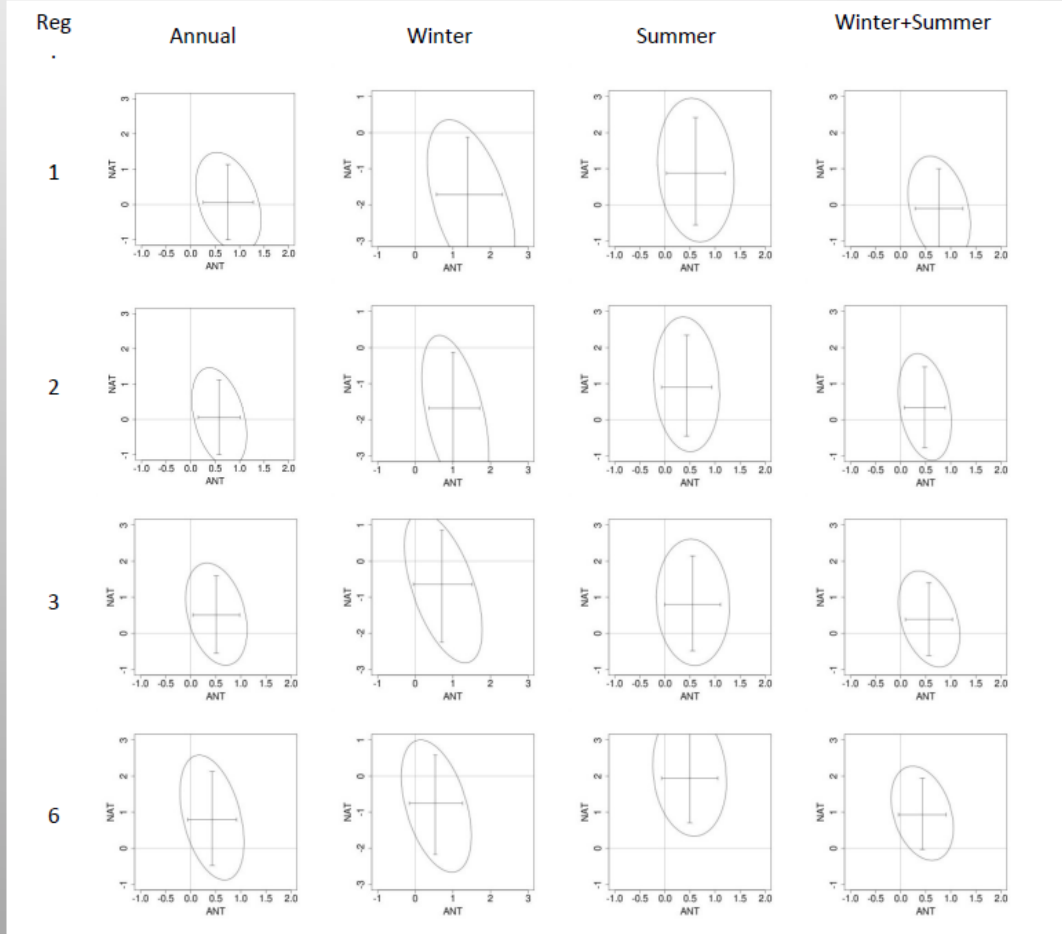


Figure 6, Zhang et al., 2014