NMF comparisons

Jerry

2025-09-11

Data Cleaning

- these data are all of the merged trailer data VOCs at 10 minutes within an hour, matched with the other compounds at that 10 minutes.

Limits of detection

- Limits of detection are defined by the instrument and can be found in Table 1a of the final report
- Note we are using 20 ppb for CO rather than 30 ppb per guidance from Gunnar

Background concentration correction

- we take the minimum concentration for each compound as the background value
- adjustments made according to paper: Gunnar's 2018 paper section 2.2 and Guha 2015 section 3.3
- Check whether chemical has background noise level that needs to be removed
- NO ADJUSTMENT if minimum value < 2xLOD and maximum value > 100xLOD

Normalize the data

Remove Ozone

Compute NMF

Compute uncertainty/weights matrix

Inverse uncertainty matrix

In the NMF library, weight is defined as 1/s_ij, at least from what their code look like. From the code, WRSS is computed using: sum(((X - fitted(object)) * weight)^2 , na.rm = TRUE)/2 This only matches the regular WRSS form if weight is 1/s_ij

Custom function to run NMF for different specs

Define global variables

NMF + random seed

NMF + nndsvd seed

LS-NMF + random seed

LS-NMF + nndsvd seed

Compare NMF

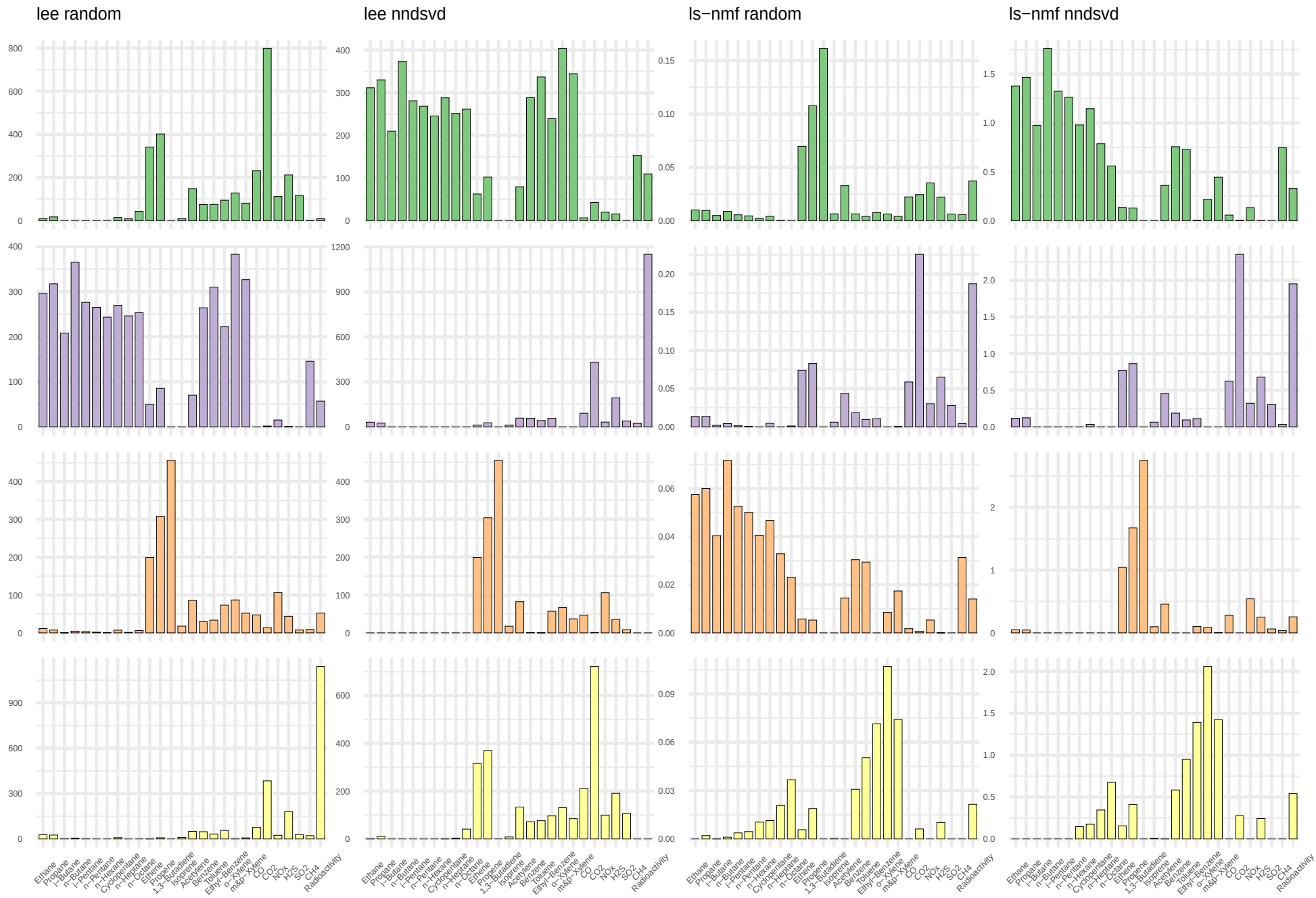
Variables for renaming

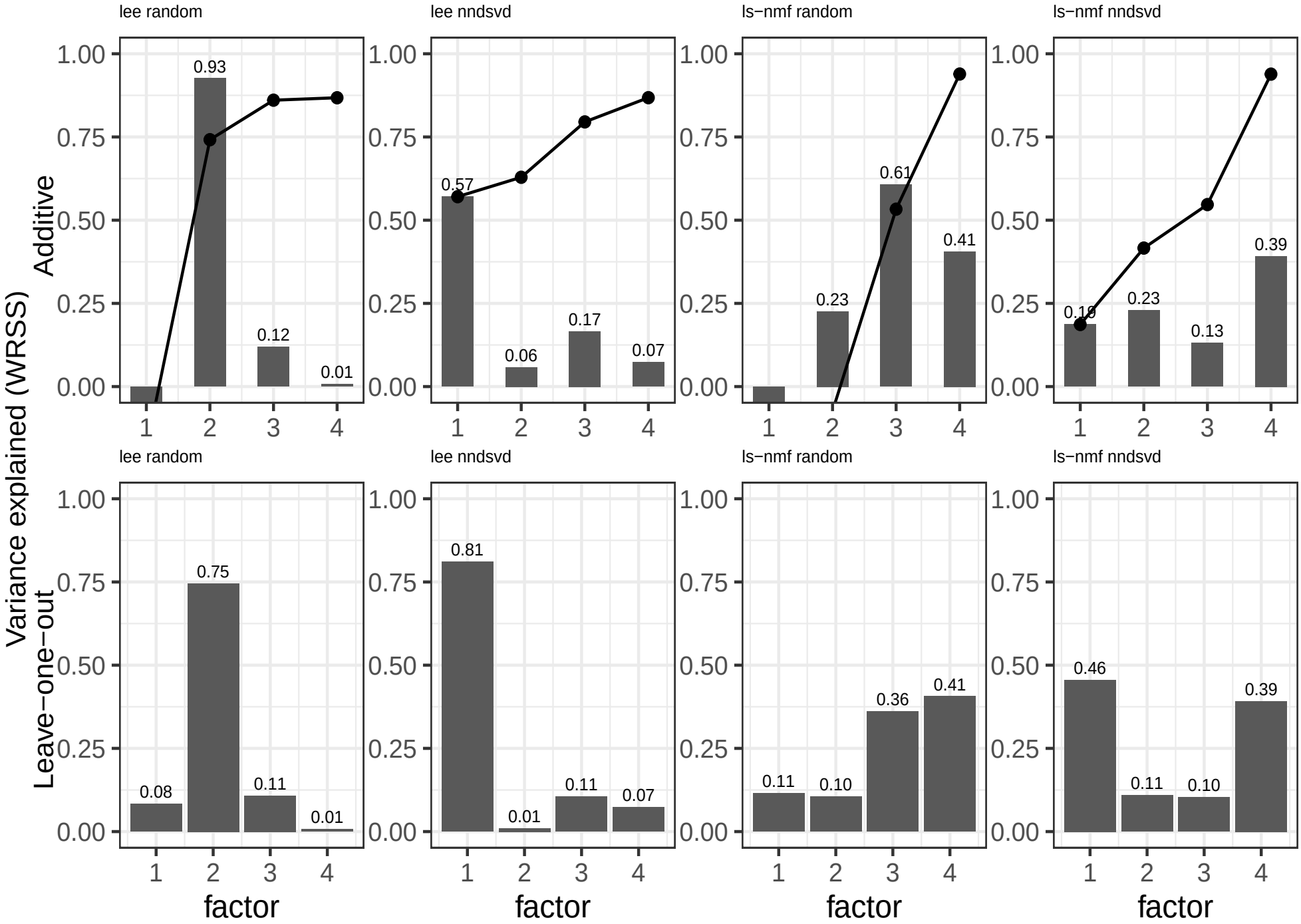
Source Contribution Plot

Variance

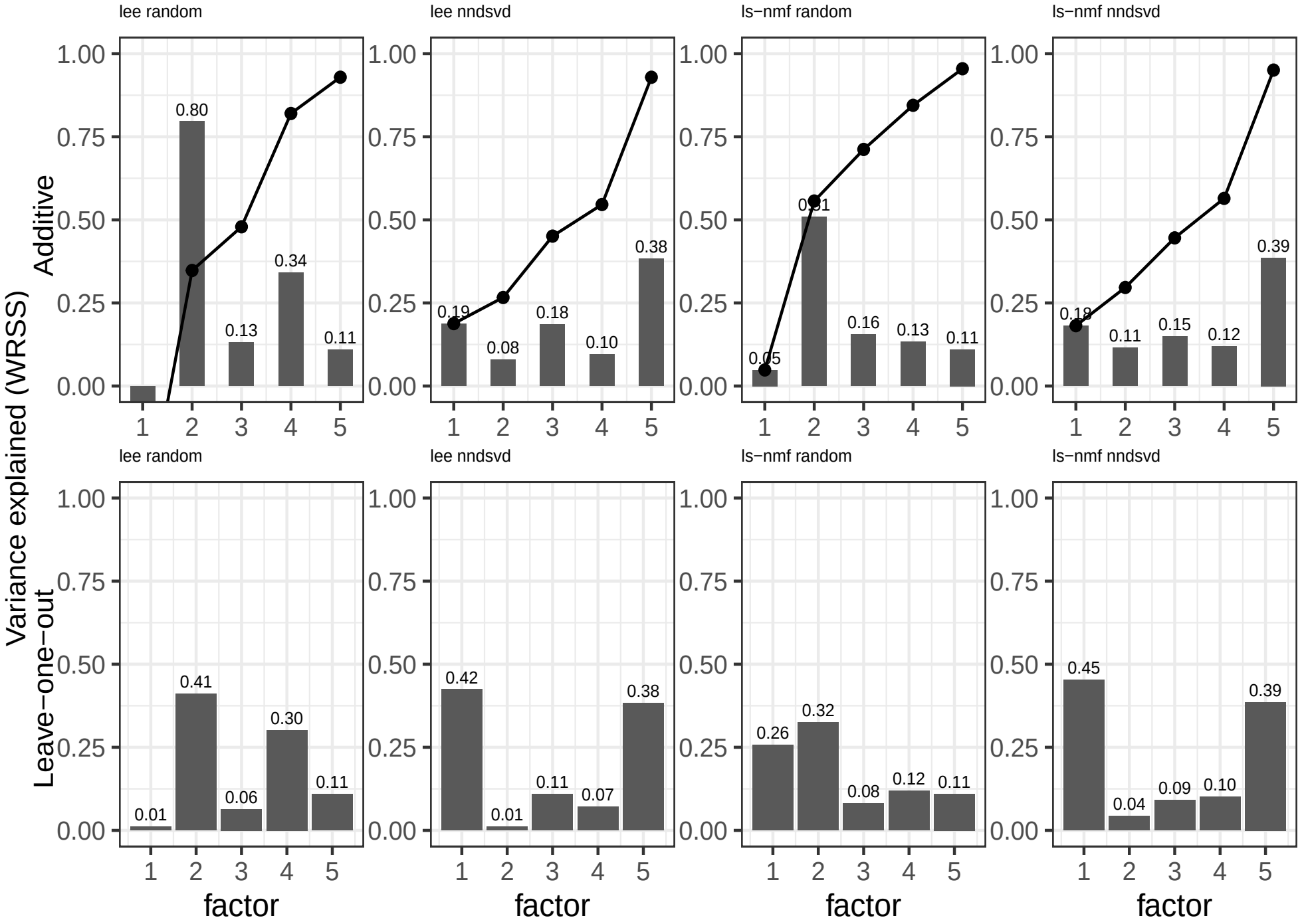
Plot factor by factor

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

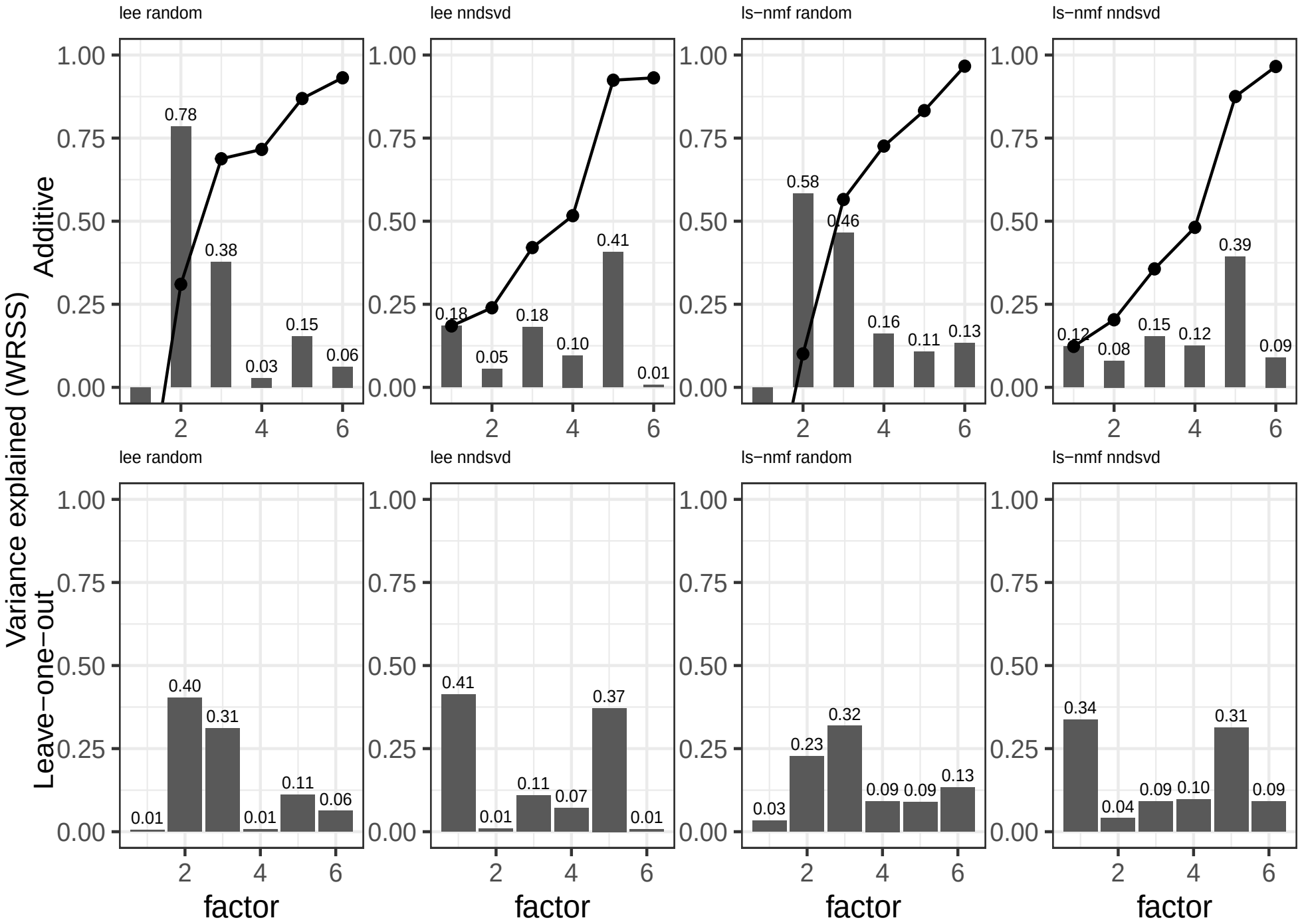

lee random
lee nndsvd
ls−nmf random
ls−nmf nndsvd
Ethane
Propane
i−Butane
n−Butane
i−Pentane
n−Pentane
n−Hexane
Cyclopentane
n−Heptane
n−Octane
Ethene
Propene
1,3−Butadiene
Isoprene
Acetylene
Benzene
Toluene
Ethyl−Benzene
o−Xylene
m&p−Xylene
CO
CO2
NOx
H2S
SO2
CH4
Radioactivity


lee random
lee nndsvd
ls−nmf random
ls−nmf nndsvd
Additive
Leave−one−out
Variance explained (WRSS)
factor
1.00
0.75
0.50
0.25
0.00
1
2
3
4
0.93
0.12
0.01
0.57
0.06
0.17
0.07
0.23
0.61
0.41
0.10
0.23
0.13
0.39
0.08
0.75
0.11
0.01
0.81
0.01
0.11
0.07
0.11
0.10
0.36
0.41
0.46
0.11
0.10
0.39

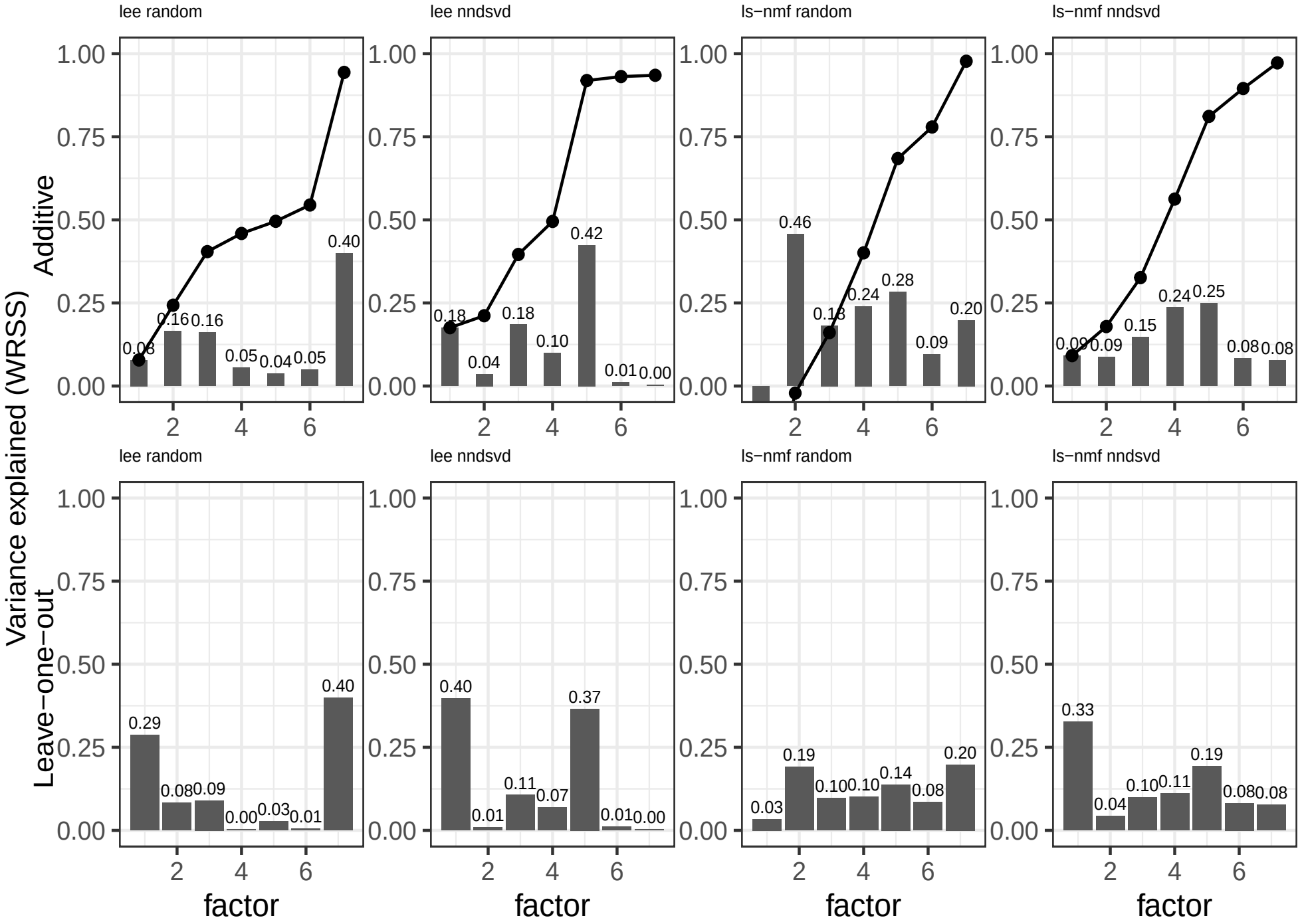

lee random
lee nndsvd
ls−nmf random
ls−nmf nndsvd
Ethane
Propane
n−Butane
i−Butane
n−Pentane
i−Pentane
n−Hexane
Cyclopentane
n−Heptane
n−Octane
Ethene
Propene
1,3−Butadiene
Isoprene
Acetylene
Benzene
Toluene
Ethyl−Benzene
m&p−Xylene
o−Xylene
CO
CO2
NOx
H2S
SO2
CH4
Radioactivity

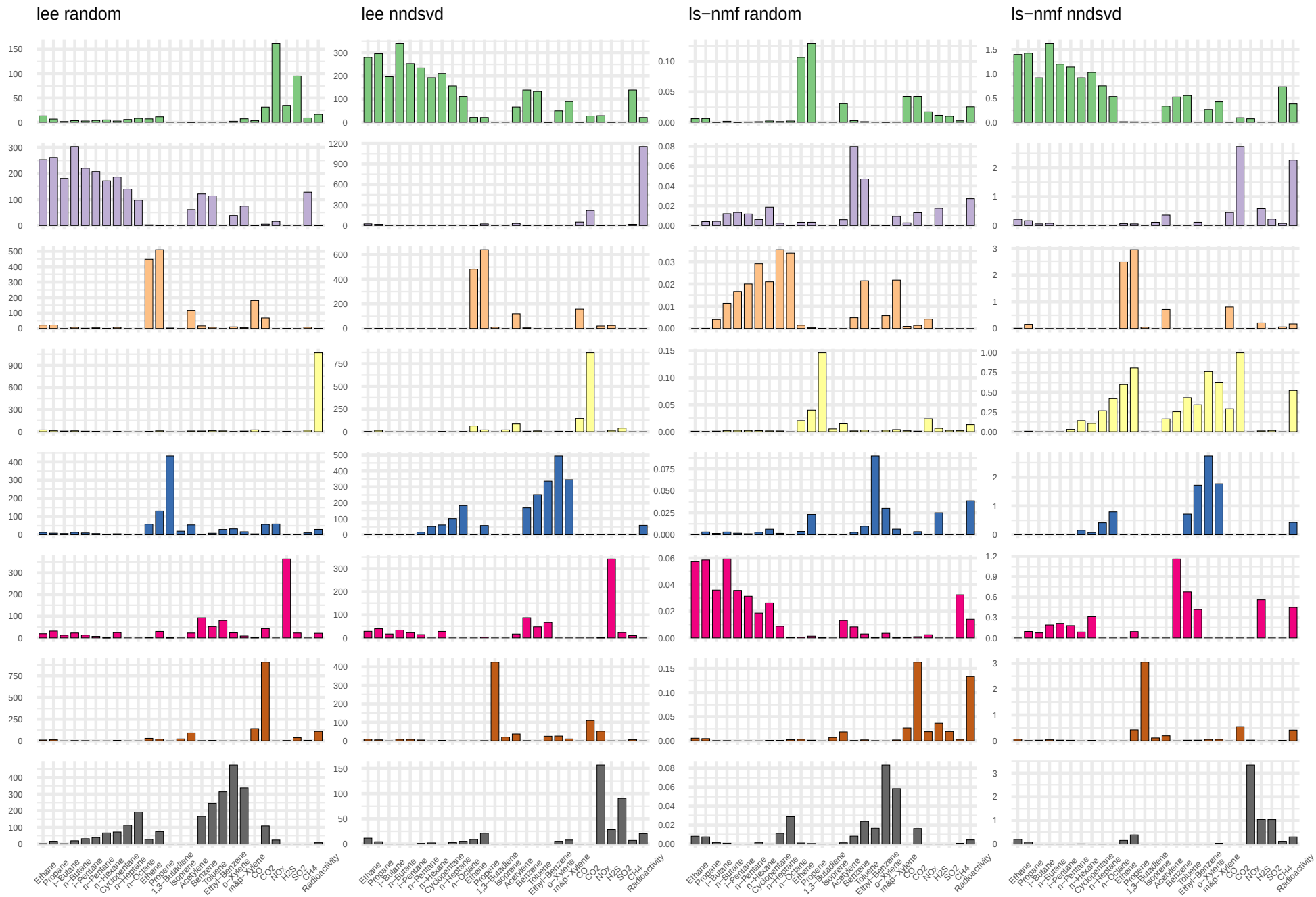

lee random
lee nndsvd
ls−nmf random
ls−nmf nndsvd
Additive
Leave−one−out
Variance explained (WRSS)
factor
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
0.80
0.13
0.34
0.11
0.19
0.08
0.18
0.10
0.38
0.05
0.51
0.16
0.13
0.11
0.18
0.11
0.15
0.12
0.39
0.01
0.41
0.06
0.30
0.11
0.42
0.01
0.11
0.07
0.38
0.26
0.32
0.08
0.12
0.11
0.45
0.04
0.09
0.10
0.39

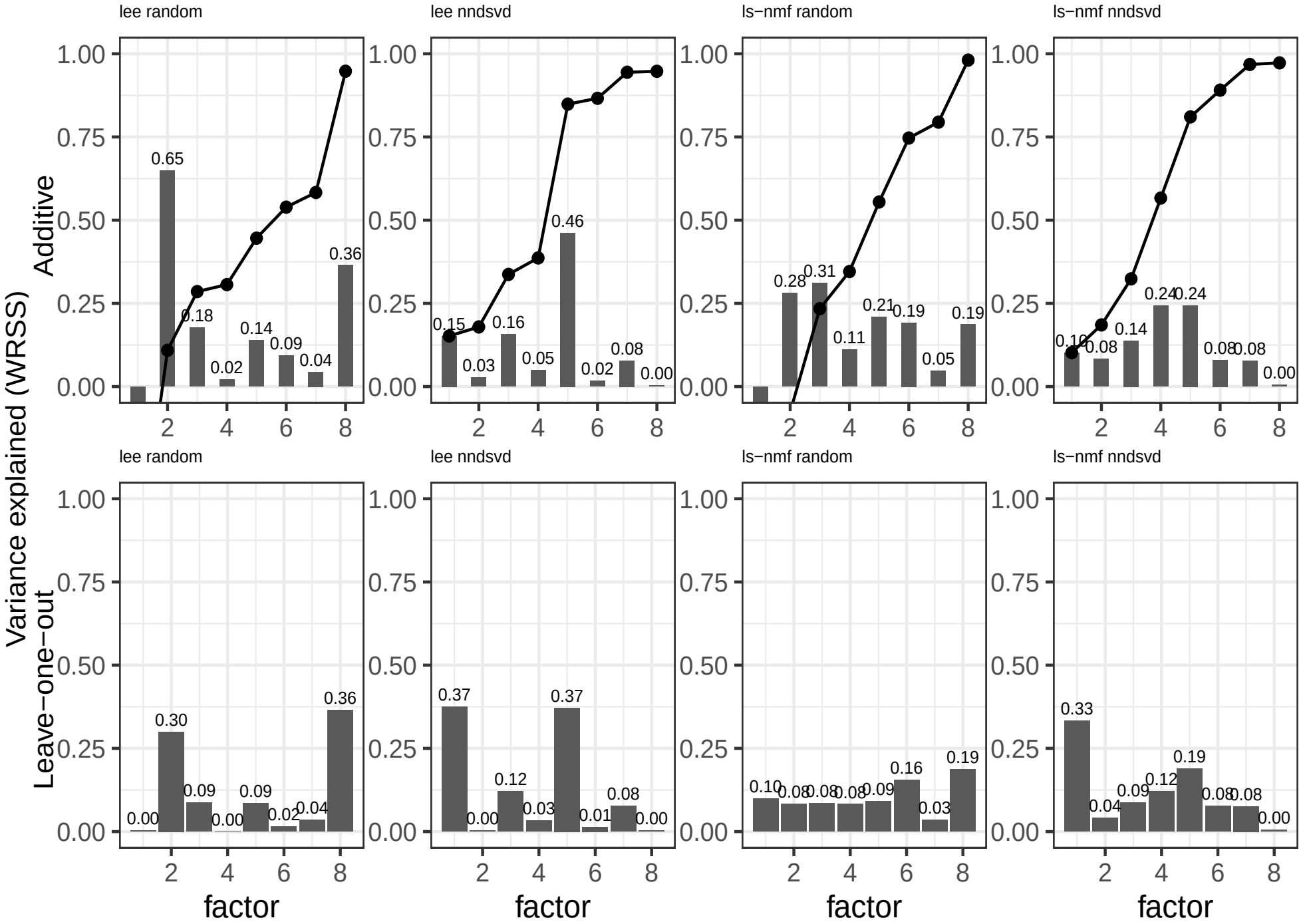

lee random
lee nndsvd
ls−nmf random
ls−nmf nndsvd
Ethane
Propane
i−Butane
n−Butane
i−Pentane
n−Pentane
n−Hexane
Cyclopentane
n−Heptane
n−Octane
Ethene
Propene
1,3−Butadiene
Isoprene
Acetylene
Benzene
Toluene
Ethyl−Benzene
o−Xylene
m&p−Xylene
CO
CO2
NOx
H2S
SO2
CH4
Radioactivity


lee random
lee nndsvd
ls−nmf random
ls−nmf nndsvd
Additive
Leave−one−out
Variance explained (WRSS)
1.00
0.75
0.50
0.25
0.00
2
4
6
factor
0.78
0.38
0.03
0.15
0.06
0.18
0.05
0.18
0.10
0.41
0.01
0.58
0.46
0.16
0.11
0.13
0.12
0.08
0.15
0.12
0.39
0.09
0.01
0.40
0.31
0.01
0.11
0.06
0.41
0.01
0.11
0.07
0.37
0.01
0.03
0.23
0.32
0.09
0.09
0.13
0.34
0.04
0.09
0.10
0.31
0.09


lee random
lee nndsvd
ls−nmf random
ls−nmf nndsvd
Ethane
Propane
i−Butane
n−Butane
i−Pentane
n−Pentane
n−Hexane
Cyclopentane
n−Heptane
n−Octane
Ethene
Propene
1,3−Butadiene
Isoprene
Acetylene
Benzene
Toluene
Ethyl−Benzene
m&p−Xylene
o−Xylene
CO
CO2
NOx
H2S
SO2
CH4
Radioactivity


lee random
lee nndsvd
ls−nmf random
ls−nmf nndsvd
Additive
Leave−one−out
Variance explained (WRSS)
1.00
0.75
0.50
0.25
0.00
2
4
6
0.08 0.16 0.16 0.05 0.04 0.05 0.40
0.18 0.04 0.18 0.10 0.42 0.01 0.00
0.46 0.13 0.24 0.28 0.09 0.20
0.09 0.09 0.15 0.24 0.25 0.08 0.08
0.29 0.08 0.09 0.00 0.03 0.01 0.40
0.40 0.01 0.11 0.07 0.37 0.01 0.00
0.03 0.19 0.10 0.10 0.14 0.08 0.20
0.33 0.04 0.10 0.11 0.19 0.08 0.08
factor


lee random
lee nndsvd
ls−nmf random
ls−nmf nndsvd
Ethane
Propane
i−Butane
n−Butane
i−Pentane
n−Pentane
n−Hexane
Cyclopentane
n−Heptane
n−Octane
Ethene
Propene
1,3−Butadiene
Isoprene
Acetylene
Benzene
Toluene
Ethyl−Benzene
m&p−Xylene
o−Xylene
CO
CO2
NOx
H2S
SO2
CH4
Radioactivity


lee random
lee nndsvd
ls−nmf random
ls−nmf nndsvd
Additive
Leave−one−out
Variance explained (WRSS)
factor
1.00
0.75
0.50
0.25
0.00
2
4
6
8
0.65
0.18
0.02
0.14
0.09
0.04
0.36
0.15
0.03
0.16
0.05
0.46
0.02
0.08
0.00
0.28
0.31
0.11
0.21
0.19
0.05
0.19
0.10
0.08
0.14
0.24
0.24
0.08
0.08
0.00
0.30
0.09
0.09
0.02
0.04
0.37
0.12
0.03
0.01
0.10
0.16
0.33