# unmixR: Hyperspectral Unmixing in *R*

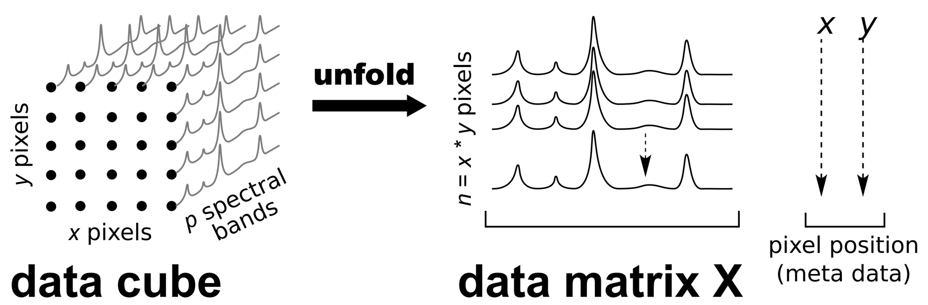
Conor McManus[1], Simon Fuller[2], Claudia Beleites[3]*, Bryan A. Hanson[4]*

1. National University of Ireland Maynooth, Maynooth/IRL 2. Computer Science Dept., National University of Ireland Maynooth/IRL
3. Leibniz Institute of Photonic Technolology, Jena/D 4. Dept. of Chemistry & Biochemistry, DePauw University, Greencastle IN USA

*Contact: claudia.beleites@ipht-jena.de & hanson@depauw.edu

## Hyperspectral Imaging

**Hyperspectral images** are 3D data sets of spectra collected over an $x, y$ grid.



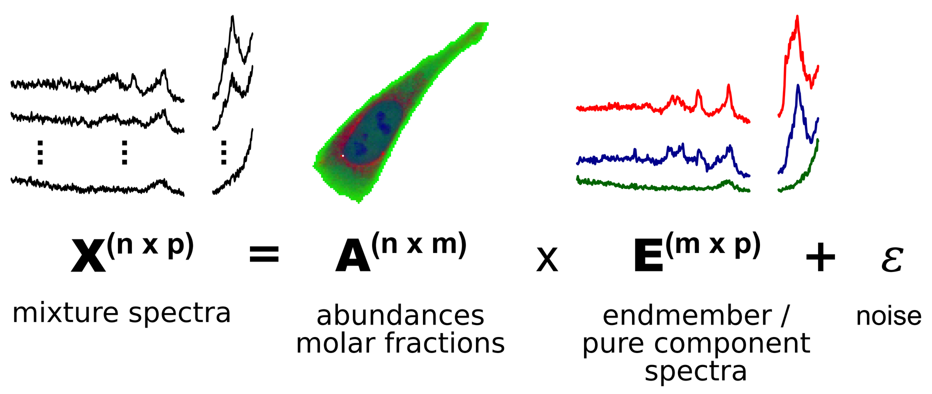

**Applications:** remote sensing/ airborne or satellite land imaging, biomedical microspectroscopy and art history investigations

**Spectra:** e.g. visible, near-infrared, mid-infrared, or Raman spectra.

## Spectral Unmixing

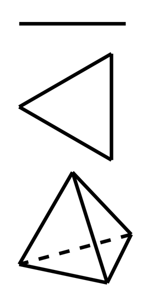
Identify $m$ pure component spectra in data, then derive respective concentrations.

**Bilinear statistical model:**




Mixture diagram for $m$ components: $(m-1)$-simplex in $m-1$ dimensions (***A***).

| | | |
|---|---|---|
| 2 components | 1-simplex | line |
| 3 components | 2-simplex | triangle |
| 4 components | 3-simplex | tetrahedron |



*Vertices are pure component spectra.*

**Assumptions:**

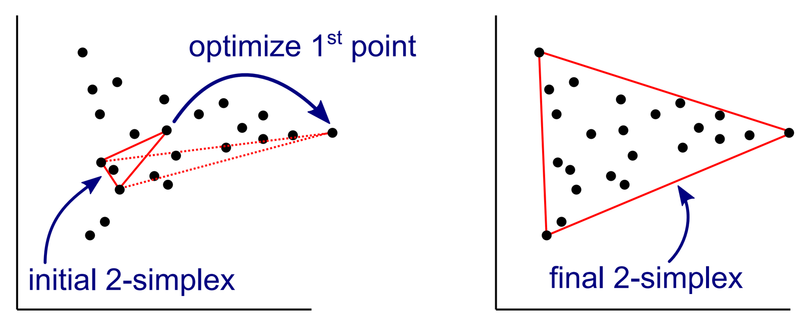
- Data consists of *mixture* spectra
- Spectra of pure components are available somewhere in the data ***X***
- Not too much noise on measurements (possibly after PCA)
- (Other methods relax assumptions 2 and 3)
- Number of pure components $m$ ("chemical rank") provided by user input
- Abundances subject to non-negativity constraint

## N-FINDR Algorithm

**Heuristic:** find $m$ spectra within data set that span $(m-1)$-simplex with *largest volume*

1. Project ***X*** into $(m-1)$-dimensional space (typically by PCA)
2. Initialize simplex with $m$ arbitrary points
3. Iteratively grow simplex:
   For each vertex point in turn:
   exchange by point that maximizes simplex volume (keeping the other $m-1$ points constant)

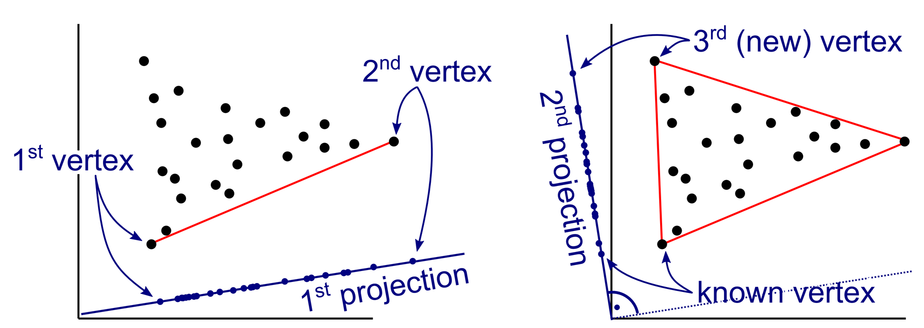
   Iterate/refine until convergence
4. Return corresponding spectra of ***X*** as endmembers
5. `predict` abundances by non-negative least squares [nnls] on found endmembers




## VCA Algorithm

**Heuristic:** projection of points onto arbitrary direction will always have 2 of the $m$ vertices as maximum and minimum.

1. Project ***X*** into $(m-1)$-dimensional space if data is considered too noisy
2. Project ***X*** onto arbitrary direction
3. Find first 2 vertices as min and max
4. Project ***X*** onto arbitrary direction orthogonal to all previously used directions
5. Find next vertex as unknown min or max
6. Repeat 4 and 5 until $m$ vertices are found
7. Return corresponding spectra of ***X*** as endmembers
8. `predict` abundances by non-negative least squares [nnls] on found endmembers




## AVIRIS Cuprite Data

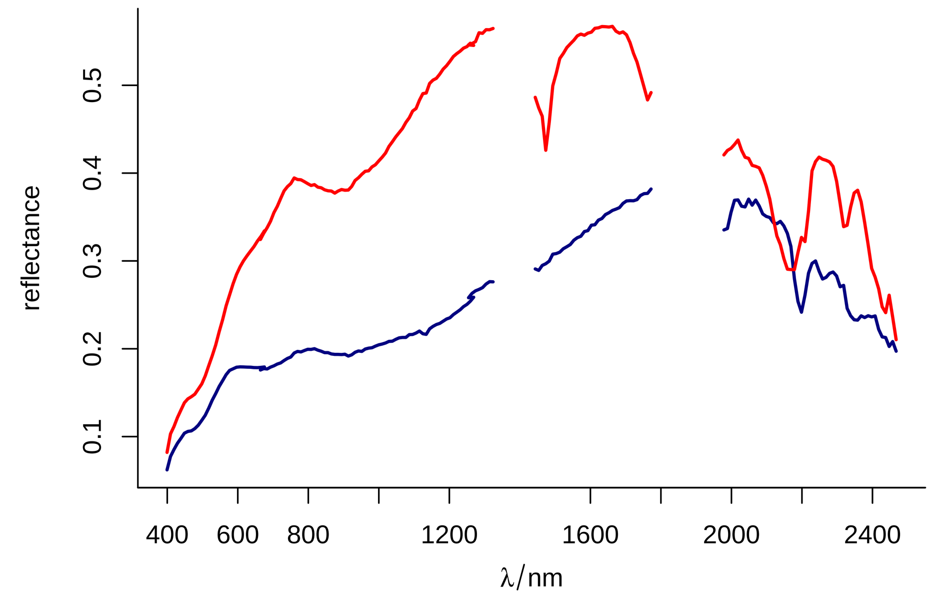
**Data Set:**

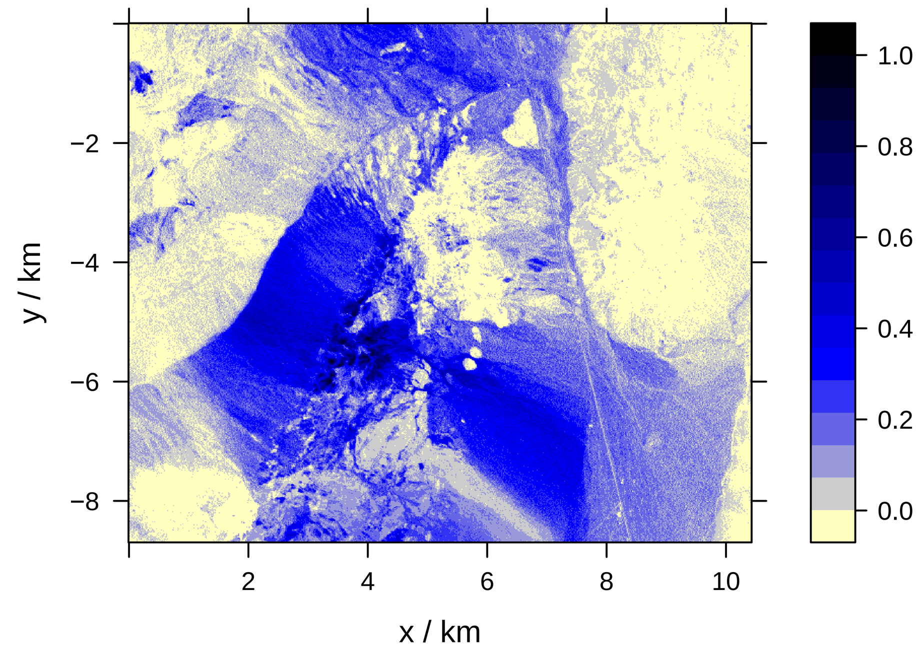
- Acquired by NASA's Airborne Visible/ InfraRed Imaging Spectrometer
- of mining region in the south of Nevada/USA
- 45 × 10 km (300 000 pixel subimage shown)
- 250 - 4 000 nm (224 spectral bands)
- Well-known ground truth

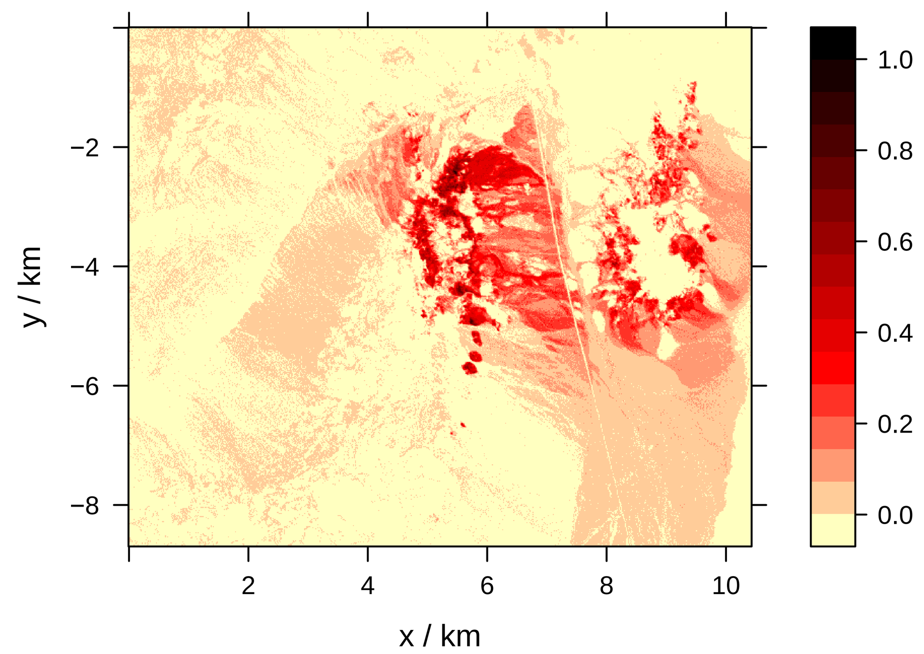
### N-FINDR with $m = 19$ endmembers

As example, we show 2 components identified as

- muscovite (mica, $KAl_2(AlSi_3O_{10})(FOH)_2$), and
- alunite (alumstone, $KAl_3(SO_4)_2(OH)_6$).



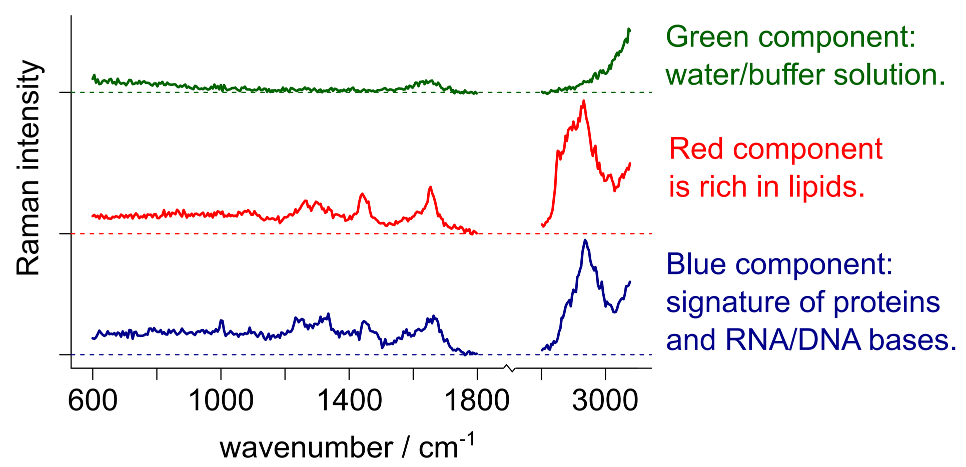



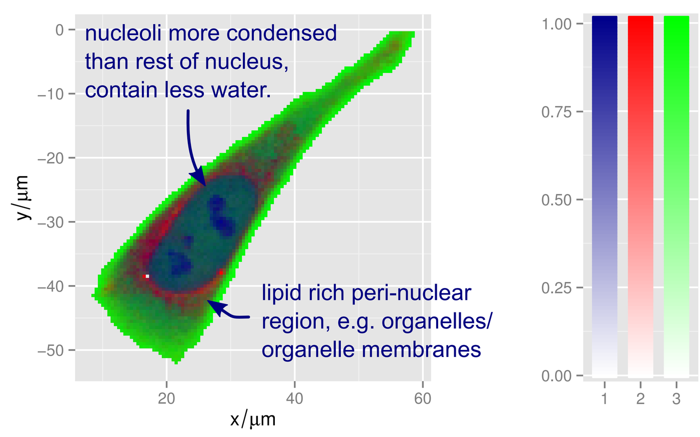

## Raman Image of HeLa Cell

**Data Set:**

- Raman spectra of HeLa cell
- Excitation: 5 mW @ 488 nm, 0.5 s/spectrum
- Spectra: 600 – 1800 + 2800 – 3075 $cm^{-1}$, 314 bands (after pre-processing)
- Area: 60 × 60 µm, step size 0.5 µm
- For details see reference [HeLa Cell].

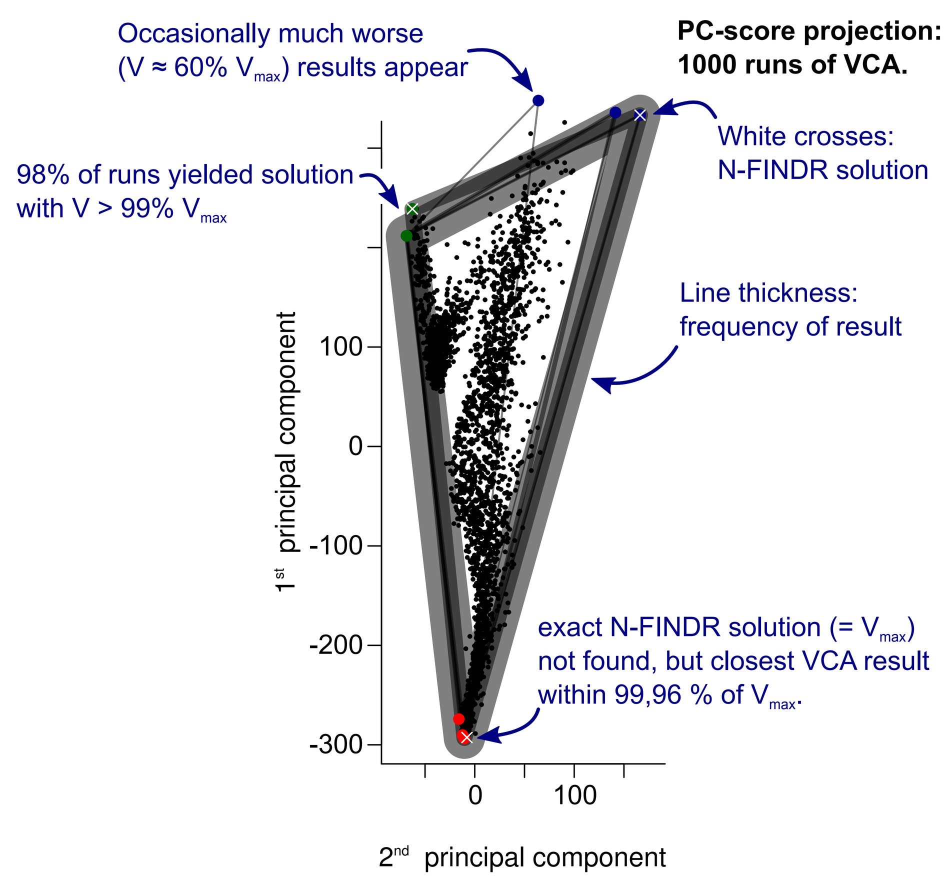
### N-FINDR with $m = 3$ endmembers




- Solution is stable: Identical results for 100 runs with random initialization

### VCA Results $m = 3$ endmembers




- VCA is expected to be less stable than N-FINDR: no refinement of tentative vertices
- VCA faster than Winter's N-FINDR, but advantage small for improved algorithms.

## R package `unmixR`

Conor McManus implemented N-FINDR [Winter, Dowler] and VCA [Nascimento, Lopez] algorithms as R package `unmixR`. He was supervised by Claudia Beleites, Simon Fuller and Bryan Hanson.

Claudia Beleites now maintains the package with help by Bryan Hanson.

The package is available at
`http://github.com/Chathurga/unmixR`

## References

**Winter** ME *et al.*: N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data, Proc SPIE, 3753, 266-275 (1999). DOI: 10.1117/12.366289

**Dowler** SW *et al.*: Reducing the complexity of the N-FINDR algorithm for hyperspectral image analysis., IEEE Trans Image Process, 22, 2835–2848 (2013). DOI: 10.1109/TIP.2012.2219546

**Nascimento** JMP *et al.*: Vertex Component Analysis: A Fast Algorithm to Unmix Hyperspectral Data, IEEE Trans Geosci. Rem Sens, 43, 898–910 (2005). DOI: 10.1109/TGRS.2005.844293 Iden

**Lopez** S *et al.*: A Low-Computational-Complexity Algorithm for Hyperspectral Endmember Extraction: Modified Vertex Component Analysis, IEEE Geosci Rem Sens Lett, 9, 502-506 (2012). DOI: 10.1109/LGRS.2011.2172771

**Cuprite:** AVIRIS data: `http://aviris.jpl.nasa.gov/data/free_data.html`, spetroscopic and geological information: `http://speclab.cr.usgs.gov/PAPERS/cuprite.clark.93/mineral_map.html`, spetroscopic reference: Clark RN *et al.*: USGS digital spectral library splib06a: U.S. Geological Survey, Digital Data Series 231 (2007).

**HeLa Cell:** Hedegaard M *et al.*: Spectral unmixing and clustering algorithms for assessment of single cells by Raman microscopic imaging, Theor Chem Acc, 130, 1249–1261 (2011). DOI: 10.1007/s00214-011-0957-1

**hyperSpec:** a package to handle hyperspectral data sets in R, Beleites C & Sergo V, Ver. 0.98-20140612 (2014).

**nnls:** The Lawson-Hanson algorithm for non-negative least squares, Mullen KM & van Stokkum IHM, Ver. 1.4 (2012).

**lattice:** Multivariate Data Visualization with R, Sarkar D, Springer (2008). Ver. 0.20-29

**ggplot2:** Elegant Graphics for Data Analysis, Wickham H, Springer (2009). Ver. 1.0.0

## Acknowledgments

Conor McManus was supported by the Google Summer of Code 2013 to implement the algorithms.
Claudia Beleites thanks the BMBF for funding via the project "RamanCTC" (13N12685).
We thank Christian Matthäus for providing us with the HeLa cell data set.