Automated processing of Raman spectral data

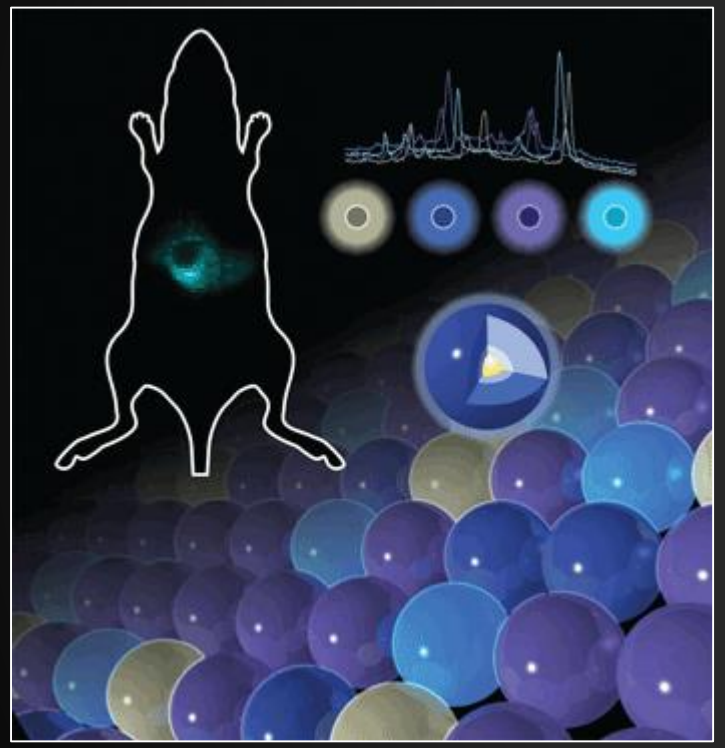
Mathematical and statistical tools for qualitative and quantitative analysis

Pablo Sobron
SETI Institute & Impossible Sensing

Impossible Sensing

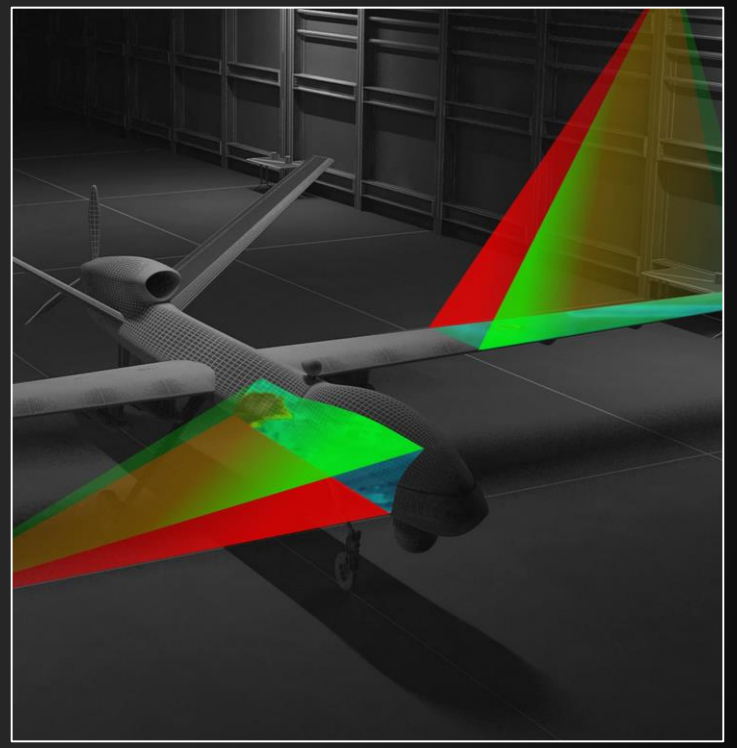
SETI Institute

Operator Limited

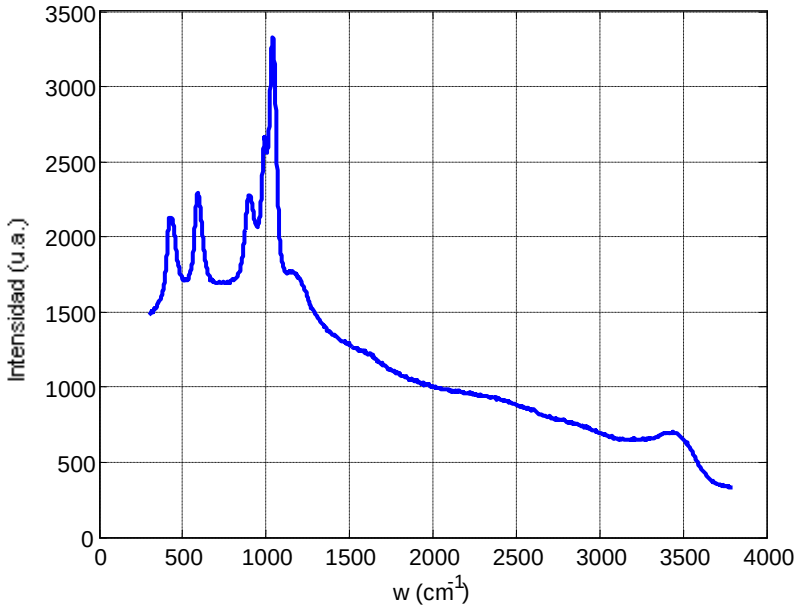


Bandwidth Limited

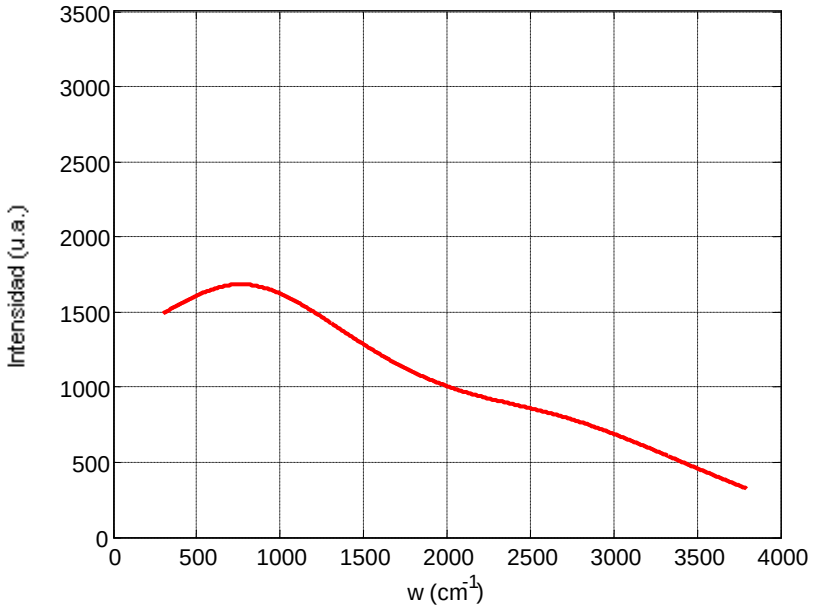
Application Driven



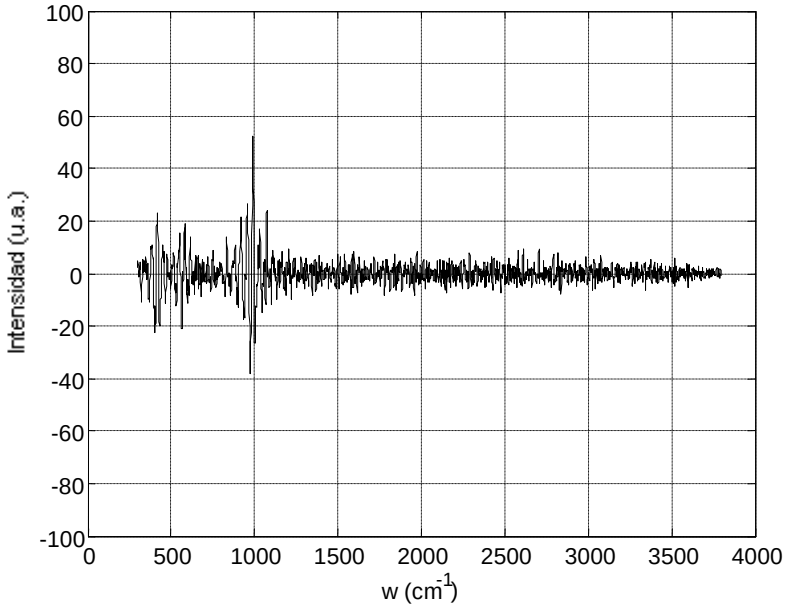
1. Pre-processing
2. Multivariate Analysis
3. Data Fusion


Intensidad (u.a.)
w (cm-1)
0
500
1000
1500
2000
2500
3000
3500
4000

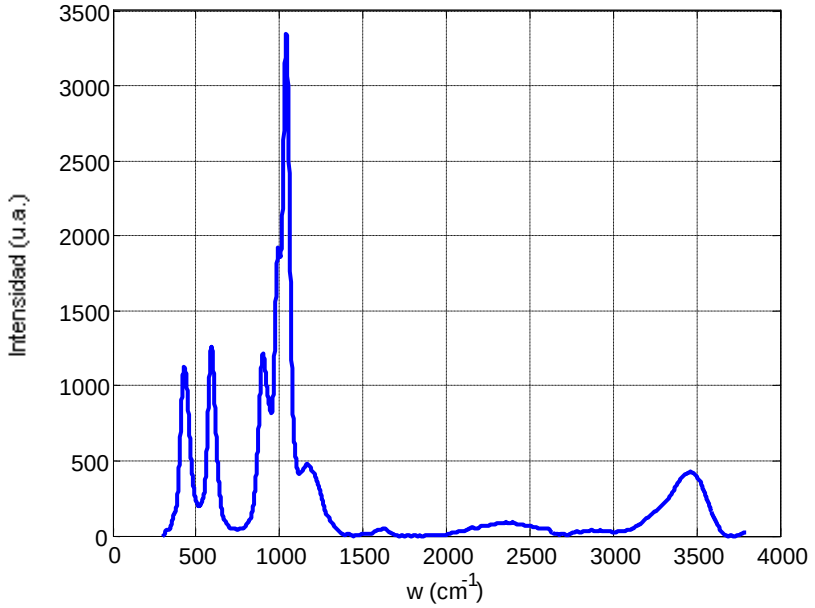
a


Intensidad (u.a.)
w (cm-1)
0
500
1000
1500
2000
2500
3000
3500
4000

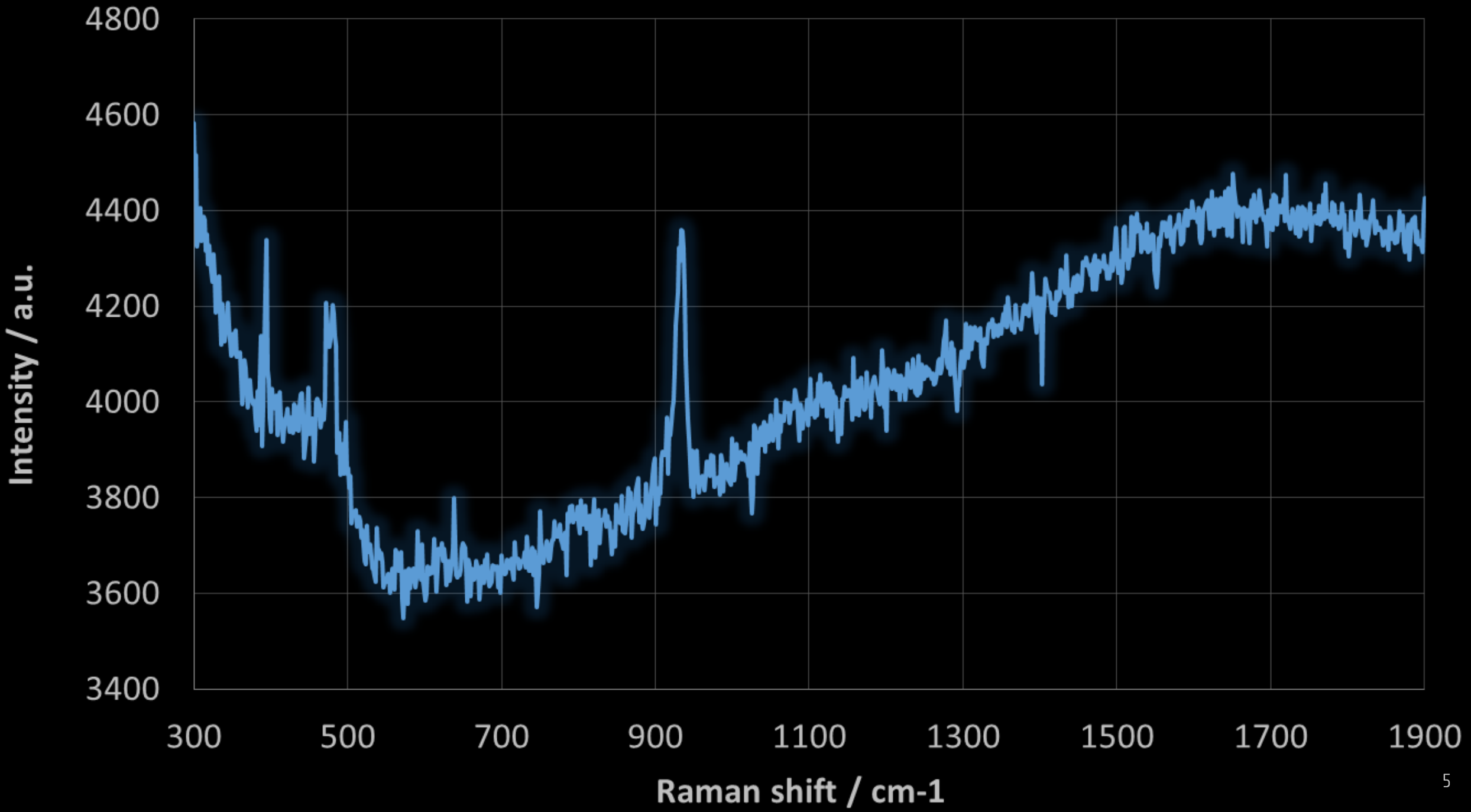
b

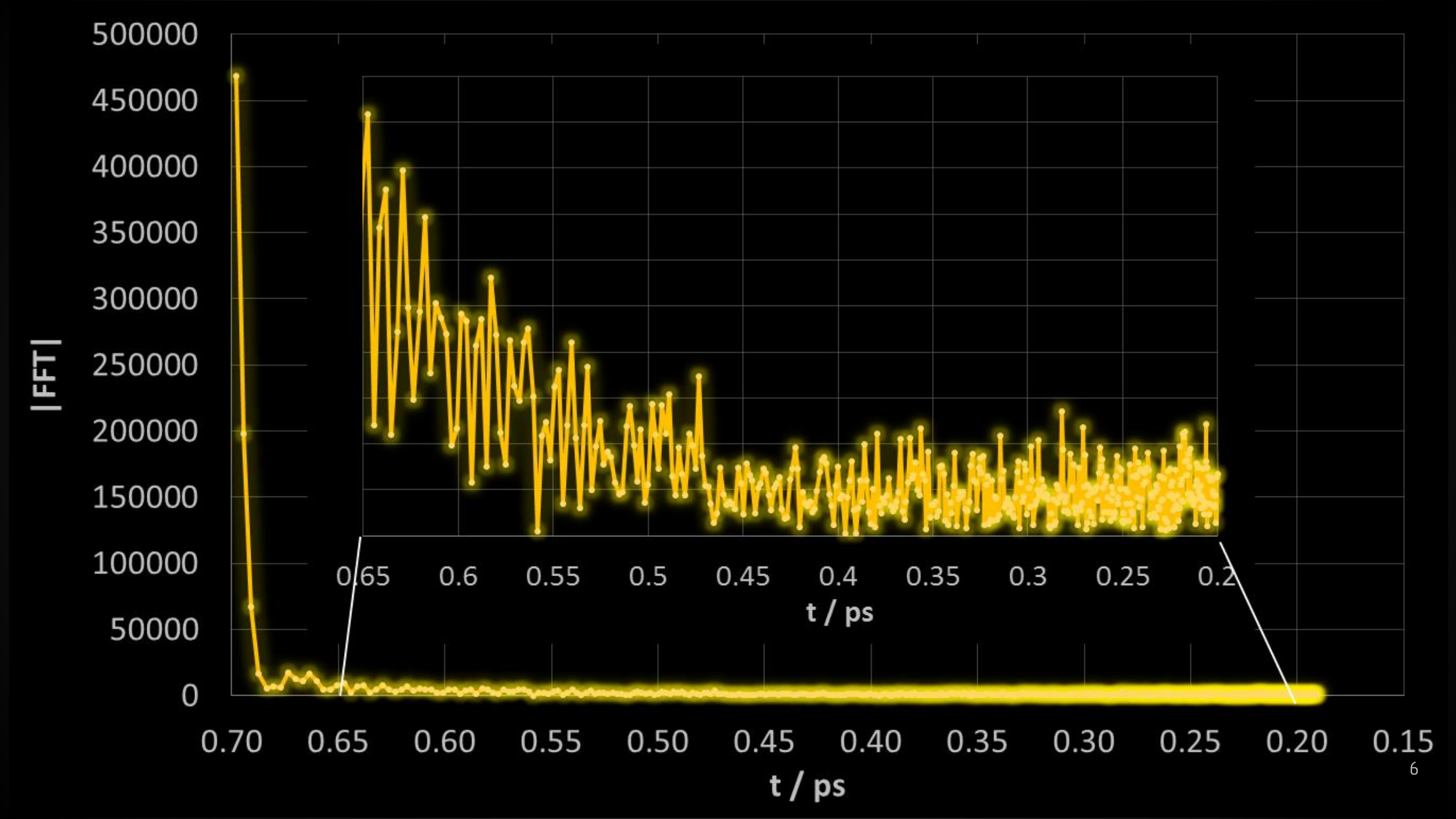

Intensidad (u.a.)
w (cm-1)
100
80
60
40
20
0
-20
-40
-60
-80
-100
0
500
1000
1500
2000
2500
3000
3500
4000

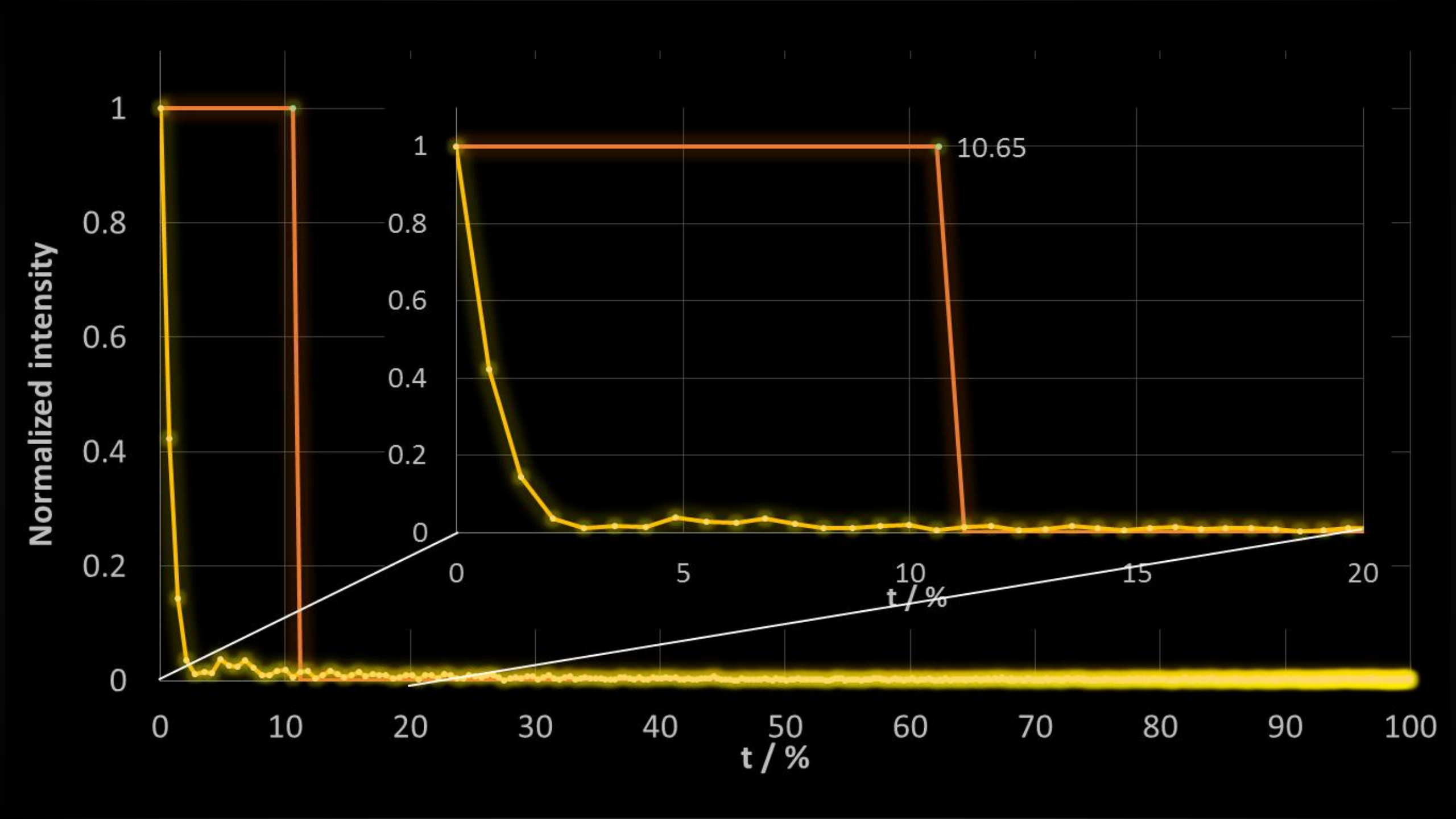
c

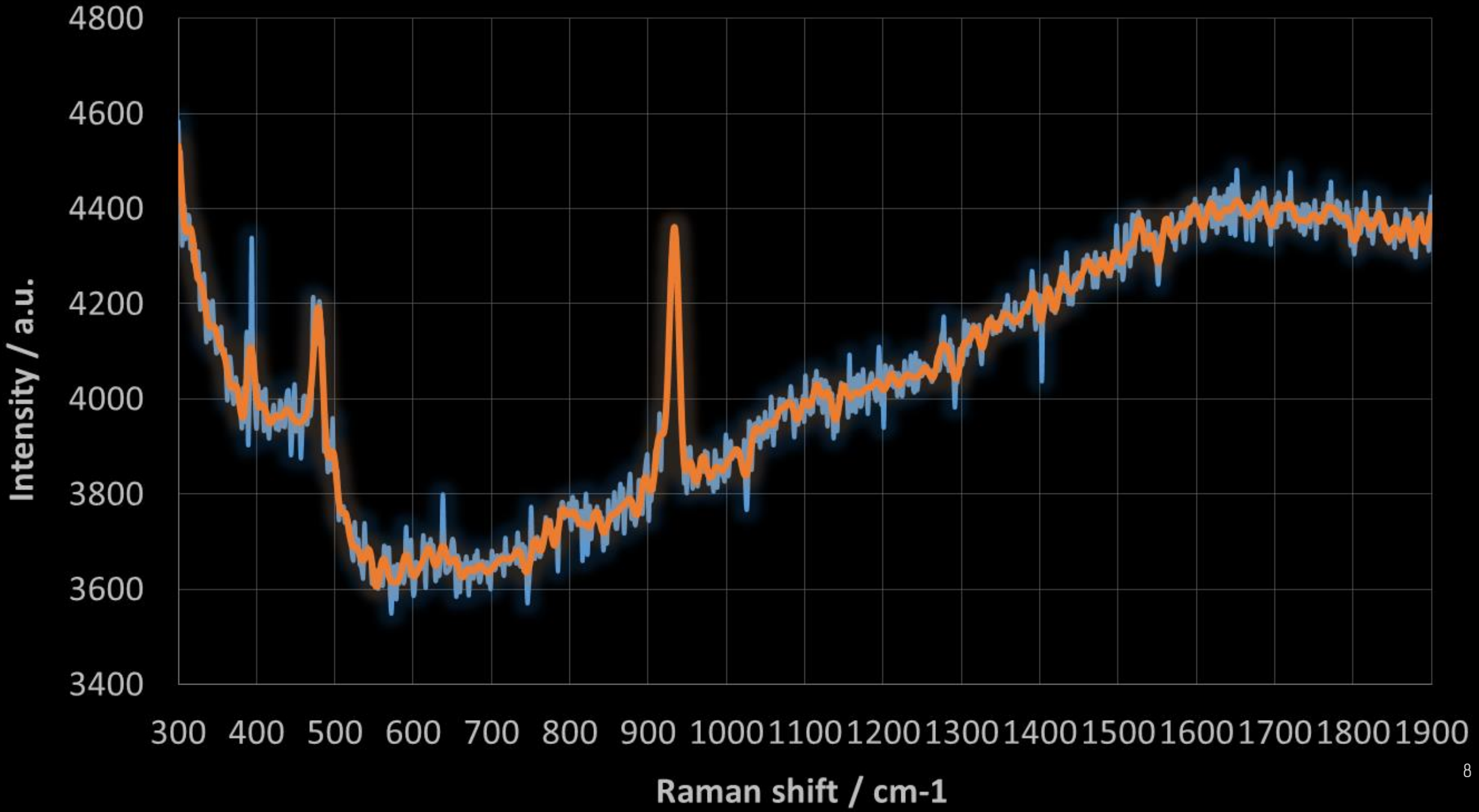

Intensidad (u.a.)
w (cm-1)
0
500
1000
1500
2000
2500
3000
3500
4000

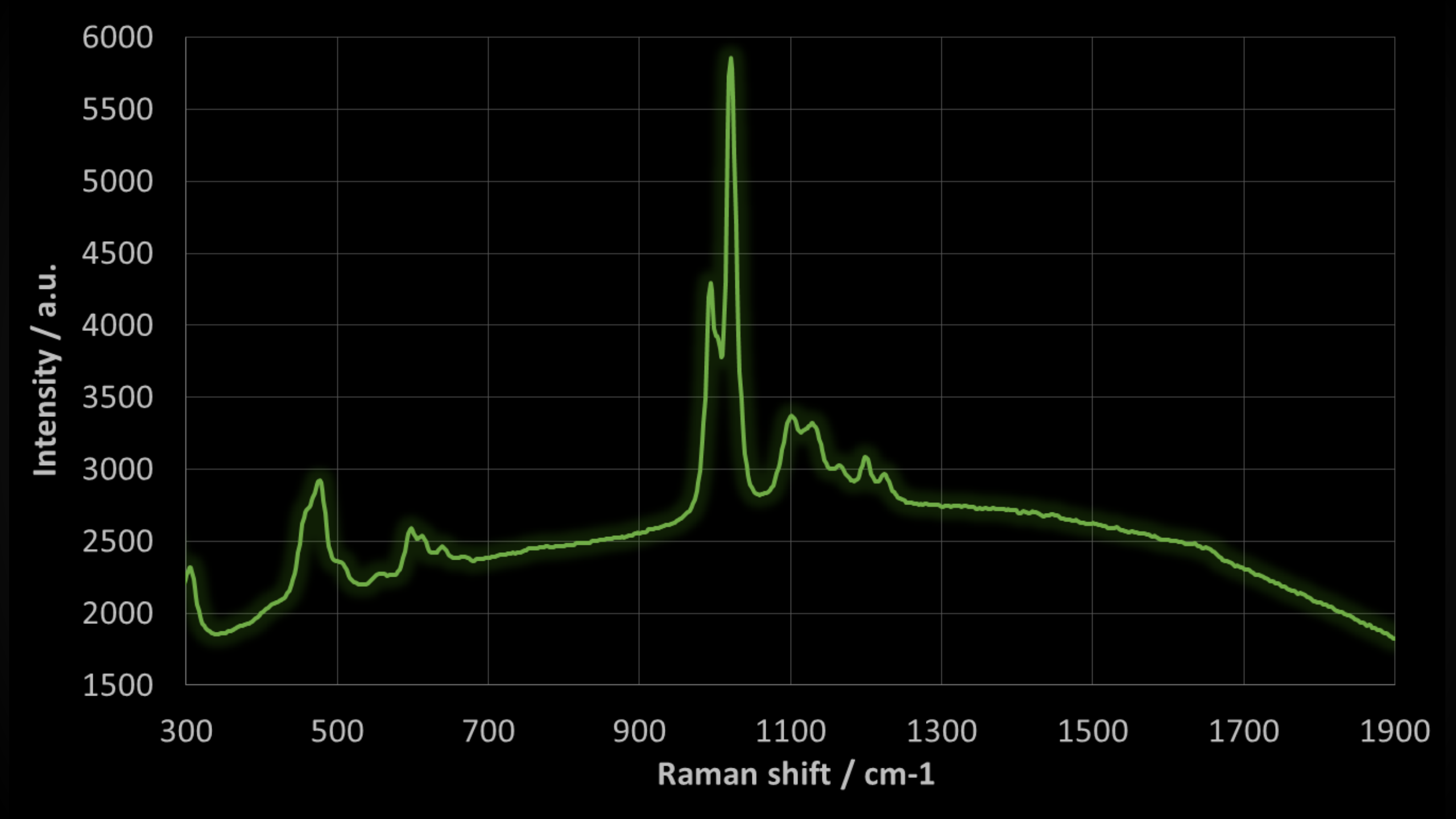
d

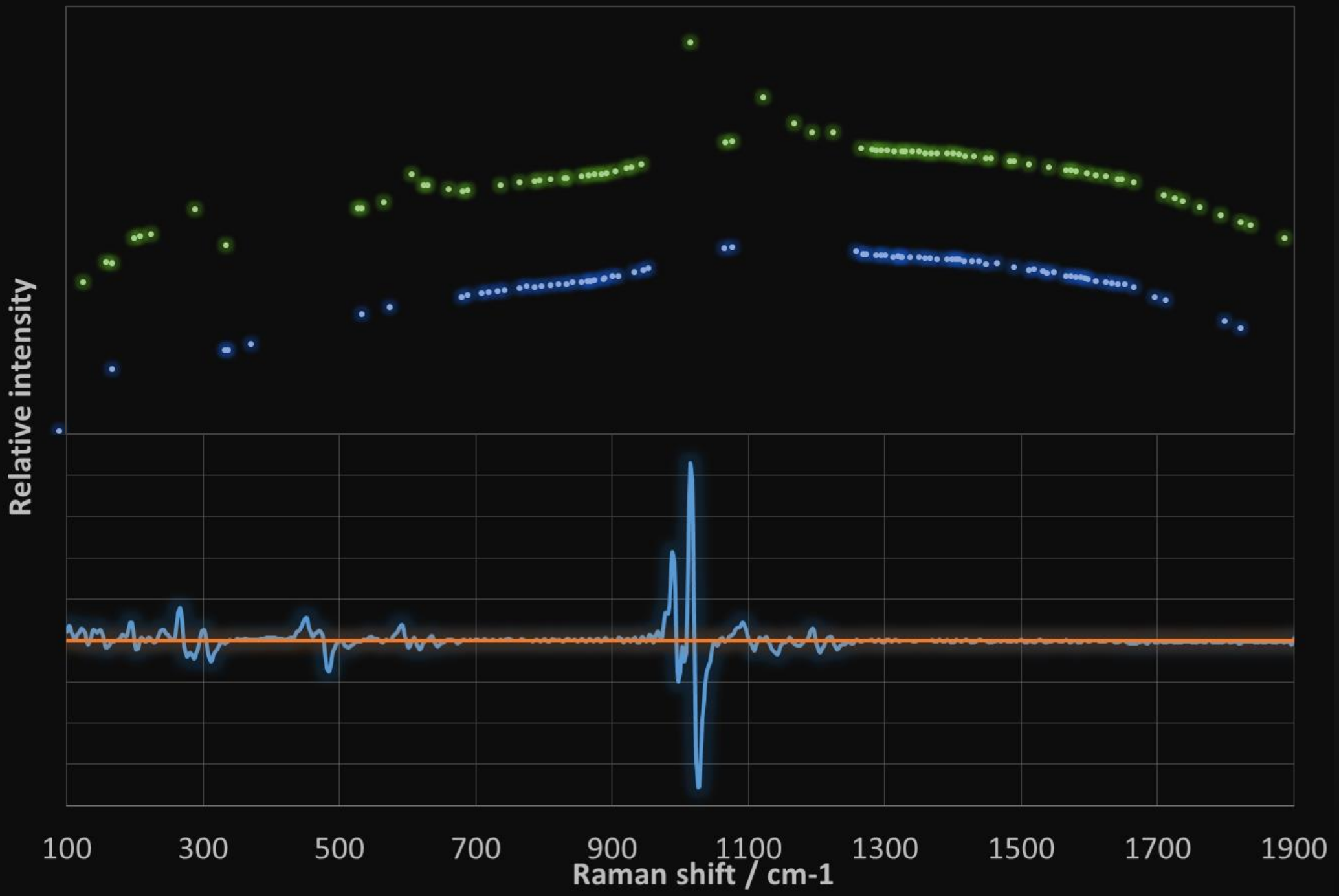

4800
4600
4400
4200
4000
3800
3600
3400
Intensity / a.u.
300
500
700
900
1100
1300
1500
1700
1900
Raman shift / cm-1

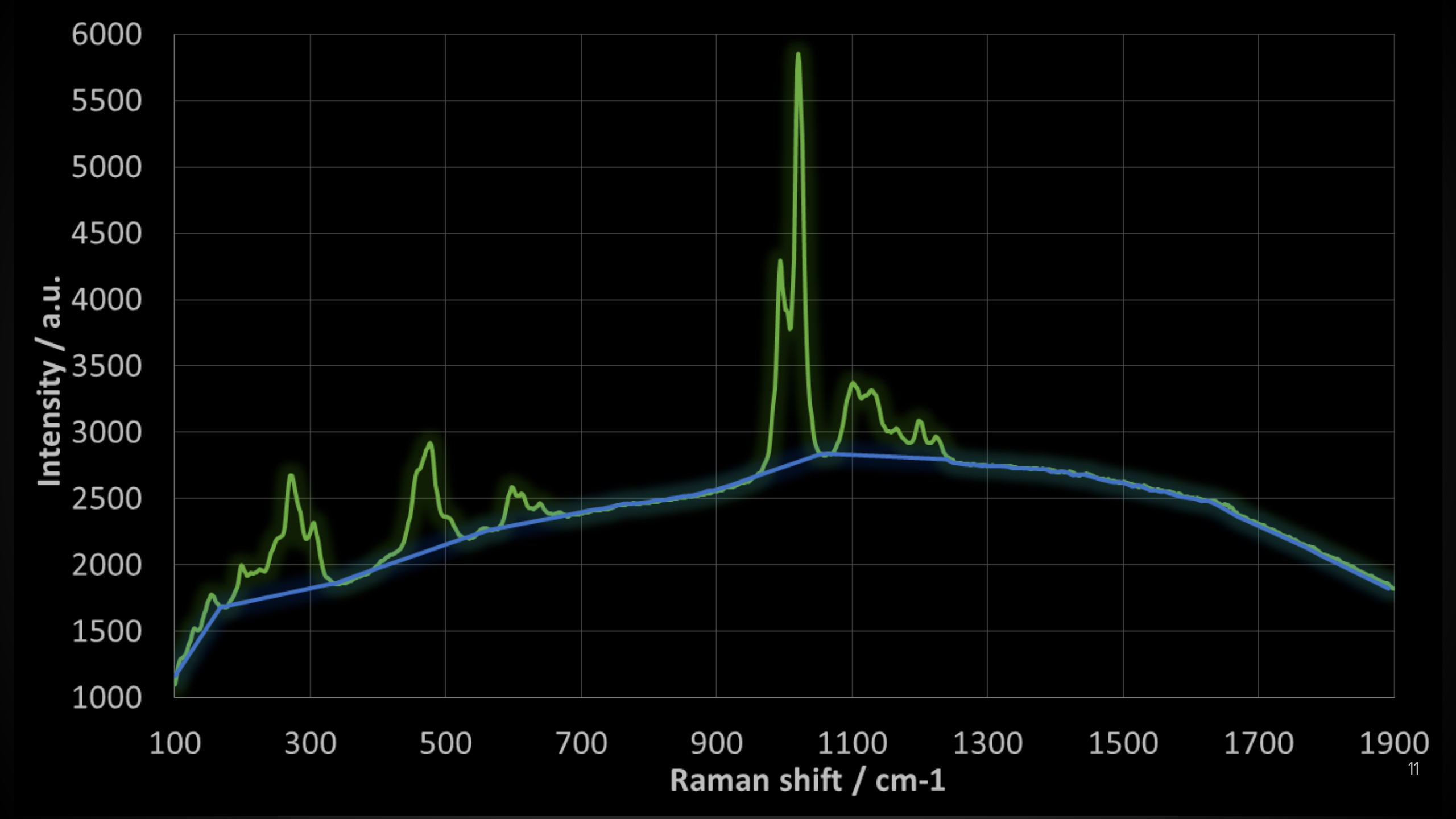

500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
|FFT|
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
t / ps
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
t / ps

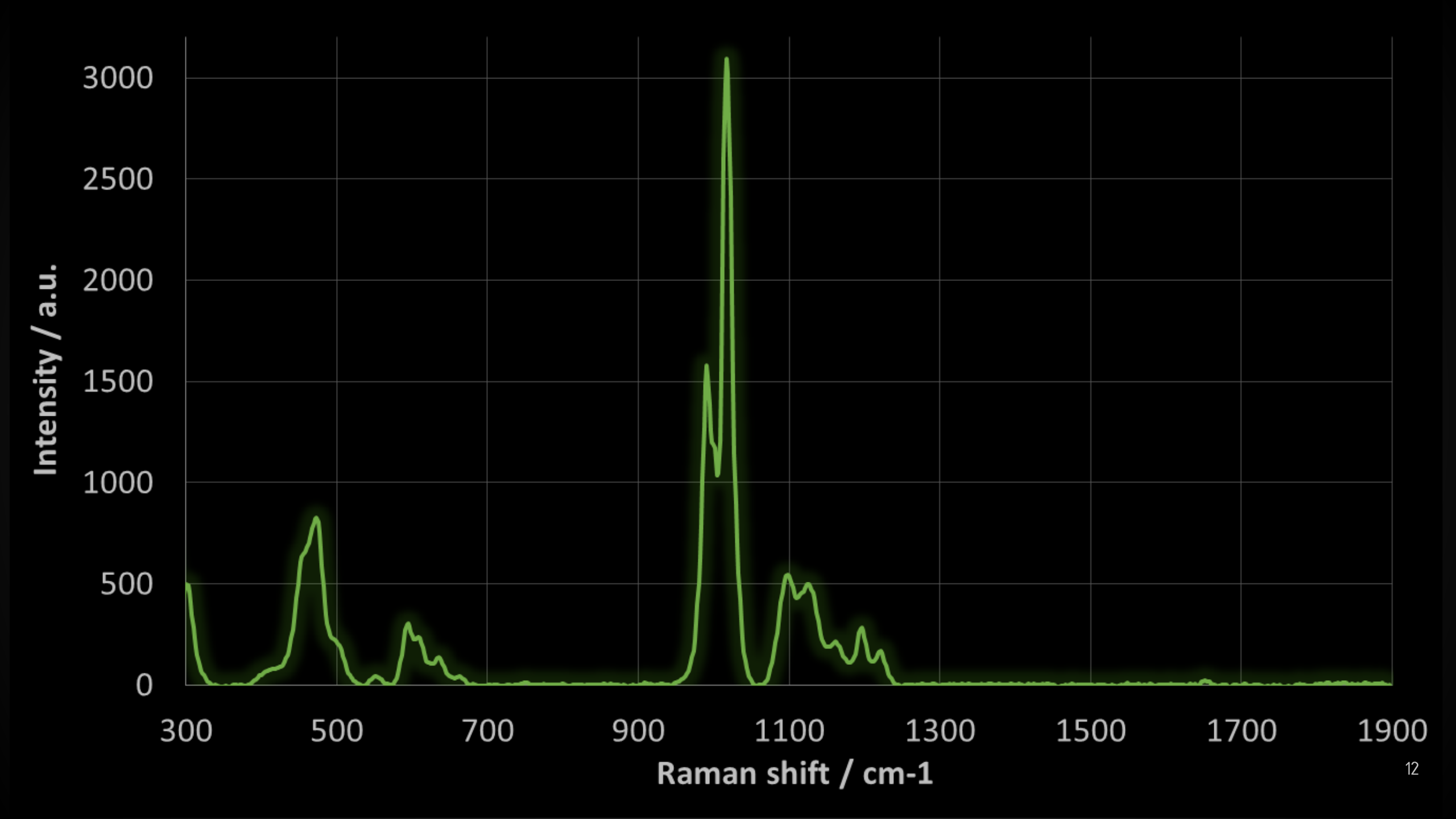

Normalized intensity
1
0.8
0.6
0.4
0.2
0
0
10
20
30
40
50
60
70
80
90
100
t / %
10.65
0
5
10
15
20
t / %


4800
4600
4400
4200
4000
3800
3600
3400
Intensity / a.u.
300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900
Raman shift / cm-1


6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
Intensity / a.u.
300
500
700
900
1100
1300
1500
1700
1900
Raman shift / cm-1

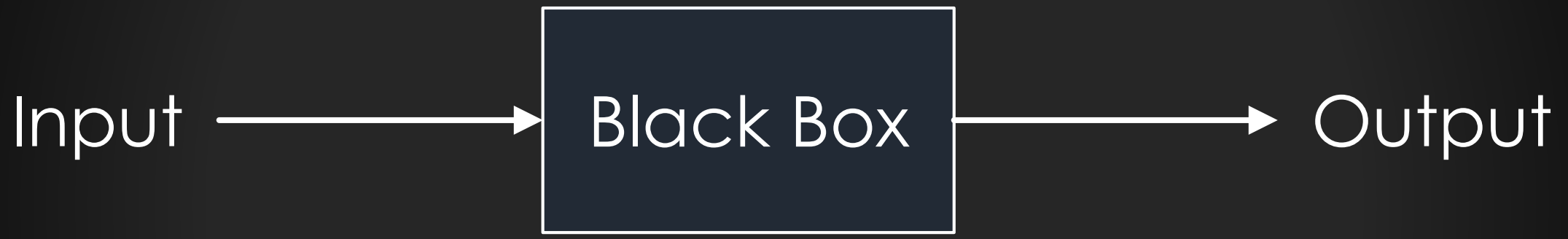

Relative intensity
100
300
500
700
900
1100
1300
1500
1700
1900
Raman shift / cm-1


6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
Intensity / a.u.
100
300
500
700
900
1100
1300
1500
1700
1900
Raman shift / cm-1


3000
2500
2000
1500
1000
500
0
Intensity / a.u.
300
500
700
900
1100
1300
1500
1700
1900
Raman shift / cm-1

Multivariate Analysis

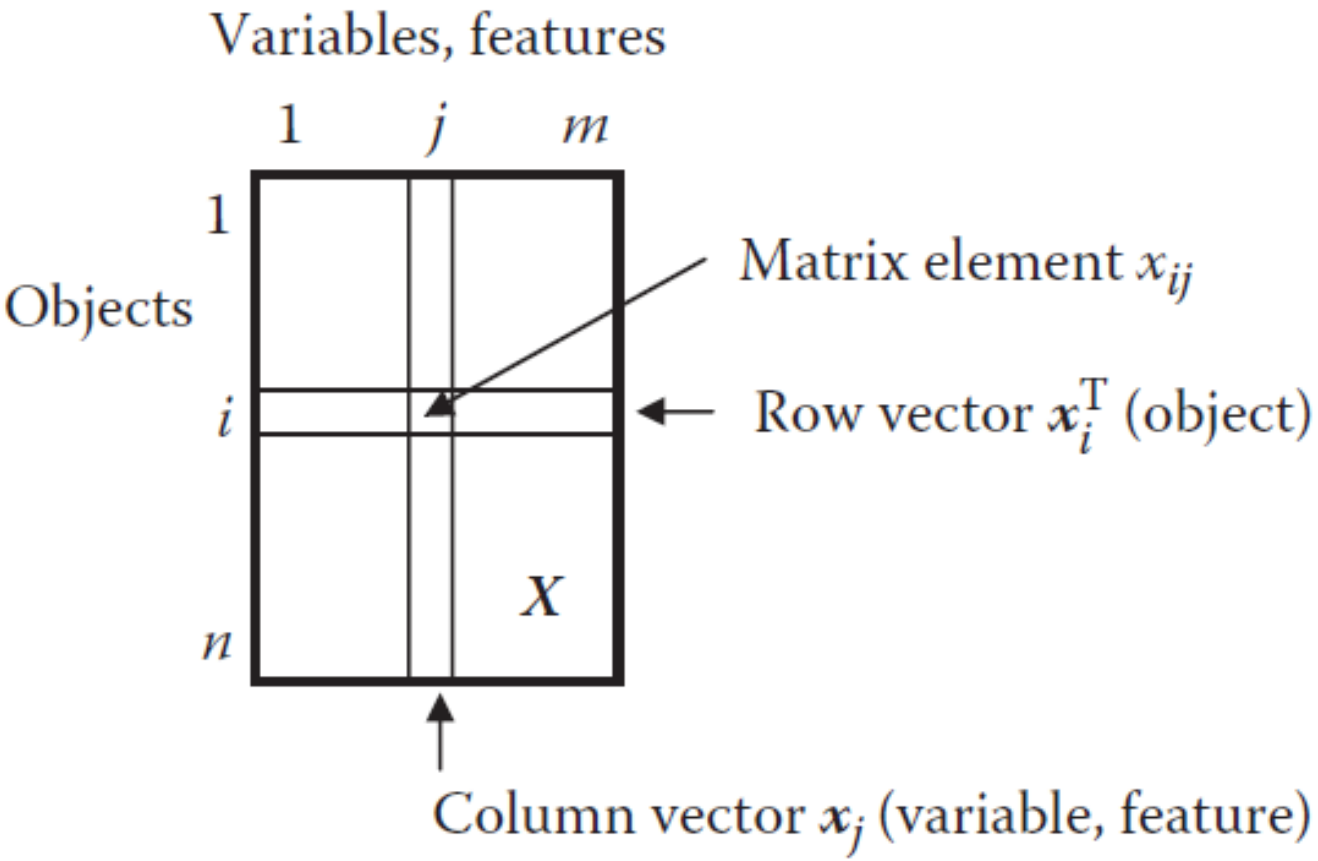
Multivariate Analysis

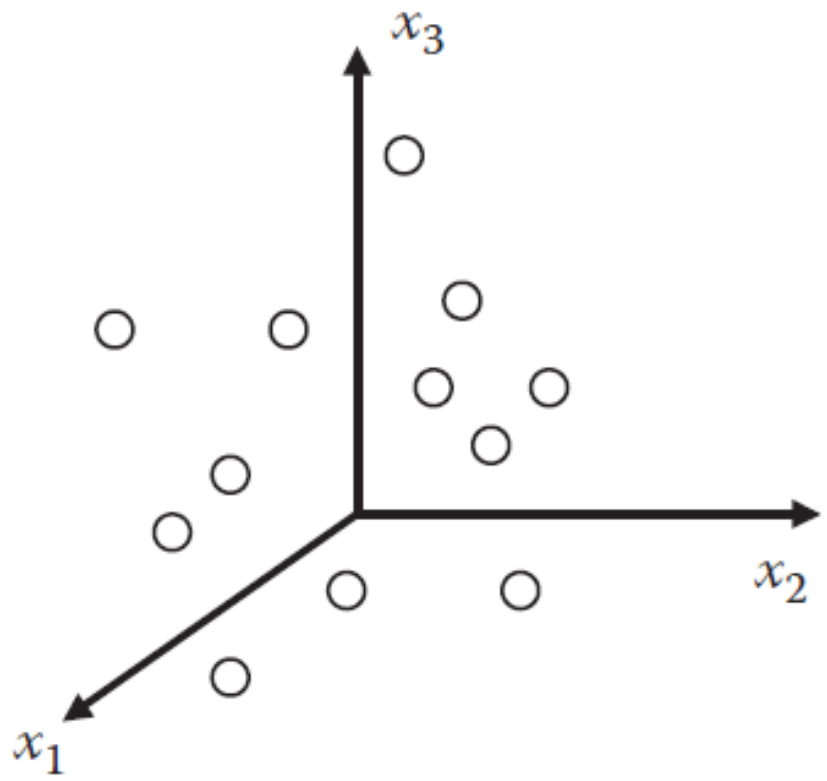


Multivariate Analysis

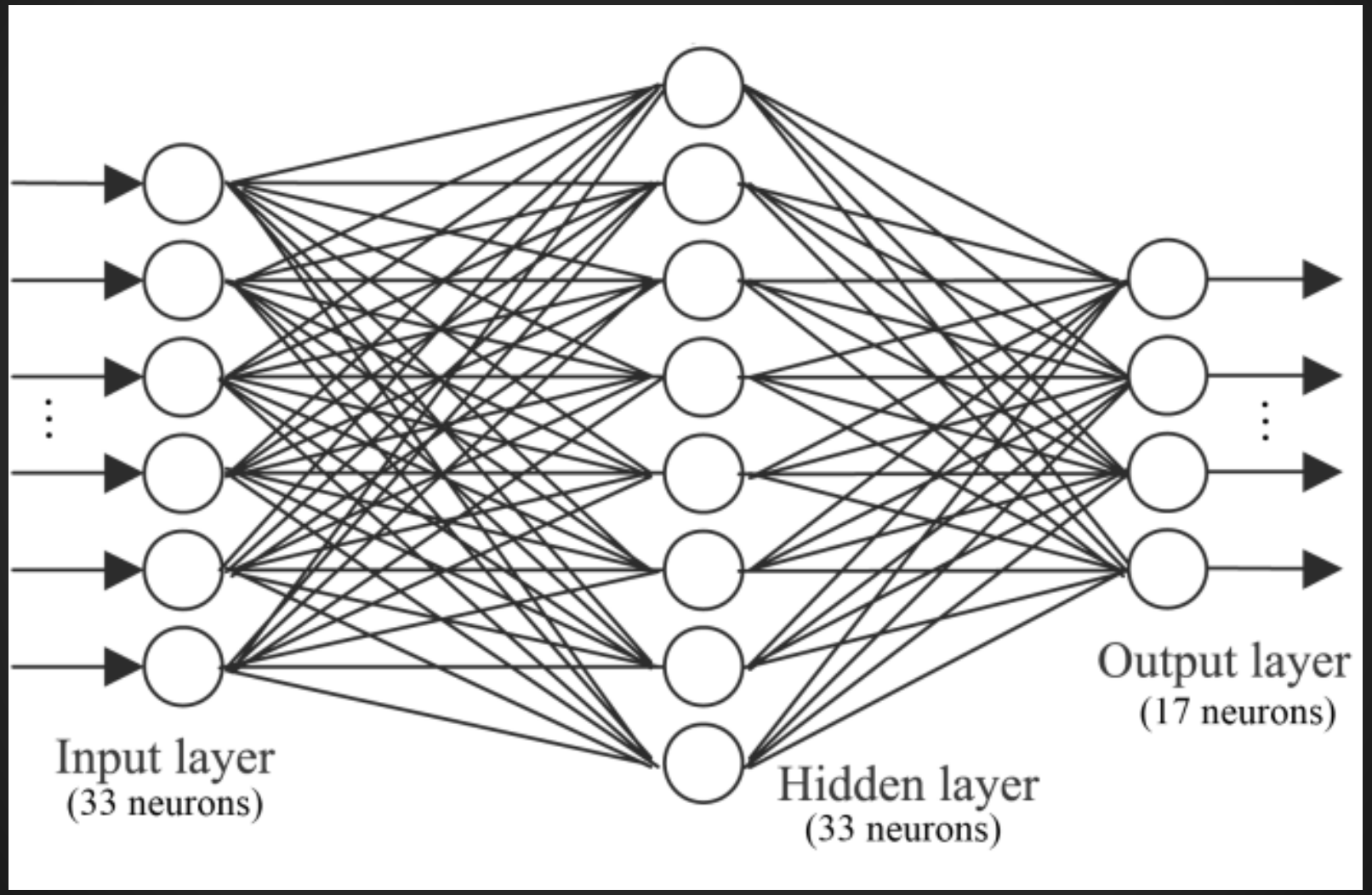
Principal Component Analysis (PCA)
Partial Least Squares Regression (PLSR)
Artificial Neural Networks (ANN)

PCA and PLSR

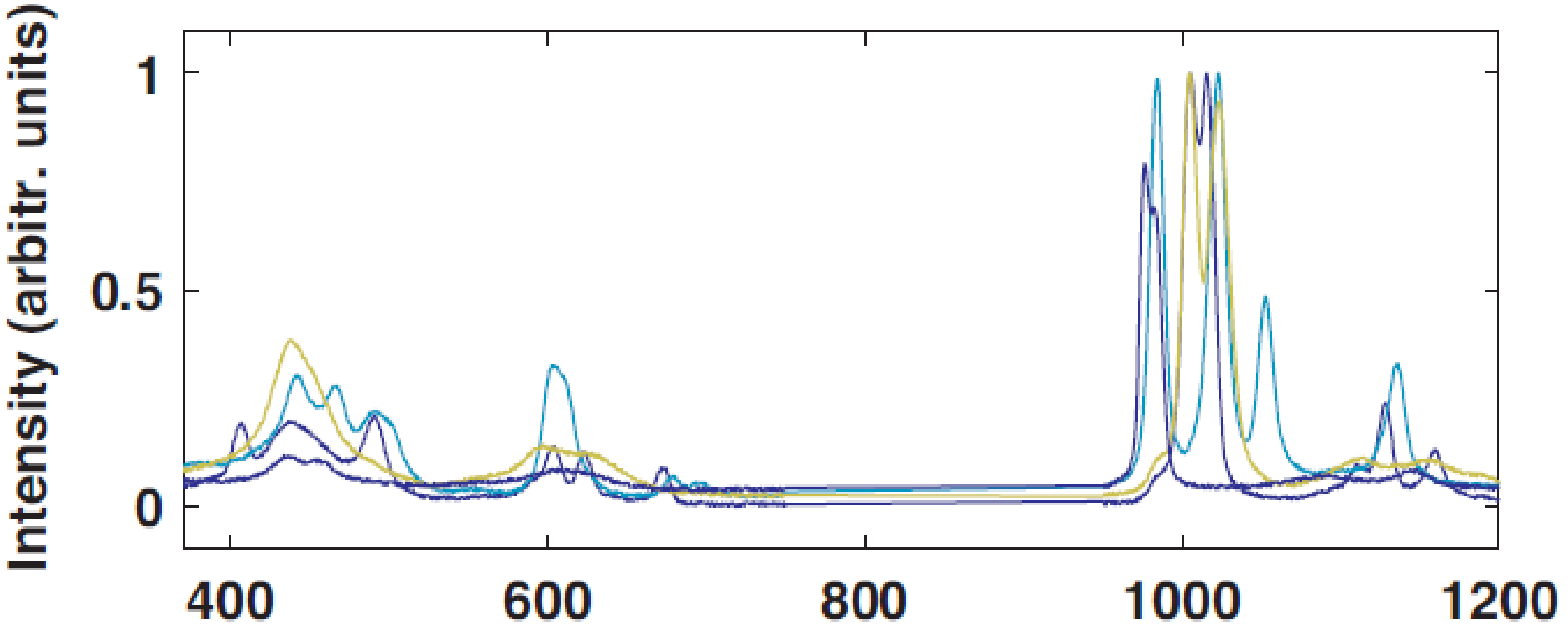

Variables, features
1
j
m
1
Objects
i
n
X
Matrix element x_{ij}
Row vector x_i^T (object)
Column vector x_j (variable, feature)

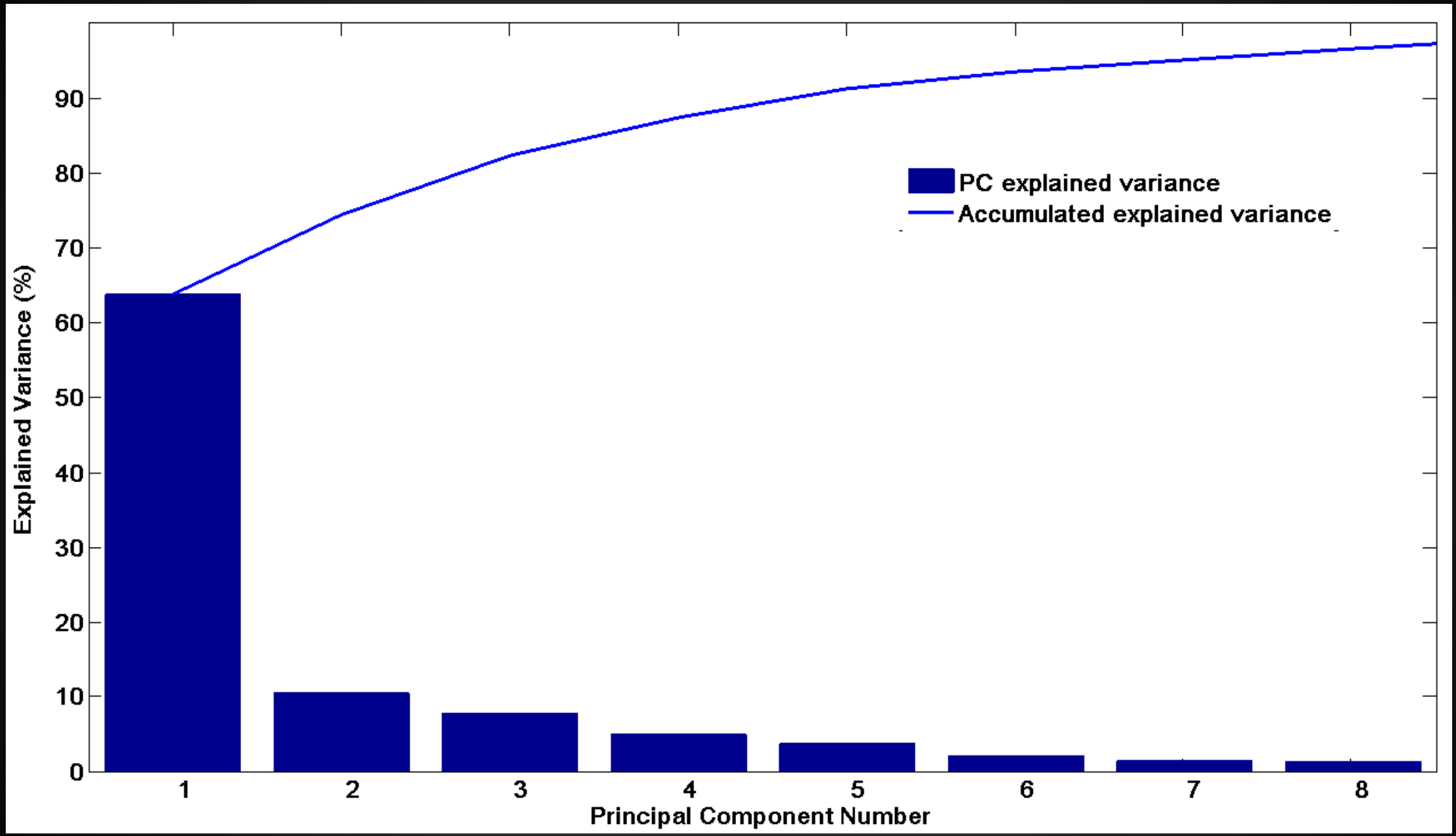

x_3
x_2
x_1

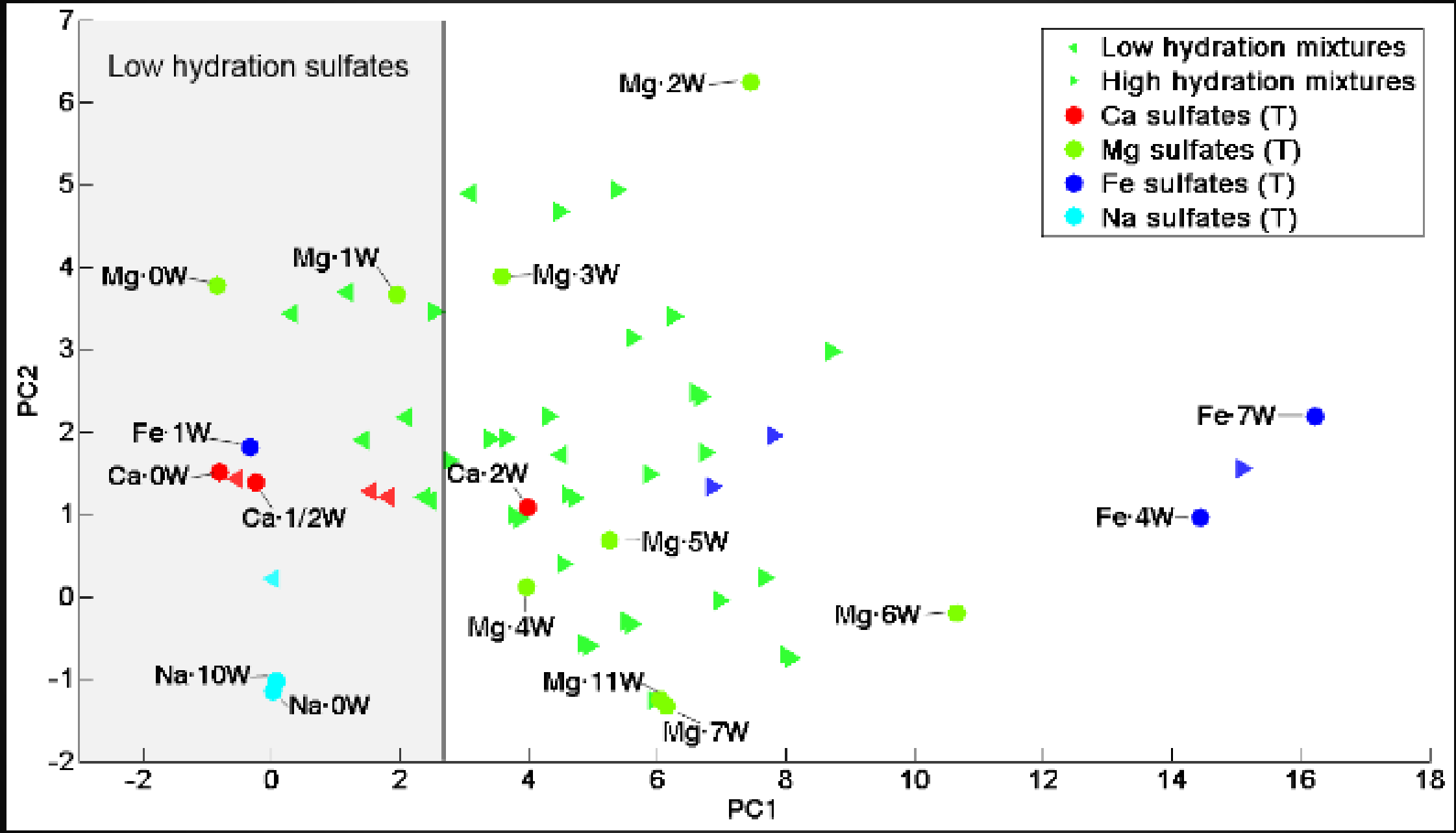
Artificial Neural Networks (ANN)

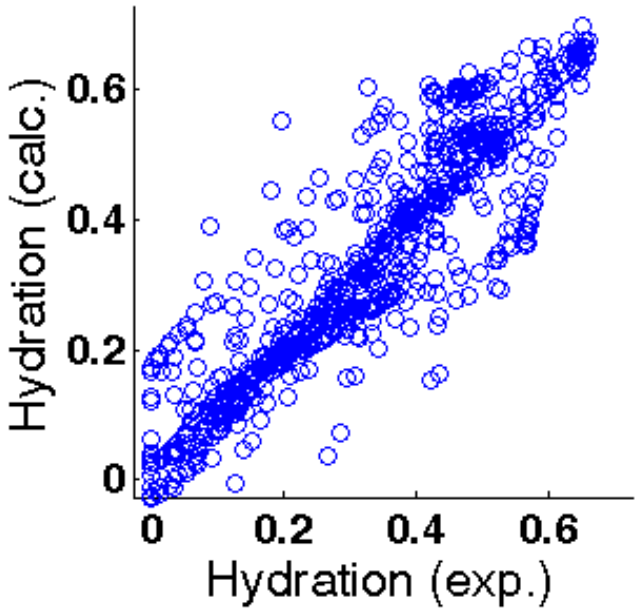


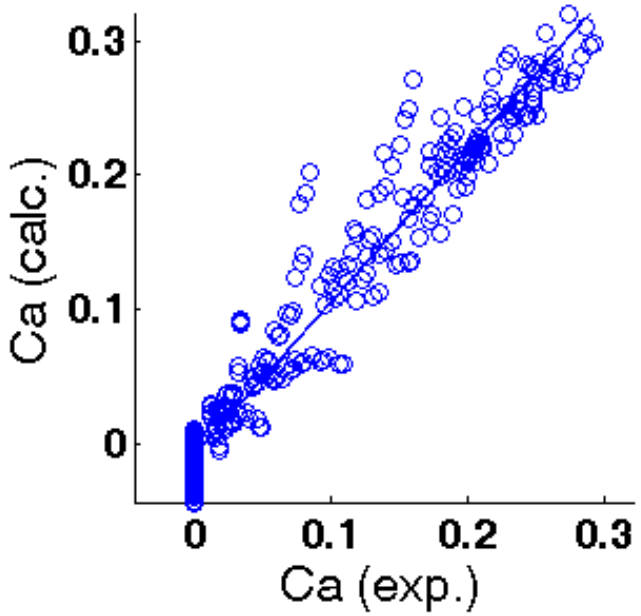
Hydration state	Mg	Ca	Fe	Na
Anhydrous	Anhydrous Mg-Sulfate	Anhydrite	--	Thenardite
1/2 H_2O	--	Bassanite	--	--
1 H_2O	Kieserite	--	Szomolnokite	--
2 H_2O	Sanderite	Gypsum	--	--
3 H_2O	Mg-sulfate tri-hydrate	--	--	--
4 H_2O	Starkeyite	--	Rozenite	--
5 H_2O	Pentahydrate	--	--	--
6 H_2O	Hexahydrate	--	--	--
7 H_2O	Epsomite	--	Melanterite	--
10 H_2O	--	--	--	Glauber's salt
11 H_2O	Meridianiite	--	--	--

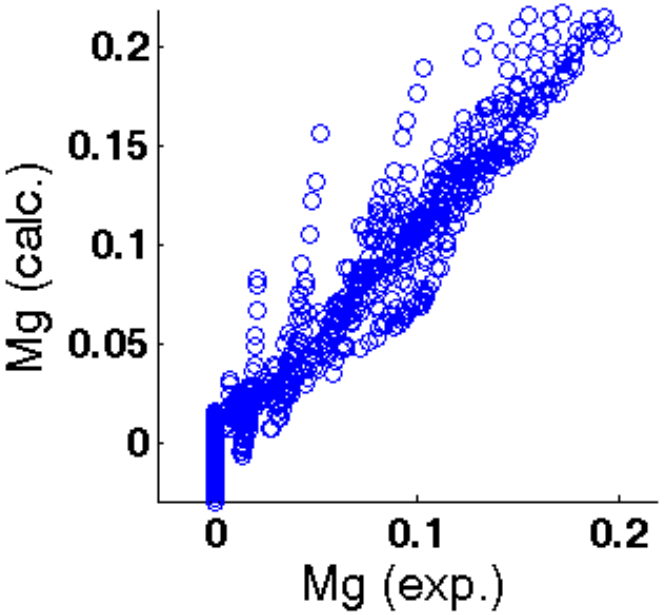

Intensity (arbitr. units)
1
0.5
0
400
600
800
1000
1200

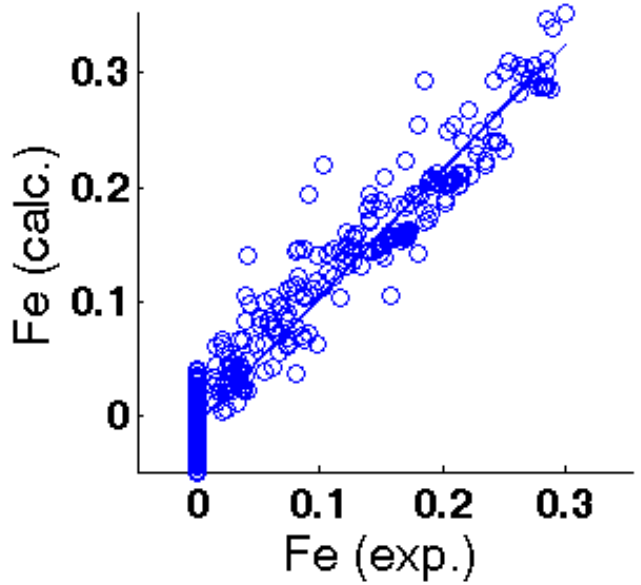

PC explained variance
Accumulated explained variance
Explained Variance (%)
0
10
20
30
40
50
60
70
80
90
1
2
3
4
5
6
7
8
Principal Component Number

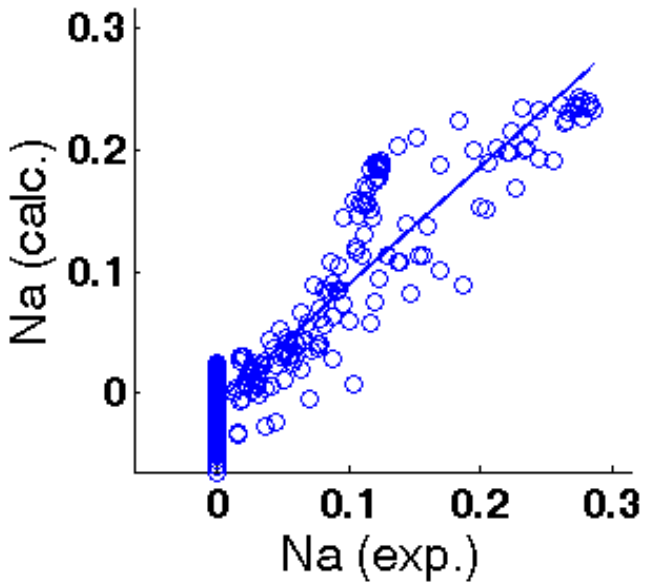

Low hydration sulfates
Low hydration mixtures
High hydration mixtures
Ca sulfates (T)
Mg sulfates (T)
Fe sulfates (T)
Na sulfates (T)
Mg·2W
Mg·0W
Mg·1W
Mg·3W
Fe·1W
Ca·0W
Ca·1/2W
Ca·2W
Fe·7W
Fe·4W
Mg·5W
Mg·4W
Mg·6W
Na·10W
Na·0W
Mg·11W
Mg·7W
PC1
PC2
7
6
5
4
3
2
1
0
-1
-2
-2
0
2
4
6
8
10
12
14
16
18

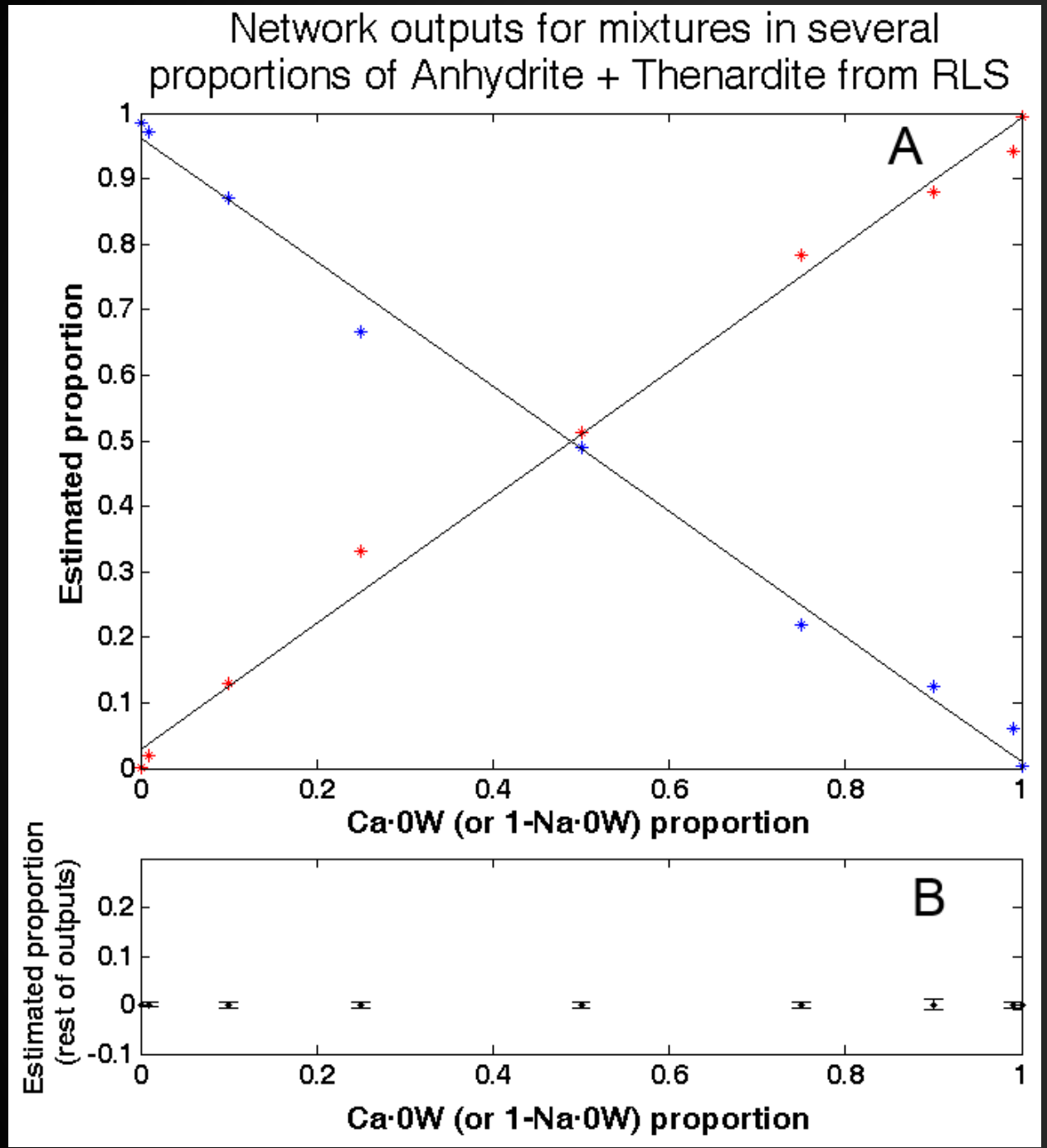

Hydration (calc.)
0
0.2
0.4
0.6
0
0.2
0.4
0.6
Hydration (exp.)

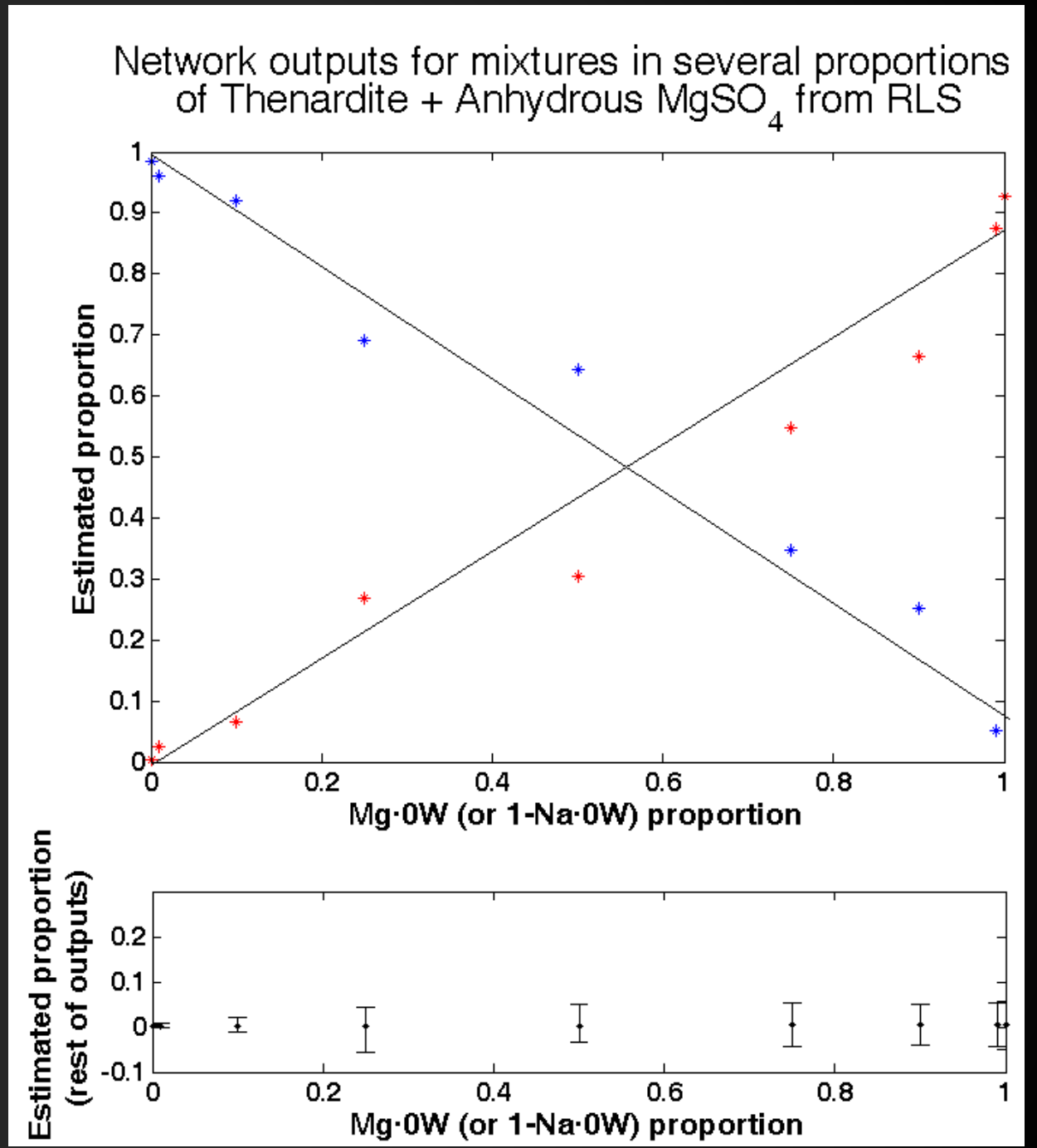

Ca (calc.)
0
0.1
0.2
0.3
0
0.1
0.2
0.3
Ca (exp.)


Mg (calc.)
0
0.05
0.1
0.15
0.2
0
0.1
0.2
Mg (exp.)


Fe (calc.)
0
0.1
0.2
0.3
0
0.1
0.2
0.3
Fe (exp.)


Na (calc.)
0
0.1
0.2
0.3
0
0.1
0.2
0.3
Na (exp.)


Network outputs for mixtures in several proportions of Anhydrite + Thenardite from RLS
A
Estimated proportion
Ca·0W (or 1-Na·0W) proportion
B
Estimated proportion (rest of outputs)
Ca·0W (or 1-Na·0W) proportion

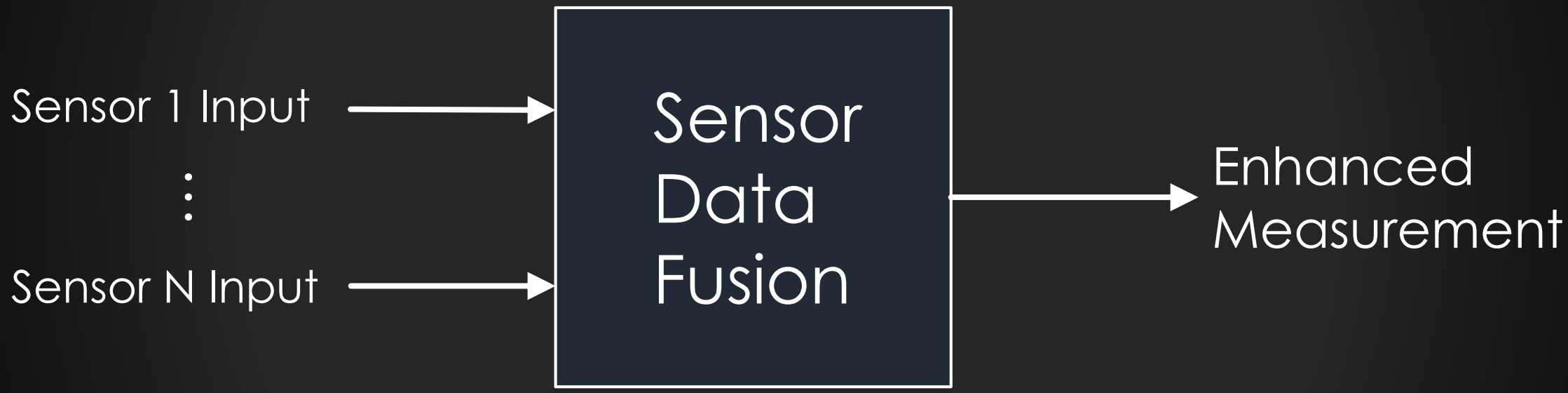

Network outputs for mixtures in several proportions of Thenardite + Anhydrous MgSO4 from RLS
Estimated proportion
Mg·0W (or 1-Na·0W) proportion
Estimated proportion (rest of outputs)
Mg·0W (or 1-Na·0W) proportion

Multivariate Analysis Summary

- Need to analyze **several thousands** of variables at the same time
- PCA and PLS calculate **principal components**
- ANNs use **non-linear** transfer functions
- PCA **separates** low- from high-hydrated sulfates
- PLS model outputs presented **good correlations** to the known responses
- ANN provides the **abundance** of the corresponding salt (**100%** accuracy)

Data Fusion

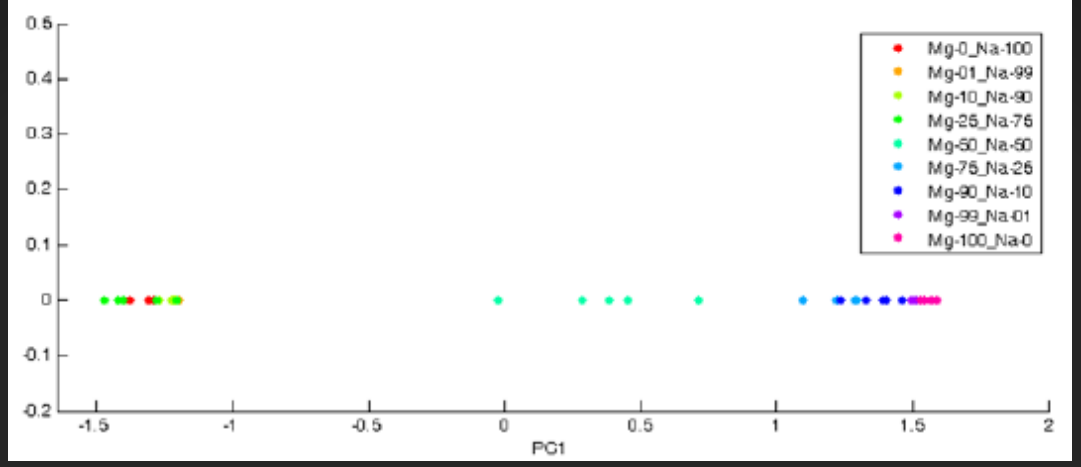
Data Fusion


Sensor 1 Input
Sensor N Input
Sensor Data Fusion
Enhanced Measurement

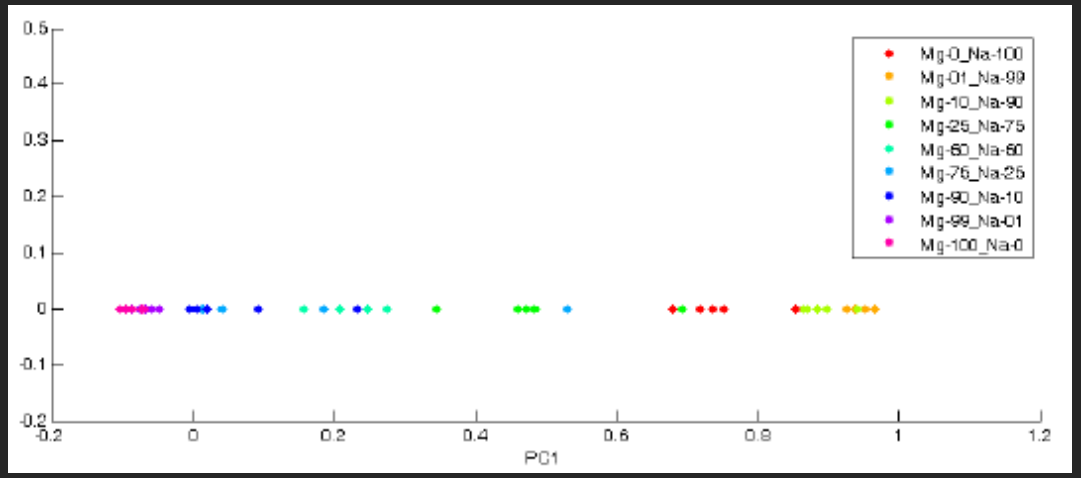
DATA FUSION IN PLANETARY LIBS+RAMAN SPECTROSCOPY

- A set of 9 binary mixtures of $MgSO_4$ and $NaSO_4$
- Collocated LIBS and Raman spectra recorded
- Low-level fusion
- 3 PCA models trained and validated

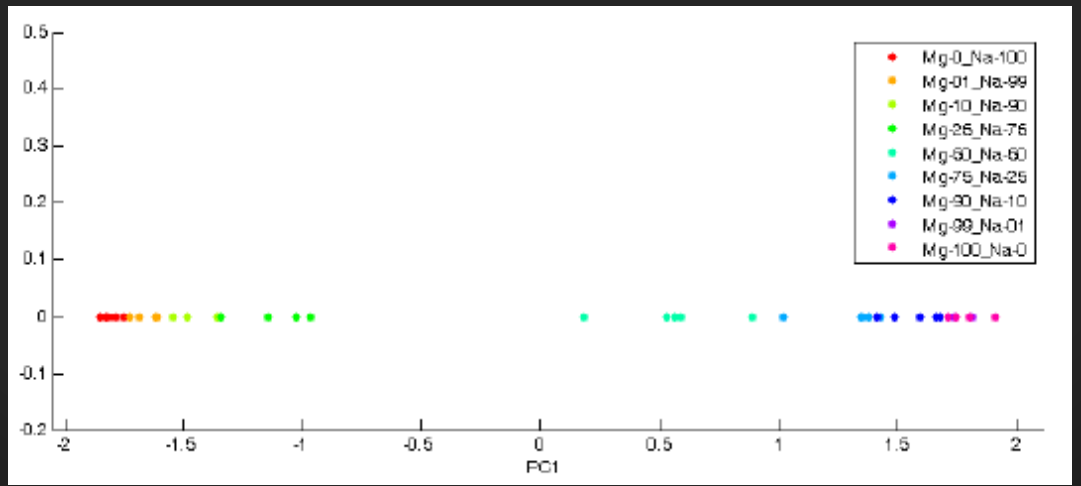
Raman Only



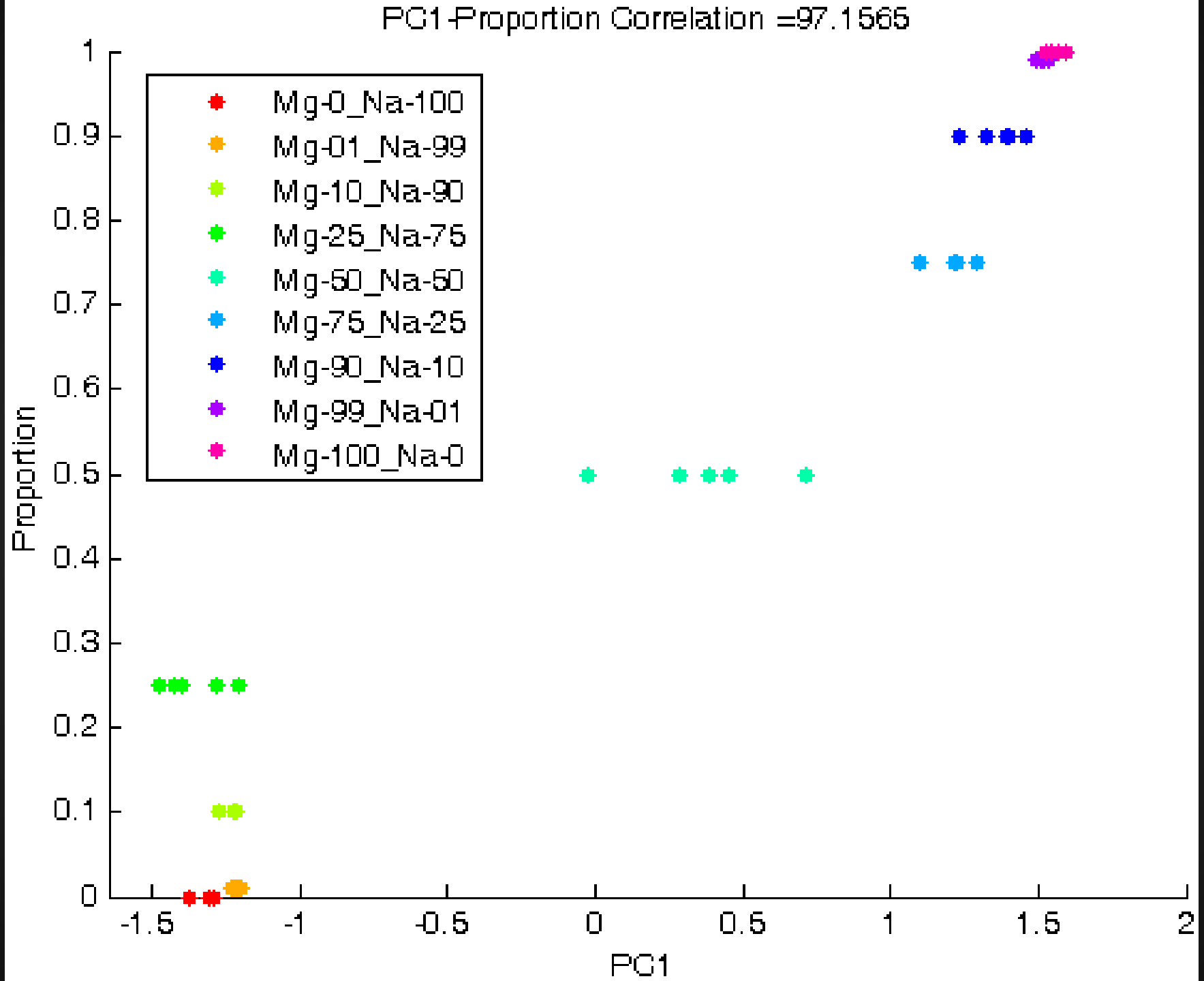
LIBS Only



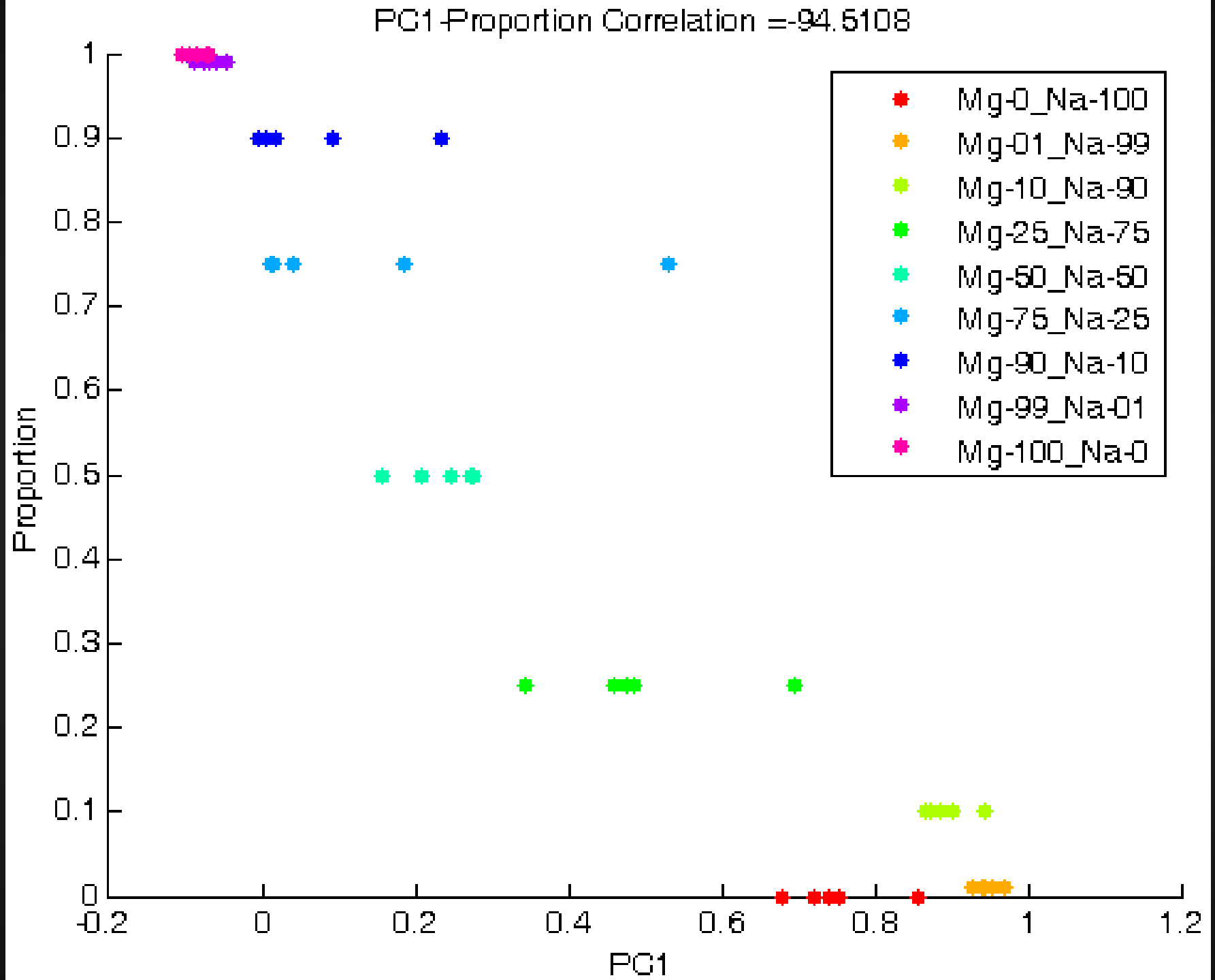
R + L



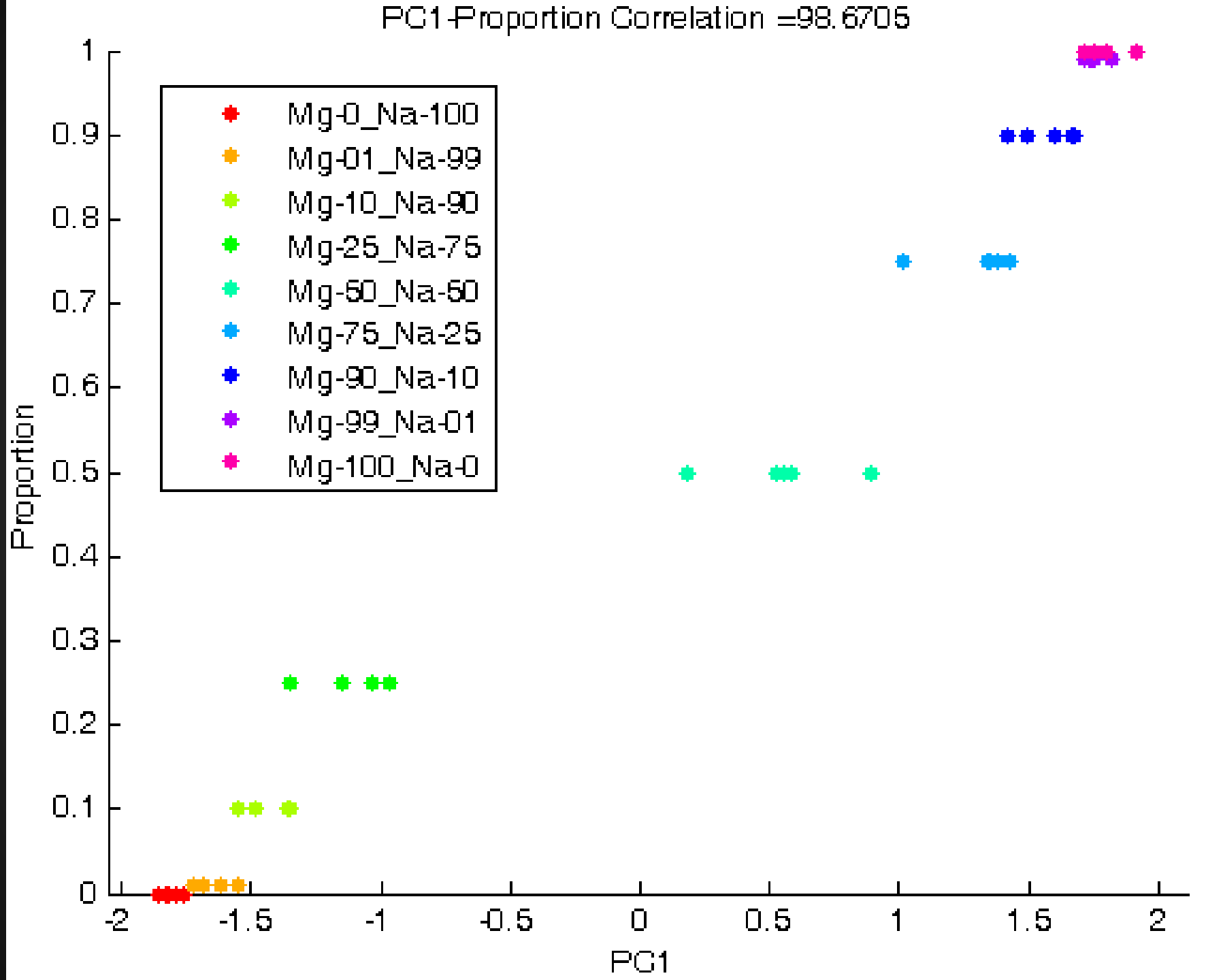
Raman Only


PC1-Proportion Correlation =97.1565
Mg-0_Na-100
Mg-01_Na-99
Mg-10_Na-90
Mg-25_Na-75
Mg-50_Na-50
Mg-75_Na-25
Mg-90_Na-10
Mg-99_Na-01
Mg-100_Na-0
Proportion
PC1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-1.5
-1
-0.5
0
0.5
1
1.5
2

LIBS Only


PC1-Proportion Correlation =-94.5108
Proportion
PC1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
Mg-0_Na-100
Mg-01_Na-99
Mg-10_Na-90
Mg-25_Na-75
Mg-50_Na-50
Mg-75_Na-25
Mg-90_Na-10
Mg-99_Na-01
Mg-100_Na-0

Raman + LIBS


PC1-Proportion Correlation =98.6705
Mg-0_Na-100
Mg-01_Na-99
Mg-10_Na-90
Mg-25_Na-75
Mg-50_Na-50
Mg-75_Na-25
Mg-90_Na-10
Mg-99_Na-01
Mg-100_Na-0
Proportion
PC1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2

Rio Tinto's Mine of the Future™

Nautilus Minerals's seafloor mining

Real-time oil analysis

Maptek's Topodrone-100 UAV

NASA etc planetary mining

Real-time leak detection