```
cd 'C:\Users\calderom\OneDrive - Dundalk Institute of Technology\PARAFACcourse\MATLAB\drEEM-0.6
dreeminstall

savelocation = ['C:\Users\calderom\OneDrive - Dundalk Institute of Technology\PARAFACcourse\Wee
```

Load EEMs

```
cd('C:\Users\calderom\OneDrive - Dundalk Institute of Technology\PARAFACcourse\Week3\AQUACOSMda
filetype=1;ext = 'csv';RangeIn='A1..AV182';headers=[1 1];display_opt=0;outdat=2;
[X,Emmat,Exmat,filelist_eem,outdata]=readineems(filetype,ext,RangeIn,headers,display_opt,outdat
Ex=Exmat(1,:); %Since all files have the same excitation wavelengths
Em=Emmat(:,1); %Since all files have the same emission wavelengths
```

Load Blanks

```
cd('C:\Users\calderom\OneDrive - Dundalk Institute of Technology\PARAFACcourse\Week3\AQUACOSMda
filetype=1;ext = 'csv';RangeIn='A1..AV182';headers=[1 1];display_opt=0;outdat=2;
[X_b,Emmat_b,Exmat_b,filelist_b,outdata_b]=readineems(filetype,ext,RangeIn,headers,display_opt,

Xin_B=assembledataset(X_b,Exmat_b(1,:),Emmat_b(:,1),'AU','filelist',filelist_b,[]); %create a c

Exb=Exmat_b(1,:); % Extract one column of emission wavelengths
Emb=Emmat_b(:,1); % Extract one row of excitation wavelengths
```

Load Absorbances

```
cd('C:\Users\calderom\OneDrive - Dundalk Institute of Technology\PARAFACcourse\Week3\AQUACOSMda
filetype='Abs';ext = 'csv';RangeIn='A1..B701';display_opt=0;outdat=0;
[S_abs,W_abs,wave_abs,filelist_abs]=readinscans(filetype,ext,RangeIn,display_opt,outdat);
A = [wave_abs; S_abs];
```

Metadata alignment

```
filename='C:\Users\calderom\OneDrive - Dundalk Institute of Technology\PARAFACcourse\Week3\AQUA
strings=[1 1 1 1 0 1 1 0 1 0 1 1 1]; %specify which columns in the log contain text
SampleLog=readlogfile(filename,strings);

% Align data according to the SampleLog. The EEMs are listed in SampleLog_EEMfile
% first align raw data from SampleLog with filelist_eem
AnalDate=   alignds(SampleLog,{'EEMfile',filelist_eem},{'DateSampleAnalysed'});
DayExp=     alignds(SampleLog,{'EEMfile',filelist_eem},{'DOE'});
Treat=      alignds(SampleLog,{'EEMfile',filelist_eem},{'Treatment'});
sampleID=   alignds(SampleLog,{'EEMfile',filelist_eem},{'SampleID'});
Rep=        alignds(SampleLog,{'EEMfile',filelist_eem},{'Replicate'});
Dilution=   alignds(SampleLog,{'EEMfile',filelist_eem},{'DilutionFactor'});

% next align previously loaded datasets with filelist_eem (ABS, Raman, Blanks, etc).
Sabs=     alignds(SampleLog,{'EEMfile',filelist_eem},{'ABSfile',filelist_abs,S_abs});
B=        alignds(SampleLog,{'EEMfile',filelist_eem},{'Blankfile',filelist_b,X_b});
% I don't have the following files!
%Sr= alignds(SampleLog,{'EEMfile',filelist_eem},{'RamanFile',filelist_R,S_R});
```

```
%Sqsuv275= alignds(SampleLog,{'EEMfile',filelist_eem},{'Qw',filelist_qsuv275,S_qsuv275});
%Sqsuv350= alignds(SampleLog,{'EEMfile',filelist_eem},{'Qw',filelist_qsuv350,S_qsuv350});

Xin=assembledataset(X,Ex,Em,'AU','filelist',filelist_eem,[])
checkdataset(Xin)
```

fdom correction

```
%Spectral correction files
cd ('C:\Users\calderom\OneDrive - Dundalk Institute of Technology\PARAFACcourse\Week3\AQUACOSMo
Excor=csvread('Excorr.csv');
Emcor=csvread('Emcorr.csv');

%Raman substraction from blanks
W350=[Emb';squeeze(X_b(:,:,Exb==350))]
[IR,IRmed,IRdiff] = ramanintegrationrange(W350,filelist_b,350,1800,4,0,0)
W = [Emb'; squeeze(B(:,:,Exb==350))];

%fdom correction
[XcRU2, Arp2, IFCmat2, BcRU2, XcQS2, QS_RU2]=fdomcorrect(Xin.X,Xin.Ex,Xin.Em,...
Emcor,Excor,W,[350 381 426],A,B,[],[],[]);
```

Assemble dataset to start PARAFAC analysis

```
Xin.X=XcRU2;
Xin.RamanArea=Arp2;
Xin.IFE=IFCmat2;

AQUAdata= assembledataset(XcRU2,Ex,Em,'RU','filelist',filelist_eem,...
    'DateSampleAnalysed',AnalDate,...
    'DOE',DayExp,...
    'Treatment', Treat,...
    'SampleID', sampleID,...
    'Replicate', Rep,...
    'DilutionFactor', Dilution,...
    [])

savelocation = ['C:\Users\calderom\OneDrive - Dundalk Institute of Technology\PARAFACcourse\Wee
save(savelocation,'AQUAdata')
```

Load the DS and start data inspection

```
clear
cd 'C:\Users\calderom\OneDrive - Dundalk Institute of Technology\PARAFACcourse\Week3'
load DS.mat
```

First inspection

```
eemview(AQUAdata,'X',[5 4])
```

Cut noise

```
Xstartinit=subdataset(AQUAdata,[],AQUAdata.Em<315,AQUAdata.Ex<250);
Xstart=subdataset(Xstartinit,[],[],Xstartinit.Ex>440);
eemview(Xstart,'X',[5 4])
eemreview(Xstart)
```

Remove scatter

```
% 1st Raman and 2nd Rayleigh
eemreview(smootheem(Xstart,[ ],[9 11],[16 14],[ ],[0 0 0 0],[],3382,'pause'))
% 1st Rayleigh
eemreview(smootheem(Xstart,[7 6],[9 11],[16 14],[ ],[0 0 0 0],[],3382,'pause'))
% 2nd Raman
eemreview(smootheem(Xstart,[7 6],[9 11],[16 14],[20 4],[0 0 0 0],[],3382,'pause'))
```

Spectralvariance --> potential first outliers (no interpolation)

```
spectralvariance(smootheem(Xstart,[7 6],[9 11],[16 14],[20 4],[0 0 0 0],[],3382,0))

eem = smootheem(Xstart,[7 6],[9 11],[16 14],[20 4],[0 0 0 0],[],3382,0)

eemreview(eem) %sample 45 = HF4d8 something weird!

eem = zap(eem,45,[], [355 365]) %remove weird EEM range!

eemreview(eem, 'sample', 45) %range of ex EEM removed correctly

scanview (eem)
```

First preliminary PARAFAC model

```
Test1=outliertest(eem,[2,2],3:7); %potentially 1 outlier HFL1d4 = 50 sample
compcorrplot(Test1,5) %component 3 and 2 highly correlated with C5
eemreview(eem,'sample',50) %nothing seems weird
eemview({Test1, Test1},{'X','Model5','error_residuals'},[1 3], 50) %nothing seems weird

Xpre=normeem(eem); %normalise data

Test1p=outliertest(Xpre,[2,2],3:7); %the same outlier
compcorrplot(Test1p,5,[],'Treatment') %by normalising even worse!

%move on with Test 1 (without normalisation)
spectralloadings(Test1,3:7) %definetly 5 components so far
loadingsandleverages(Test1,5) %ok
specsse(Test1, 3:6) %sample 50 outlier --> good but somehow influencing the model
errorsandleverages(Test1,5)
coreandvar(Test1)

%remove sample 50 before "real test"
out=ismember(eem.i,[50]);
eem=subdataset(eem,out,[],[]);
```

```
%try again without sample 50
Test2=outliertest(eem,[2,2],4:6); %no outliers now!
compcorrplot(Test2,5,[],'Treatment') %comp3vs1 slightly correlated but its ok
loadingsandleverages(Test2,5) %ok
spectralloadings(Test2,5)
fingerprint(Test2,5) %feels good!
```

PARAFAC MODEL

```
[LSmodel,all,details]=randinitanal(eem,4:6,50,'nonnegativity',1e-8);

spectralloadings(LSmodel,4:6)
comparespectra(LSmodel,4:6) %maybe model with 4 components better than others!
fingerprint(LSmodel,4) %good
fingerprint(LSmodel,5) %mmmm also good
fingerprint(LSmodel,6) %not good
loadingsandleverages(LSmodel,4) %ok
loadingsandleverages(LSmodel,5) %ok
compare2models(LSmodel,4,5,0.05) %difficult to say, both model residuals very similar (maybe mo
eemview({LSmodel, LSmodel},{'X','Model4','error_residuals'},[4 3]) %less noise...
eemview({LSmodel, LSmodel},{'X','Model5','error_residuals'},[4 3])

coreandvar(LSmodel) %better 4 in general...
```

INTERNAL VALIDATION (with model 4 and model 5)

```
%try different splitds
metadata(eem)
%1.alternating splitds
S1=splitds(eem,[],4,'alternating',{[1 2],[3 4],[1 3],[2 4],[1 4],[2 3]});
%2.bytreatment splitds
S2=splitds(eem,'Treatment',[],'exact', {[1 2], [3 4]});
%3.continguous blocks
S3=splitds(eem,[],4,'contiguous',{[1 2],[3 4],[1 3],[2 4],[1 4],[2 3]});
%4.random
S4=splitds(eem,[],4,'random',{[1 2],[3 4],[1 3],[2 4],[1 4],[2 3]});


%model with 4 components%.................................%
%WITH S1
A1=splitanalysis(S1,4,'starts',50,'constraints','nonnegativity','convgcrit',1e-8); %like randir
val4A1=splitvalidation(A1,4,[1 2;3 4;5 6],{'AB','CD','AC','BD','AD','BC'}, LSmodel); %overall r
%WITH S2
A2=splitanalysis(S2,4,'starts',50,'constraints','nonnegativity','convgcrit',1e-8); %like randir
val4A2=splitvalidation(A2,4,[1 2], {'AB','CD'}, LSmodel); %overall results = not validated
%WITH S3
A3=splitanalysis(S3,4,'starts',50,'constraints','nonnegativity','convgcrit',1e-8); %like randir
val4A3=splitvalidation(A3,4,[1 2;3 4;5 6],{'AB','CD','AC','BD','AD','BC'}, LSmodel); %overall r
%WITH S4
A4=splitanalysis(S4,4,'starts',50,'constraints','nonnegativity','convgcrit',1e-8); %like randir
val4A4=splitvalidation(A4,4,[1 2;3 4;5 6],{'AB','CD','AC','BD','AD','BC'}, LSmodel); %overall r

%model with 5 components%...........only with S1.....................%
```

```
A2=splitanalysis(S1,5,'starts',50,'constraints','nonnegativity','convgcrit',1e-8); %increase co
val5A2=splitvalidation(A2,5,[1 2;3 4;5 6],{'AB','CD','AC','BD','AD','BC'}, LSmodel); %overall r

val5all = splitvalidation(A2,5); close,close %validate again with DEFAULT option
r = relcomporder(val5all,1)

r(2,1)=1;
r(2,3)=3;
r(2,4)=4;
r(2,5)=5;
r(4,1)=1;
r(4,3)=4;
r(4,5)=5;
r(6,1)=1;
r(6,3)=3;
r(6,4)=4
spectralloadings(val5,5,1:6,r) %weird because some components seem ok but others most of the sp
```

EXPORT MODEL

```
%Get the peaks
picklist = pickpeaks(eem) %not working! pickpeaks requires the Curve Fitting Toolbox!

%Get the slopes --> modify mydata to add absorbance data!
```

1. Absorbance data is to be stored in the field `DS.Abs` with the dimensions `[DS.nSample x numel(DS.Abs_wave]`
2. The field `DS.Abs_wave` with the dimensions of `[size(DS.Abs,2) x 1]`.
3. [slopes,metadata,model] = slopefit(DS,Name,Value)

```
%describecomp
peaks=describecomp(LSmodel,4,[])

%convert raw PARAFAC model scores to Fmax
[F4,scores4]=scores2fmax(val4A1,4) %results copied in excel!

%exports scores with Fmax of model --> modify for OpenFluor external validation!
cd 'C:\Users\calderom\OneDrive - Dundalk Institute of Technology\PARAFACcourse\Week3'
modelexport(val4A1,4,'AQUAfv_modelexport.xlsx');

%Also option to export scores of all samples, even the removed ones
%Extra care must be taken when interpreting the scores for outlier samples.
[FMax,B,C,FMaxFull,Proj]=modelexport(val4A1,4,'AQUAoutlier_modelexport.xlsx', %?%); %not workin

%Also option to export metadata info (without outliers!)
[FMax,B,C]=modelexport(val4A1,4,'AQUAmeta_modelexport.xlsx',[], {'SampleID','Treatment','Replic
[FMax,B,C]=modelout(val4A1,4,'AQUAmetaout_modelexport.csv',[], {'SampleID','Treatment','Replica
```