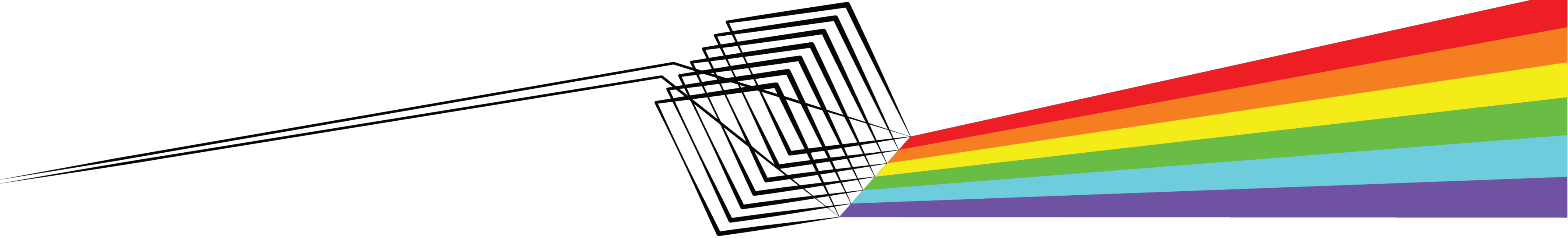


Subpixel detection of peanut in wheat flour using near infrared hyperspectral imaging

Antoine Laborde[1], Benoît Jaillais[2], Delphine Jouan-Rimbaud Bouveresse[3], Anthony Boulanger[1], Christophe B.Y. Cordella[3]
[1]GreenTropism, 75116 Paris, France.
[2]StatSC INRA/ONIRIS, 44322 Nantes, France.
[3]UMR914 PNCA INRA/AgroParisTech, 75231 Paris, France.

Chimiométrie XIX - 2019

30 Janvier - 1er Février 2019 - Montpellier SupAgro

Peanut in wheat flour

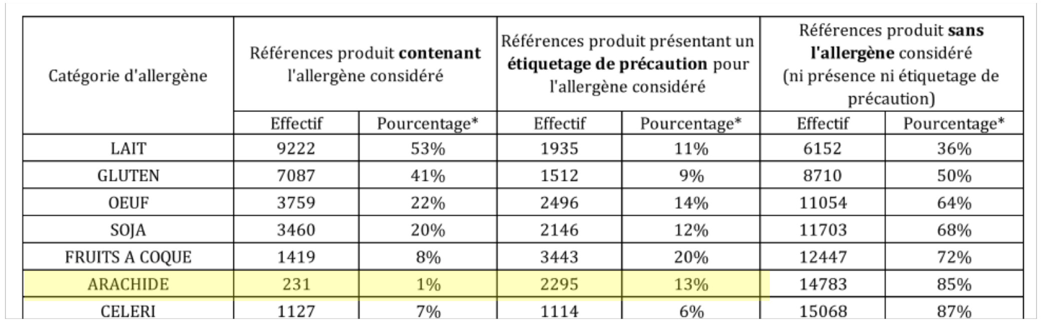
- **Peanut allergy** is a **major health concern**:
    - High sensitivity for very small dose (a few mg)
    - Allergic population on the rise ?
    - Widely used in the food industry

- **Advisory labelling** for food allergen are **not a complete solution** for allergic people

- Allergen contamination could be detected using **imaging technique** in raw material as wheat flour

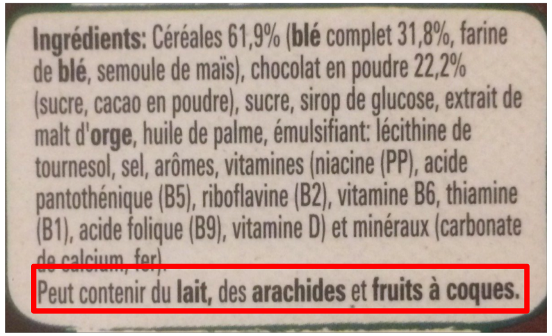
References
W. Burks, “Peanut allergy,” *Lancet*, vol. 371, no. 9623, pp. 1538–1546, 2008.
E. W. Lusas, “Food uses of peanut protein,” *J. Am. Oil Chem. Soc.*, vol. 56, no. 3, pp. 425–430, 1979.
C. Hadley, “Food allergies on the rise?,” *EMBO Rep.*, vol. 7, no. 11, pp. 1080–1083, 2006.

From Oqali, A study on allergens occurance in available transformed products on the French marker between 2008 and 2012. (“Etude des allergènes dans les produits transformés disponibles sur le marché français entre 2008 et 2012,” Paris, 2015.)



Catégorie d'allergène	Références produit **contenant** l'allergène considéré		Références produit présentant un **étiquetage de précaution** pour l'allergène considéré		Références produit **sans** l'allergène considéré (ni présence ni étiquetage de précaution)	
	Effectif	Pourcentage*	Effectif	Pourcentage*	Effectif	Pourcentage*
LAIT	9222	53%	1935	11%	6152	36%
GLUTEN	7087	41%	1512	9%	8710	50%
OEUF	3759	22%	2496	14%	11054	64%
SOJA	3460	20%	2146	12%	11703	68%
FRUITS A COQUE	1419	8%	3443	20%	12447	72%
ARACHIDE	231	1%	2295	13%	14783	85%
CELERI	1127	7%	1114	6%	15068	87%


[redacted]
[redacted]

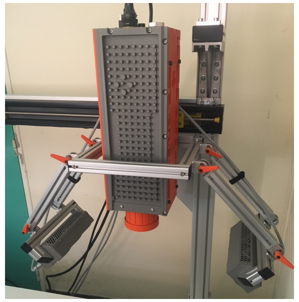


Ingrédients: Céréales 61,9% (**blé** complet 31,8%, farine de **blé**, semoule de maïs), chocolat en poudre 22,2% (sucre, cacao en poudre), sucre, sirop de glucose, extrait de malt d'**orge**, huile de palme, émulsifiant: lécithine de tournesol, sel, arômes, vitamines (niacine (PP), acide pantothénique (B5), riboflavine (B2), vitamine B6, thiamine (B1), acide folique (B9), vitamine D) et minéraux (carbonate de calcium, fer).
Peut contenir du **lait**, des **arachides** et **fruits à coques.**

*May contain milk, peanut and nuts.

Peanut and wheat flour in near infrared

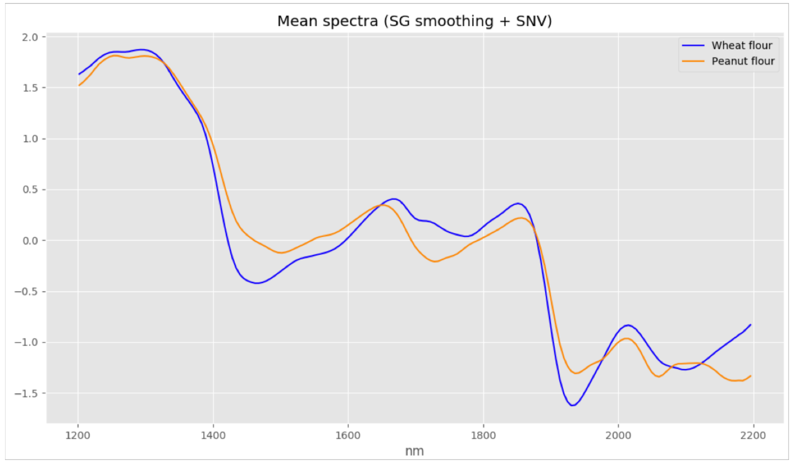
- Peanut allergy is associated with **proteins** (Ara h 1 -> Ara h 8*)
- Peanut **NIR spectral signature is not specific to proteins** (influence of fatty acids)
- **Peanut flour** is defatted and show less NIR pattern → more difficult to distinguish from wheat
- Peanut flour and wheat flour pure spectra **do not show clear specific spectral patterns** -> challenging unmixing problem



Specim SWIR

Defatted peanut flour

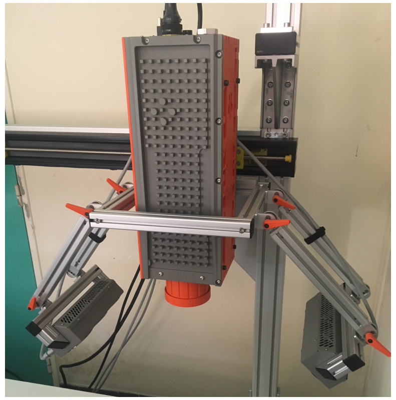
White wheat flour

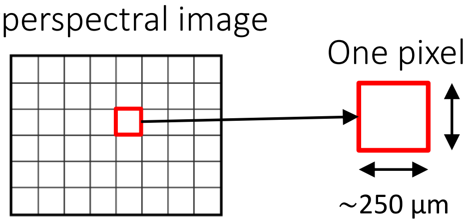


---

* Proteins from *Arachis hypogaea* (peanuts)

Subpixel issue





The camera provides the average spectrum of the pixel field of view (~250 µm square).

Particle size < 200 µm

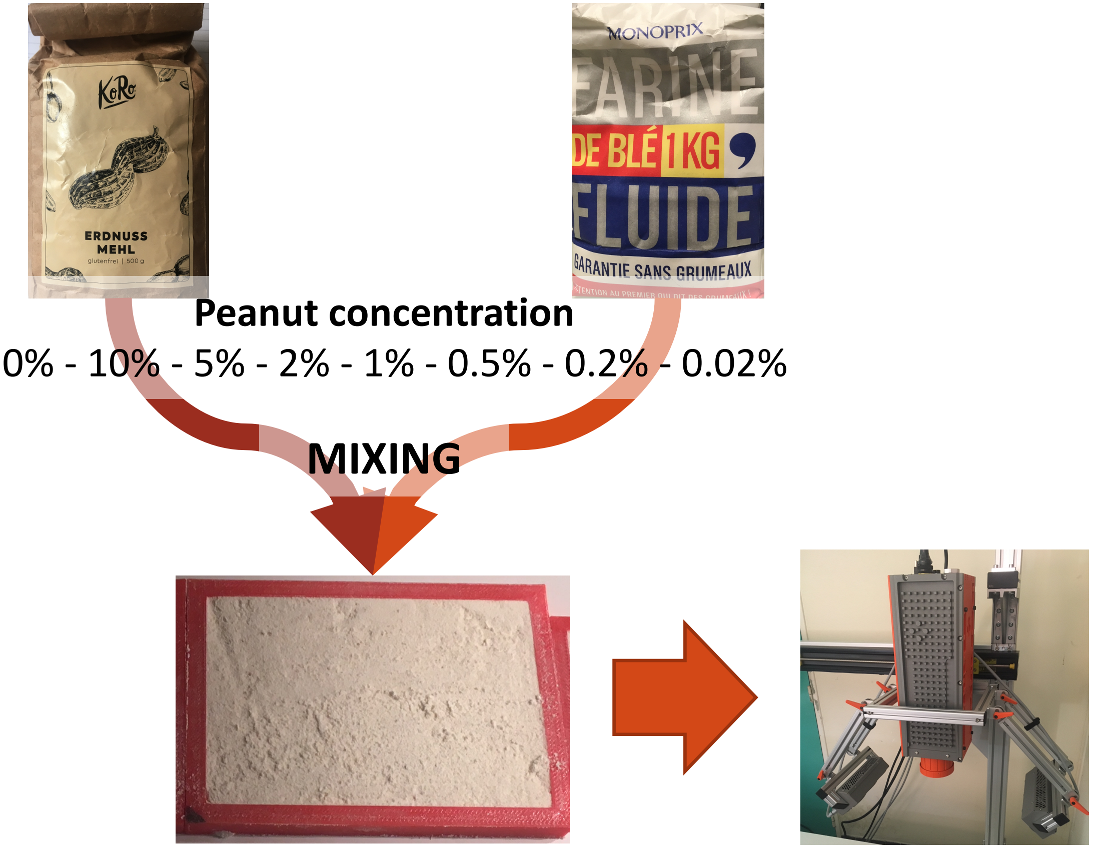
One pixel contains several particles

- Each pixel spectrum is a **mix of peanut and wheat**
- Spatial pattern cannot be used as the mixture is assumed to be homogeneous
- No spectral data with different concentrations of peanut available
- No reference value for individual pixel

Experimental setup

« Target »

« Background »



Peanut concentration

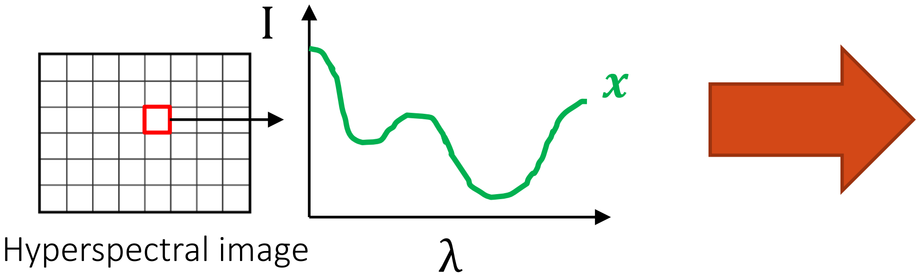
20% - 10% - 5% - 2% - 1% - 0.5% - 0.2% - 0.02%

MIXING

Specim SWIR

- Flour mixtures: from 20% down to 0.02% (200 ppm) of peanut in wheat
- Mass measurement with 0.01 mg precision
- Manually blended flour mixtures
- 3 replica for each sample andpure samples are measured with the camera
- Samples considered to be homogeneous

Target correlation approach



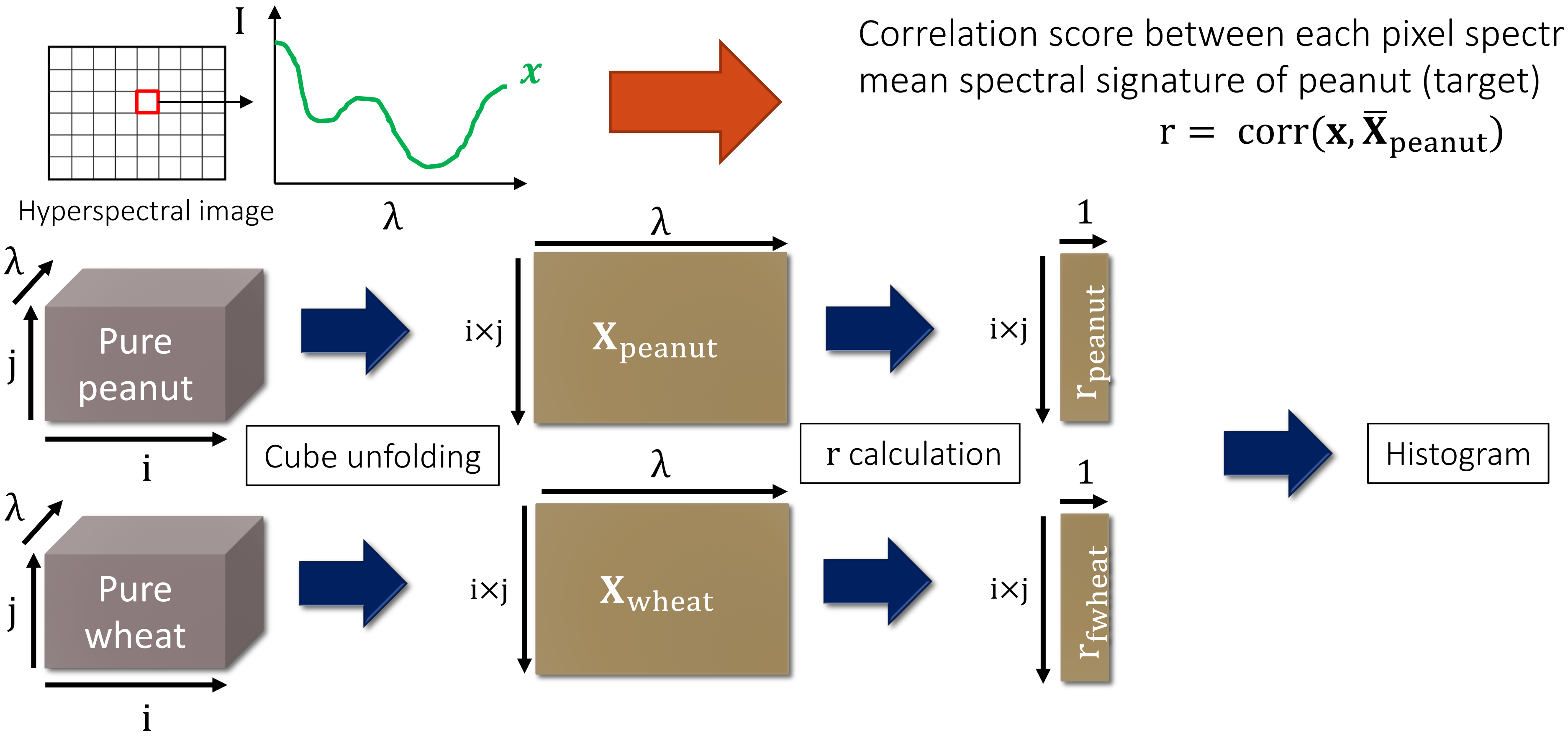
Correlation score between each pixel spectrum and the mean spectral signature of peanut (target)

$$r = \operatorname{corr}(\mathbf{x}, \bar{\mathbf{X}}_{\text{peanut}})$$

Target correlation approach

Correlation score between each pixel spectrum and the mean spectral signature of peanut (target)

$$r = \mathrm{corr}(\mathbf{x}, \bar{\mathbf{X}}_{\mathrm{peanut}})$$

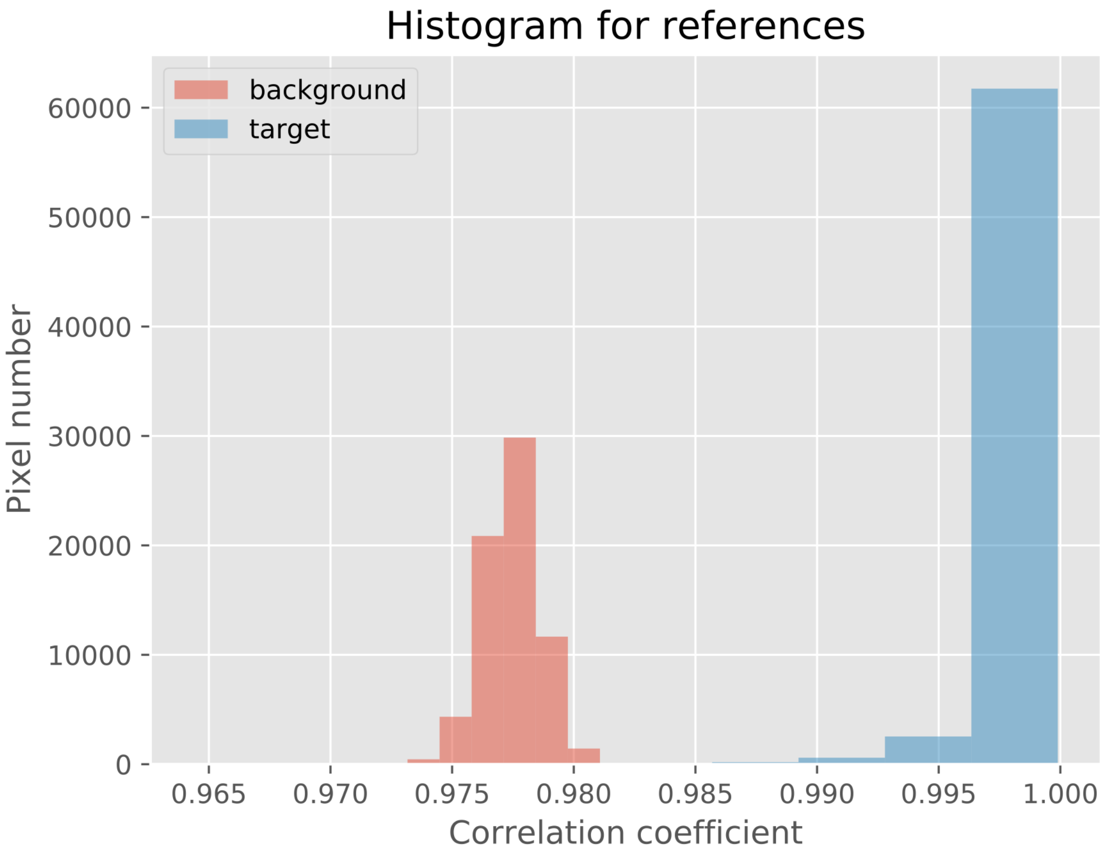


Cube unfolding

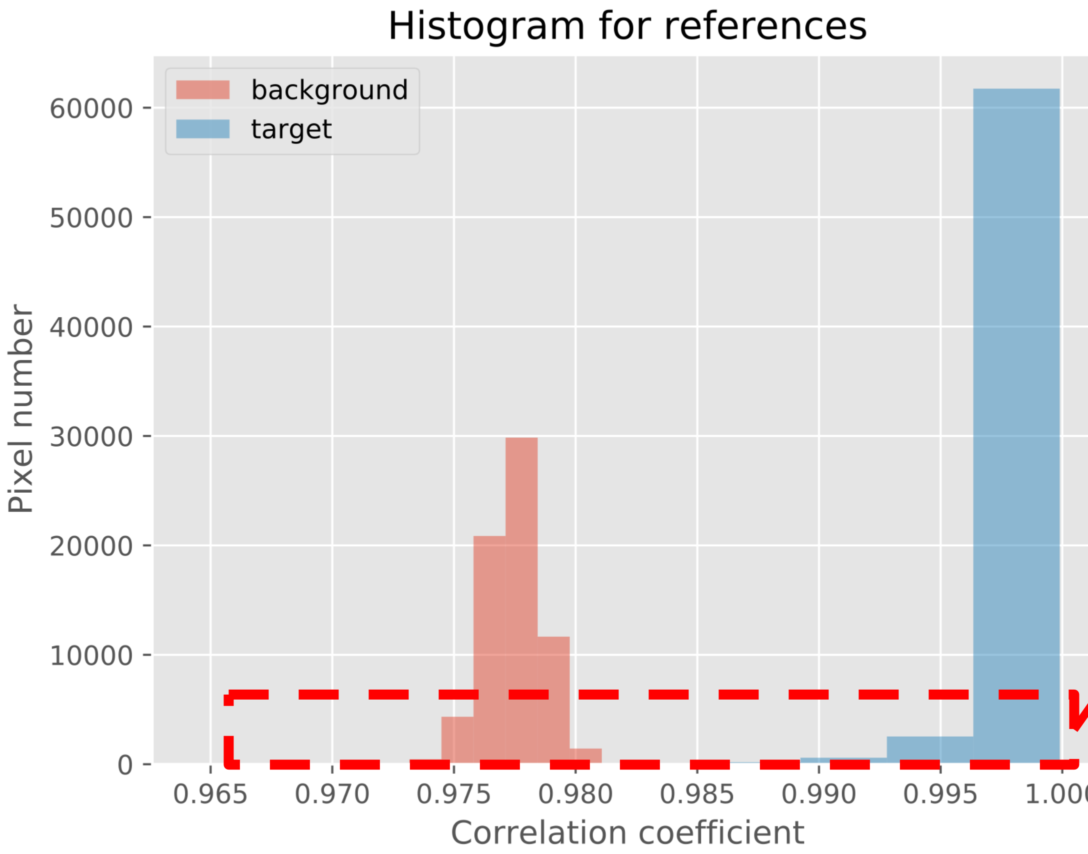
r calculation

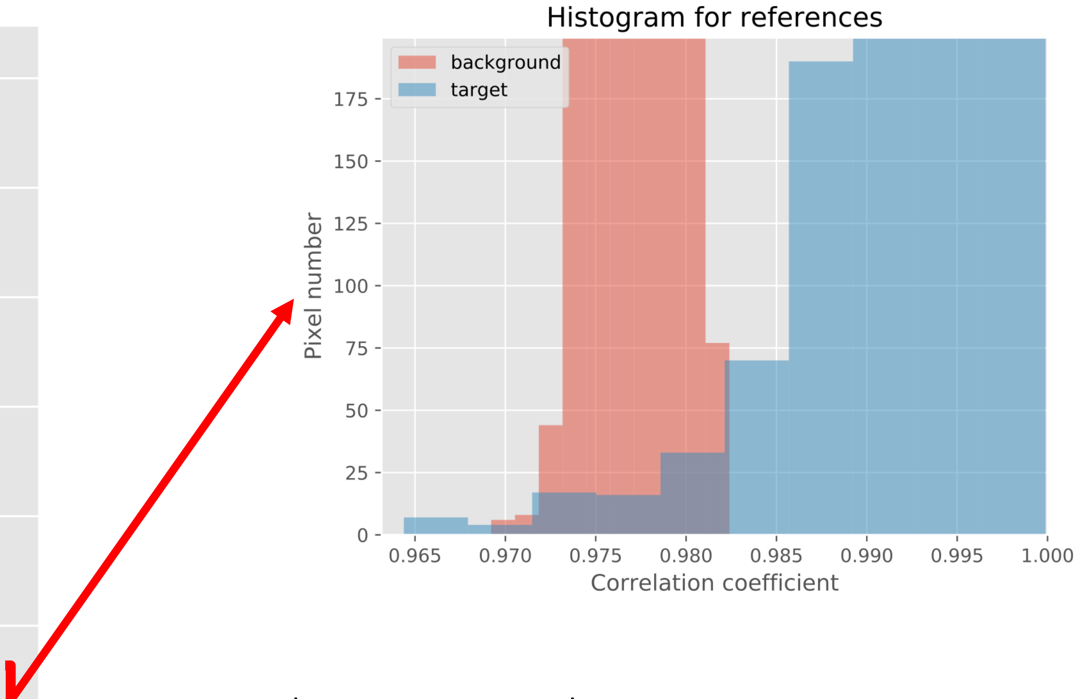
Histogram

Target correlation approach



Target correlation approach





- Correlation approach not appropriate
- Subspaces needed to model spectral variability

Adaptive Matched Subspace Detector (AMSD)

- AMSD is an **algorithm** used for **detection**: the output is **detected** (target) or **not detected** (background)

D. Manolakis, C. Siracusa, and G. Shaw, “Hyperspectral subpixel target detection using the linear mixing model,” IEEE Trans. Geosci. Remote Sens., vol. 39, no. 7, pp. 1392–1409, 2001.

Adaptive Matched Subspace Detector (AMSD)

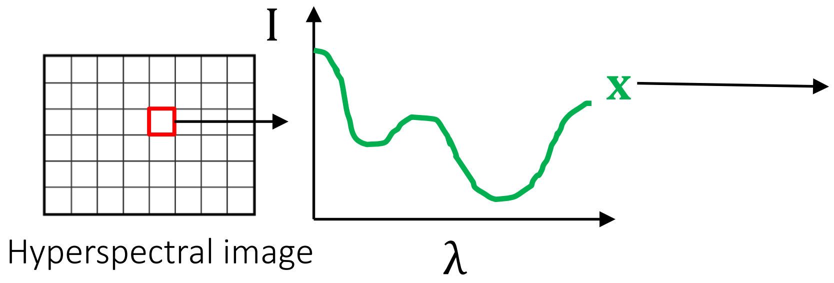
- AMSD is an **algorithm** used for **detection**: the output is **detected** (target) or **not detected** (background)

- AMSD is used to assess the presence of peanut in each pixel

D. Manolakis, C. Siracusa, and G. Shaw, “Hyperspectral subpixel target detection using the linear mixing model,” IEEE Trans. Geosci. Remote Sens., vol. 39, no. 7, pp. 1392–1409, 2001.

Adaptive Matched Subspace Detector (AMSD)

- AMSD is an **algorithm** used for **detection**: the output is **detected** (target) or **not detected** (background)
- AMSD is used to assess the presence of peanut in each pixel



$$x = \sum_{k=1}^{M} a_k s_k + r$$

Each pixel is assumed to be a **linear combination** of endmembers (s_k) according to some proportions (a_k). r holds for the residuals of the model.

D. Manolakis, C. Siracusa, and G. Shaw, “Hyperspectral subpixel target detection using the linear mixing model,” IEEE Trans. Geosci. Remote Sens., vol. 39, no. 7, pp. 1392–1409, 2001.

Adaptive Matched Subspace Detector (AMSD)

- AMSD approach is performed in 3 steps:

  1. Construction of the **target** (S_t) and the **background** (S_b) **subspaces**

  2. Design of the detector:
     - n stands for the target and background **subspace dimensionality**
     - θ stands for the **detection threshold**

  3. **Application** on test images

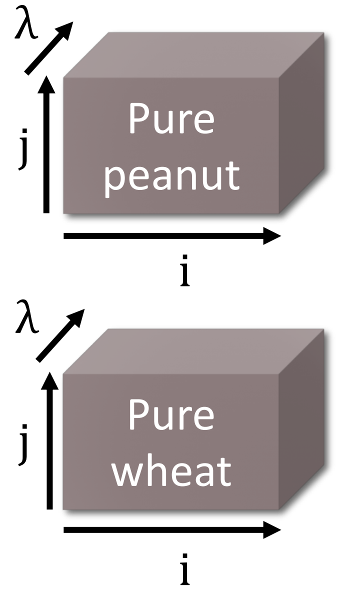
Adaptive Matched Subspace Detector (AMSD)

- AMSD approach is performed in 3 steps:

  1. Construction of the **target** (S_b) and the **background** (S_t) **subspaces**

Subspace design

How to obtain the subspaces $\mathrm{S_b}$ and $\mathrm{S_t}$?

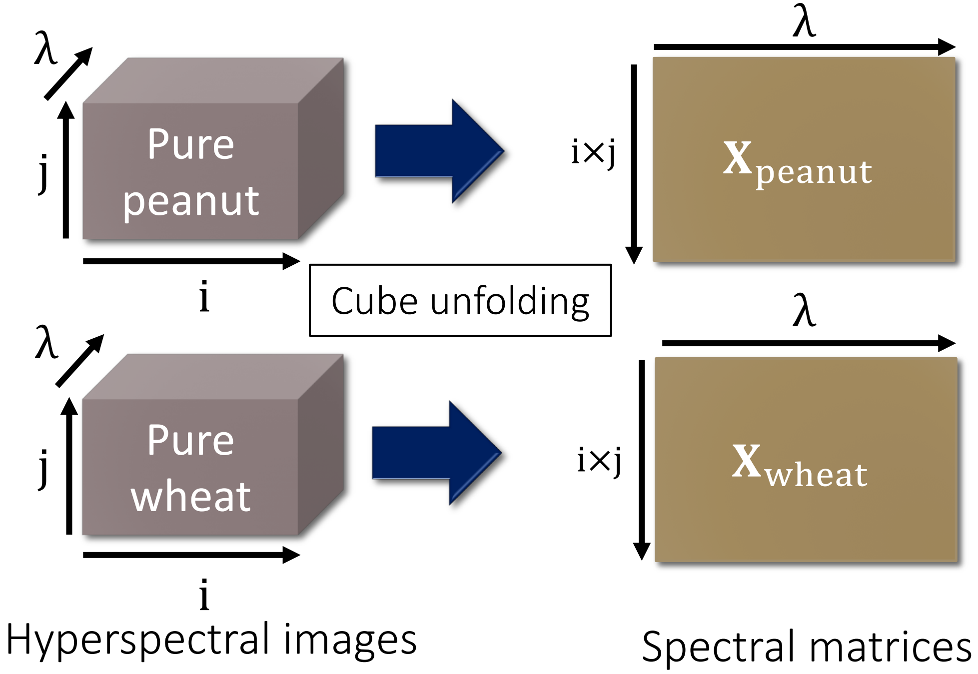


Hyperspectral images

- Pure peanut flour and pure wheat flour images are acquired to measure the **sample variability**
- $\mathrm{S_b}$ models the variability of **wheat flour**
- $\mathrm{S_t}$ models the variability of **peanut flour**

Subspace design

How to obtain the subspaces S_b and S_t ?



Pure peanut

Pure wheat

Cube unfolding

X_{peanut}

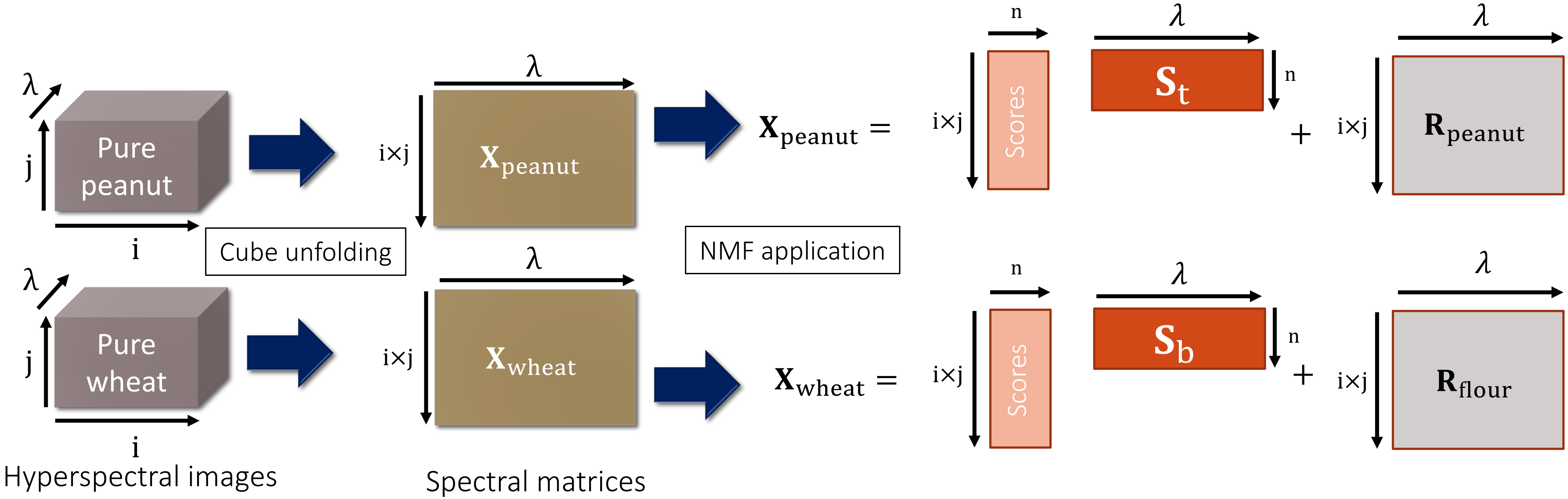
X_{wheat}

Hyperspectral images

Spectral matrices

Subspace design

How to obtain the subspaces S_b and S_t ?



Adaptive Matched Subspace Detector (AMSD)

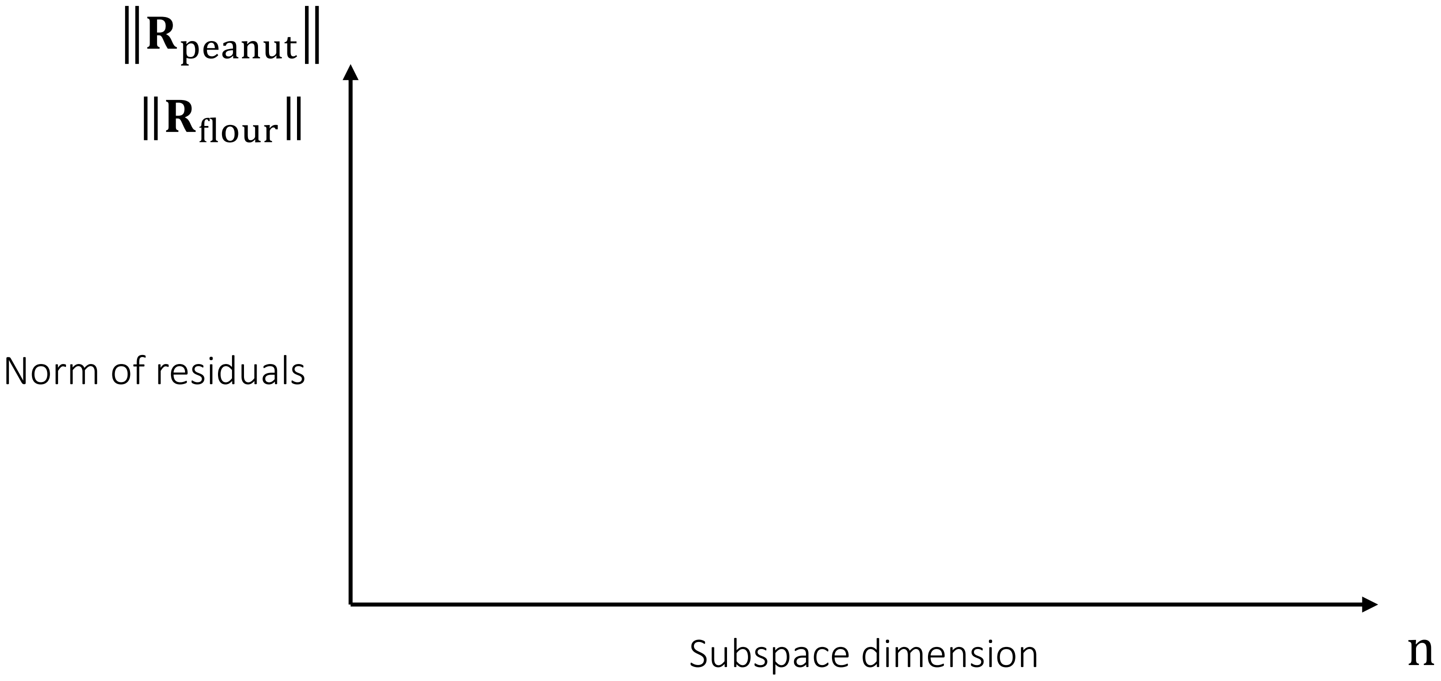
- AMSD approach is performed in 3 steps:

1. Construction of the **target** **(S_b)** and the **background** **(S_t)** **subspaces**

2. Design of the detector:
    - n stands for the target and background **subspace dimensionality**
    - θ stands for the **detection threshold**

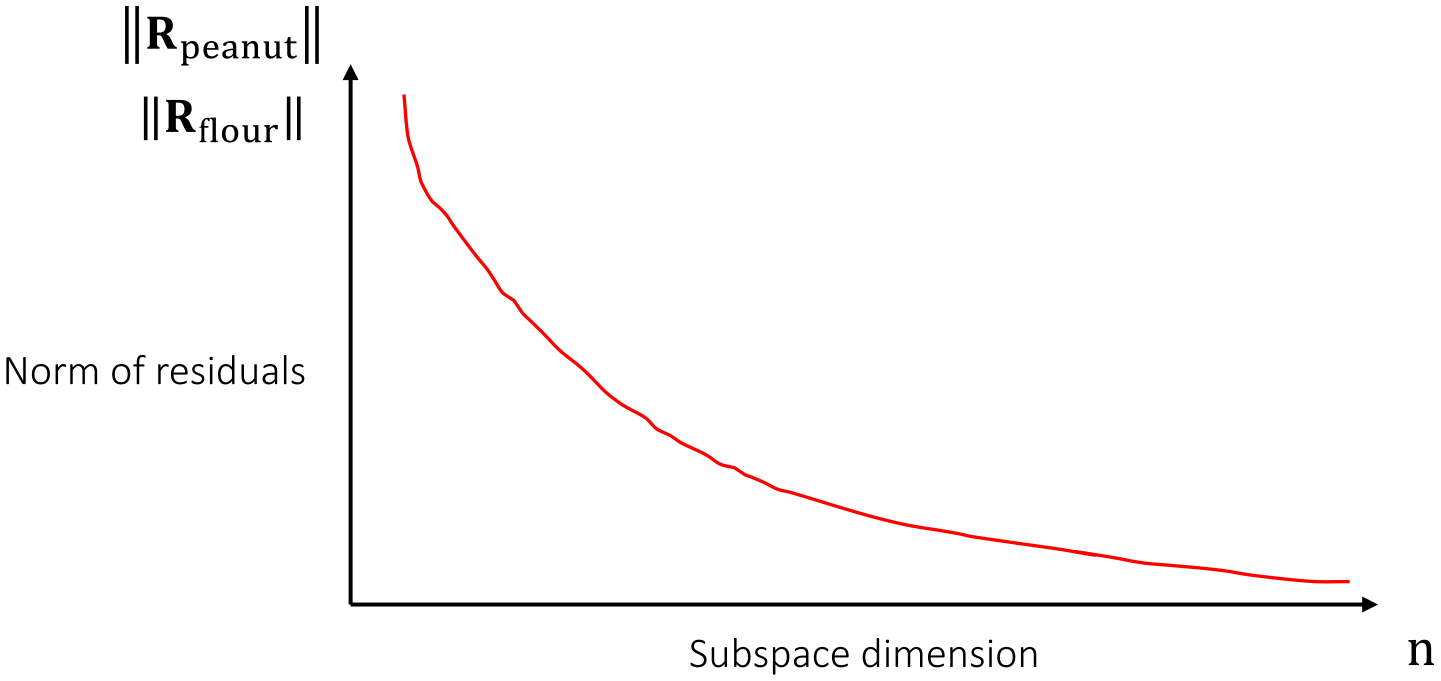
Subspace design

How to choose the subspace dimension $\mathbf{n}$?



Subspace design

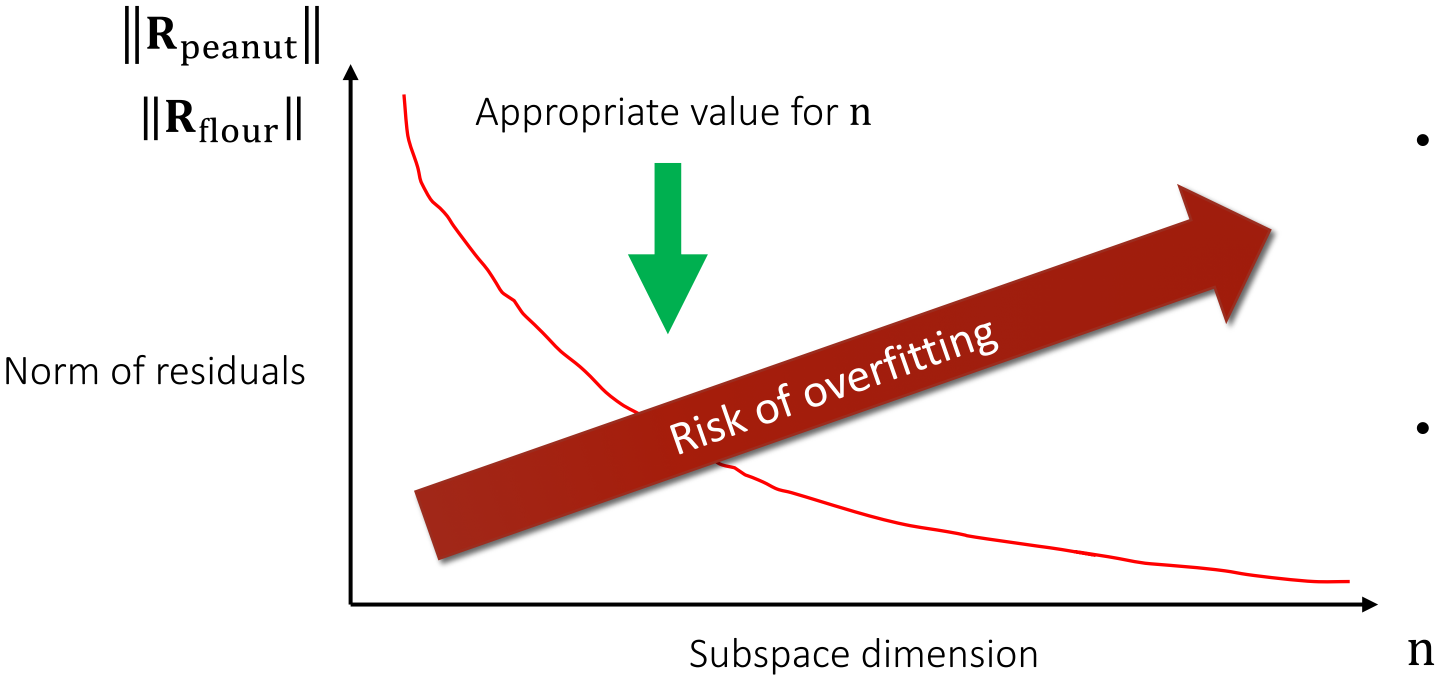
How to choose the subspace dimension $\mathbf{n}$?



- With high $\mathbf{n}$, the variability of pure samples is taken into account

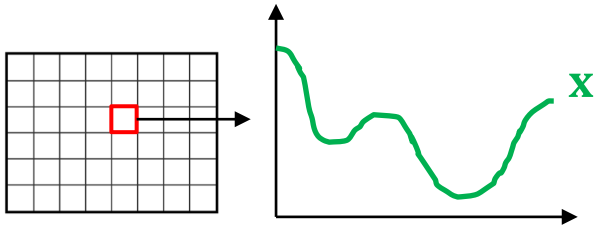
Subspace design

How to choose the subspace dimension $\mathbf{n}$?



- With high $\mathbf{n}$, the variability of pure samples is taken into account
- High $\mathbf{n}$ may lead to detector **overfitting**: the measurement noise is confused with sample variability
- A tradeoff must be found by checking the detector histograms and detection map

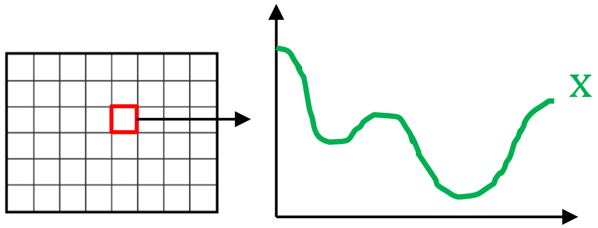
Apply AMSD ratio



Hyperspectral images

Decompose $\mathbf{x}$ on $\mathbf{S}_\mathrm{b}$ using Non Negative Least Square (NNLS) to find abundance vector

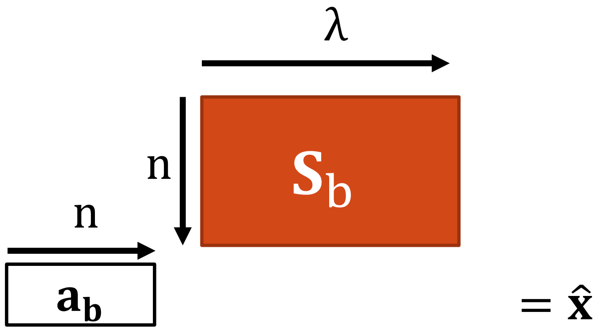
Apply AMSD ratio



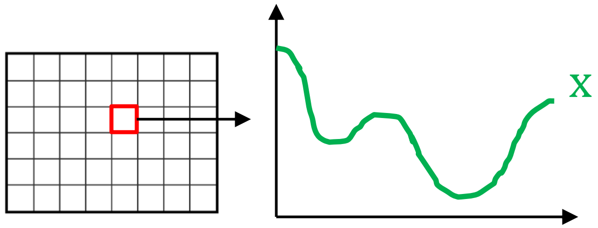
Hyperspectral images

Decompose $\mathbf{x}$ on $\mathbf{S_b}$ using Non Negative Least Square (NNLS) to find abundance vector

Reconstruct an estimation of x using $\mathbf{S_b}$ and $\mathbf{a_b}$



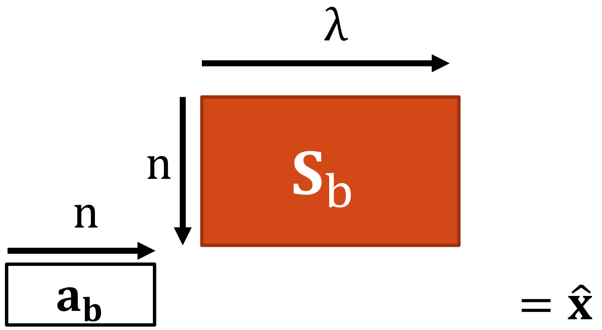
Apply AMSD ratio



Hyperspectral images

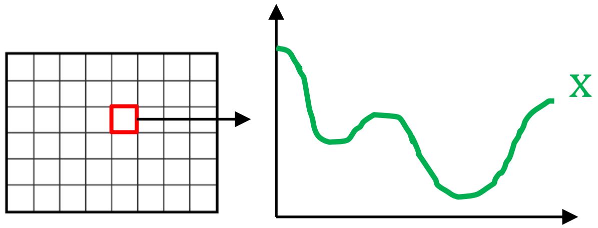
Decompose $\mathbf{x}$ on $\mathbf{S_b}$ using Non Negative Least Square (NNLS) to find abundance vector

Reconstruct an estimation of $\mathbf{x}$ using $\mathbf{S_b}$ and $\mathbf{a_b}$



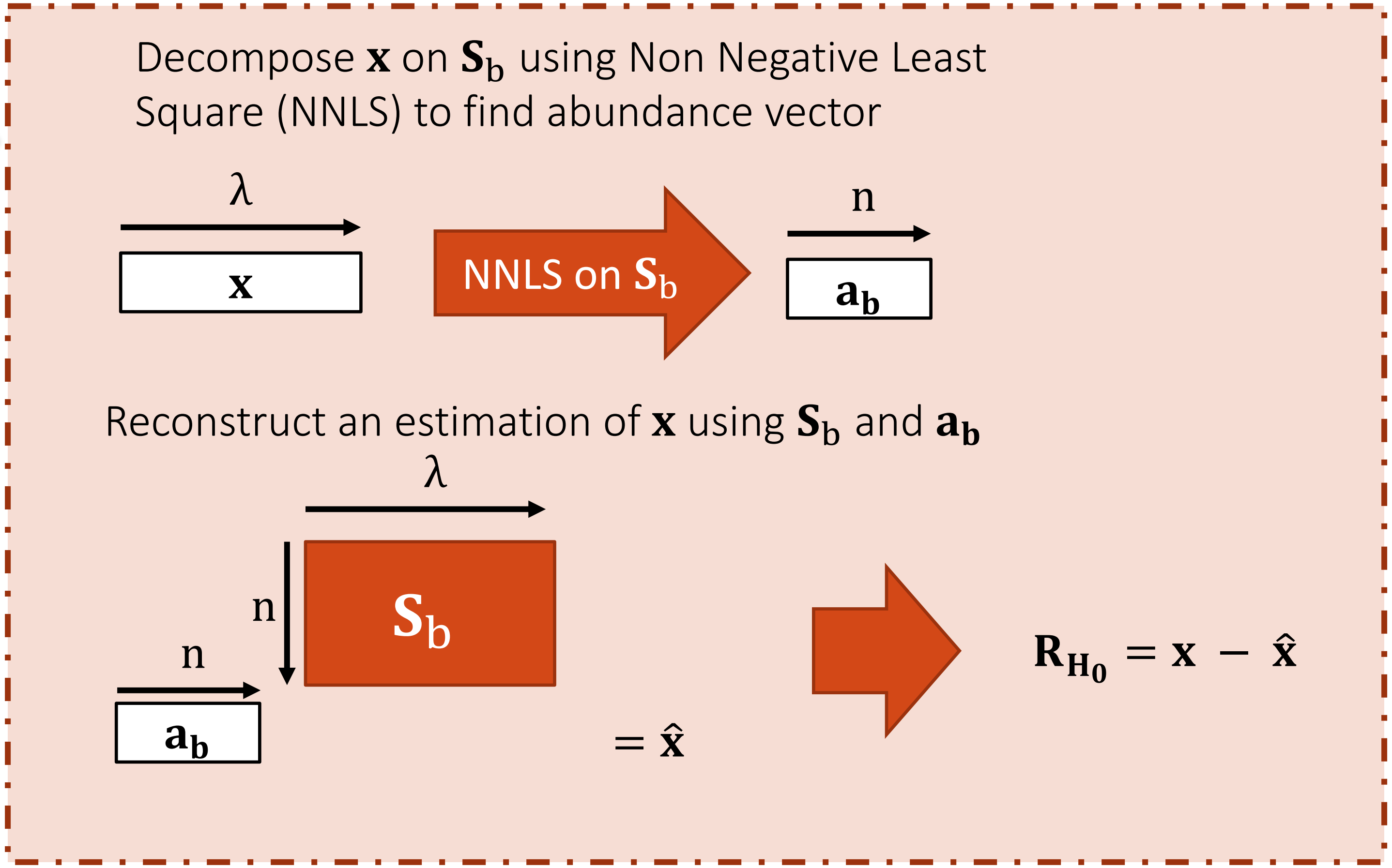
$$\mathbf{R_{H_0}} = \mathbf{x} - \hat{\mathbf{x}}$$

Apply AMSD ratio



Hyperspectral images

H_0 : target is **absent**



Decompose $\mathbf{x}$ on $\mathbf{S_b}$ using Non Negative Least Square (NNLS) to find abundance vector

λ

$\mathbf{x}$

NNLS on $\mathbf{S}_\mathrm{b}$

n

$\mathbf{a_b}$

Reconstruct an estimation of $\mathbf{x}$ using $\mathbf{S_b}$ and $\mathbf{a_b}$

λ

n

n

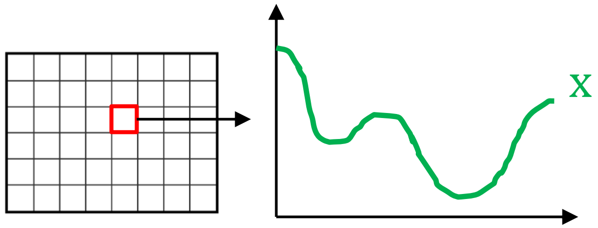
$\mathbf{a_b}$

$\mathbf{S}_\mathrm{b}$

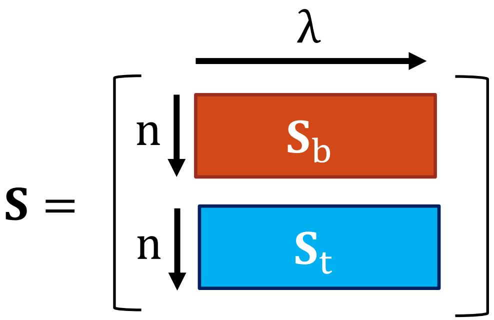
$= \hat{\mathbf{x}}$

$$\mathbf{R}_{\mathbf{H_0}} = \mathbf{x} - \hat{\mathbf{x}}$$

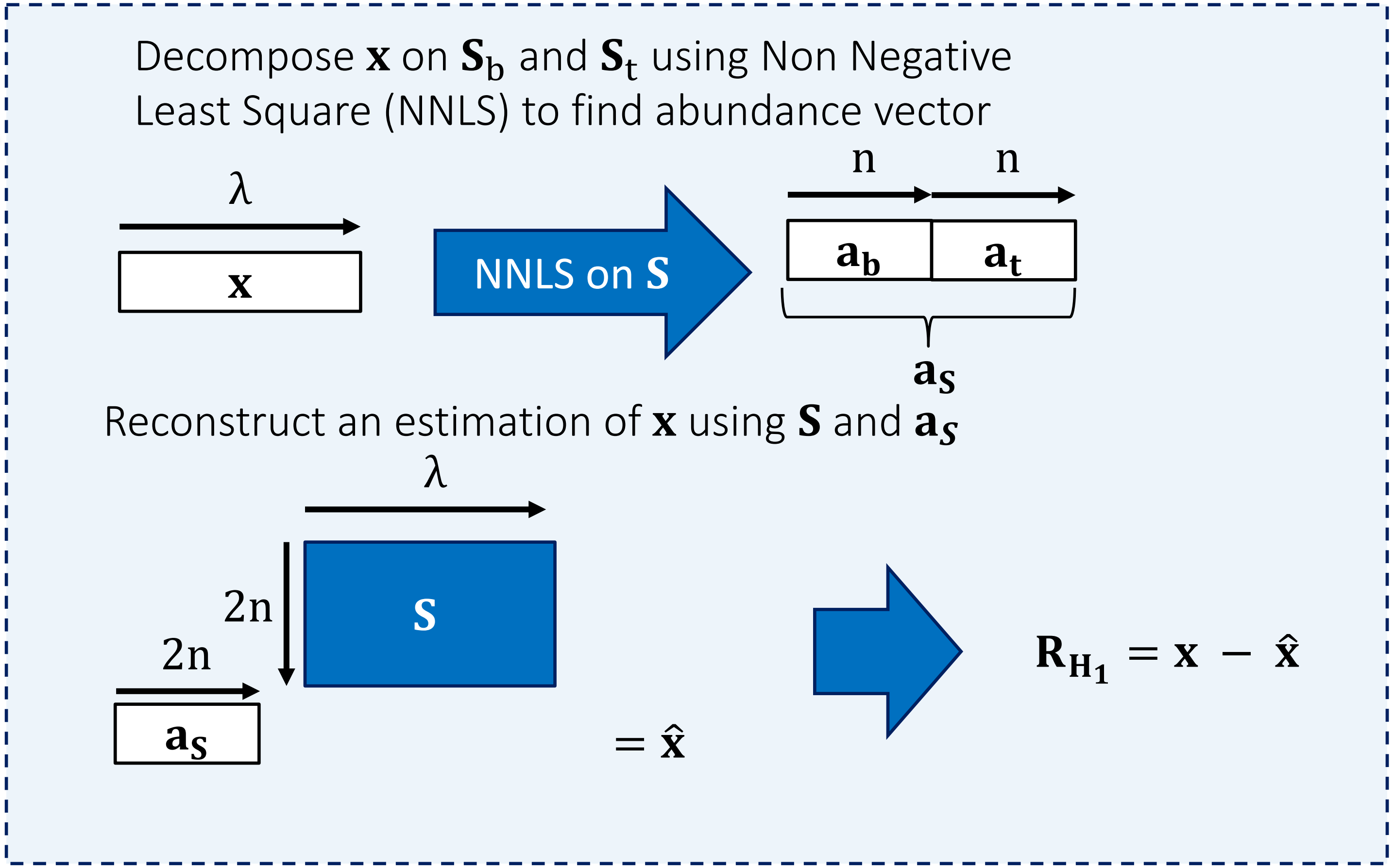
Apply AMSD ratio



Hyperspectral images



$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_b \\ \mathbf{S}_t \end{bmatrix}$$



Decompose $\mathbf{x}$ on $\mathbf{S}_b$ and $\mathbf{S}_t$ using Non Negative Least Square (NNLS) to find abundance vector

Reconstruct an estimation of $\mathbf{x}$ using $\mathbf{S}$ and $\mathbf{a}_S$

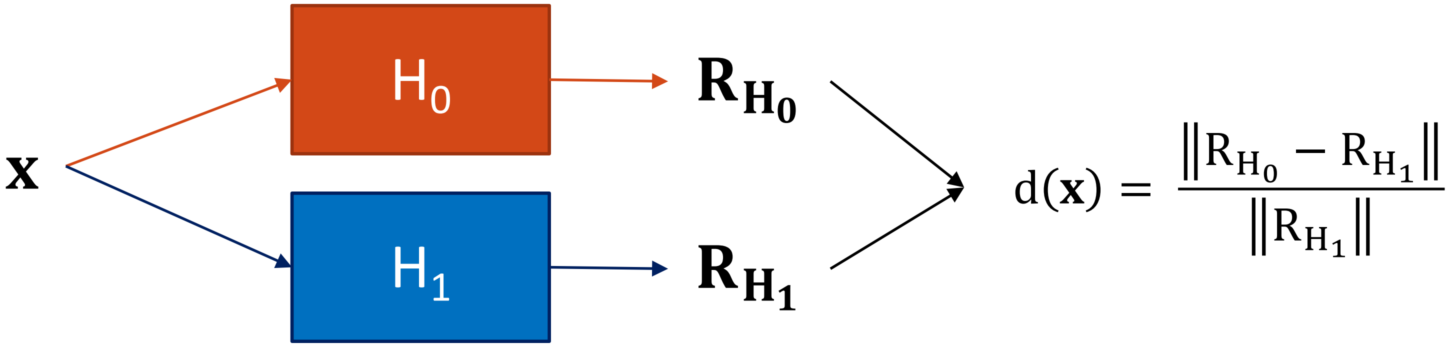
$= \hat{\mathbf{x}}$

$$\mathbf{R}_{\mathbf{H}_1} = \mathbf{x} - \hat{\mathbf{x}}$$

H_1 : target is **present**

Apply AMSD ratio

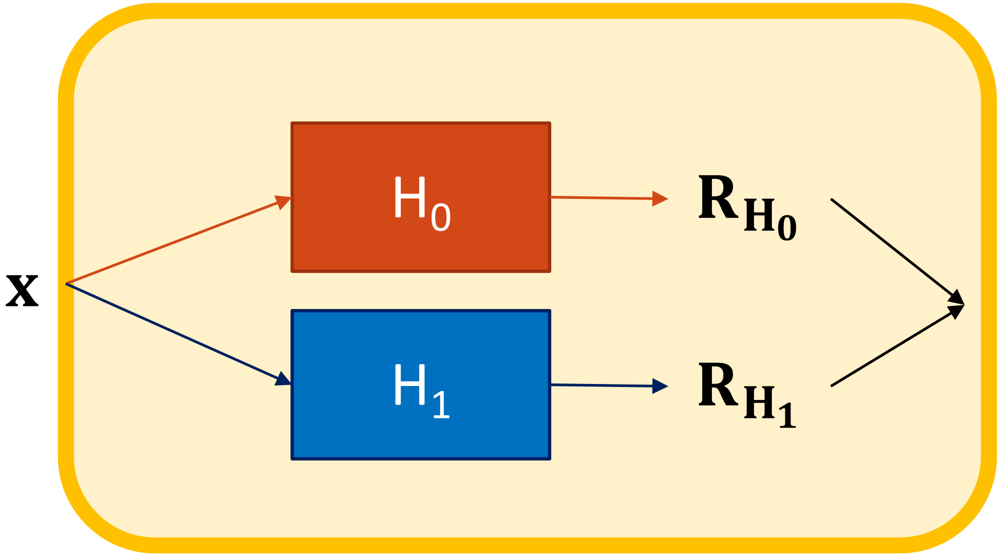
- The AMSD gives a metric for the detection which is a relative comparison of residuals between both models



$$d(\mathbf{x}) = \frac{\|R_{H_0} - R_{H_1}\|}{\|R_{H_1}\|}$$

Apply AMSD ratio

- The AMSD gives a metric for the detection which is a relative comparison of residuals between both models

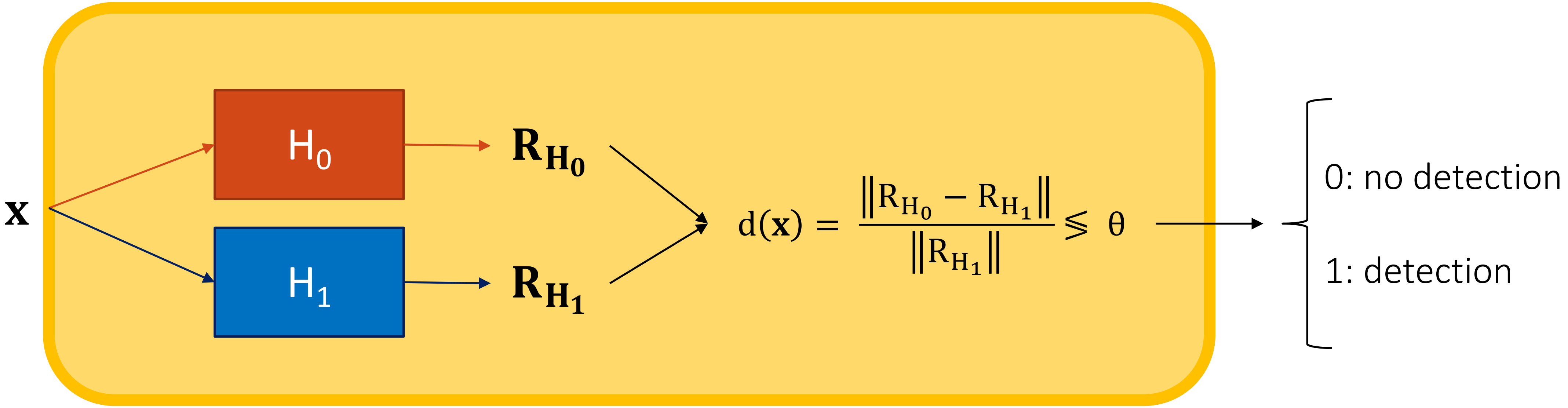


$$d(\mathbf{x}) = \frac{\left\| R_{H_0} - R_{H_1} \right\|}{\left\| R_{H_1} \right\|}$$

AMSD algorithm (first part)

Apply AMSD ratio

- The AMSD gives a metric for the detection which is a relative comparison of residuals between both models



$\mathbf{x}$

H_0

H_1

$\mathbf{R}_{\mathbf{H_0}}$

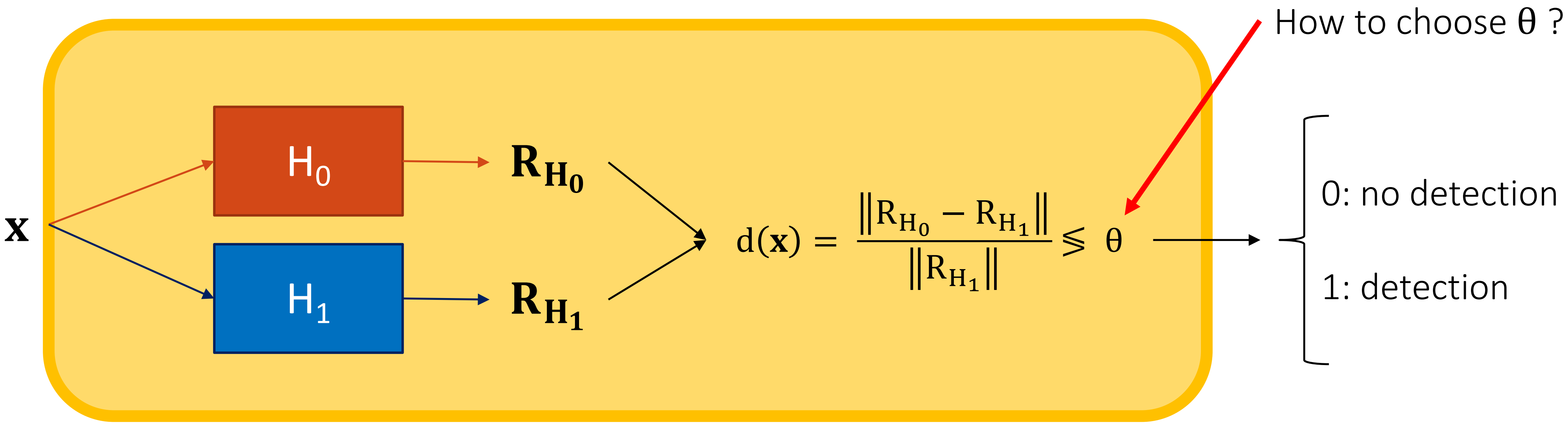
$\mathbf{R}_{\mathbf{H_1}}$

$$d(\mathbf{x}) = \frac{\|R_{H_0} - R_{H_1}\|}{\|R_{H_1}\|} \lessgtr \theta \longrightarrow \begin{cases} 0\text{: no detection} \\ 1\text{: detection} \end{cases}$$

AMSD algorithm (complete)

Apply AMSD ratio

- The AMSD gives a metric for the detection which is a relative comparison of residuals between both models



$$d(\mathbf{x}) = \frac{\|R_{H_0} - R_{H_1}\|}{\|R_{H_1}\|} \lessgtr \theta$$

How to choose θ ?

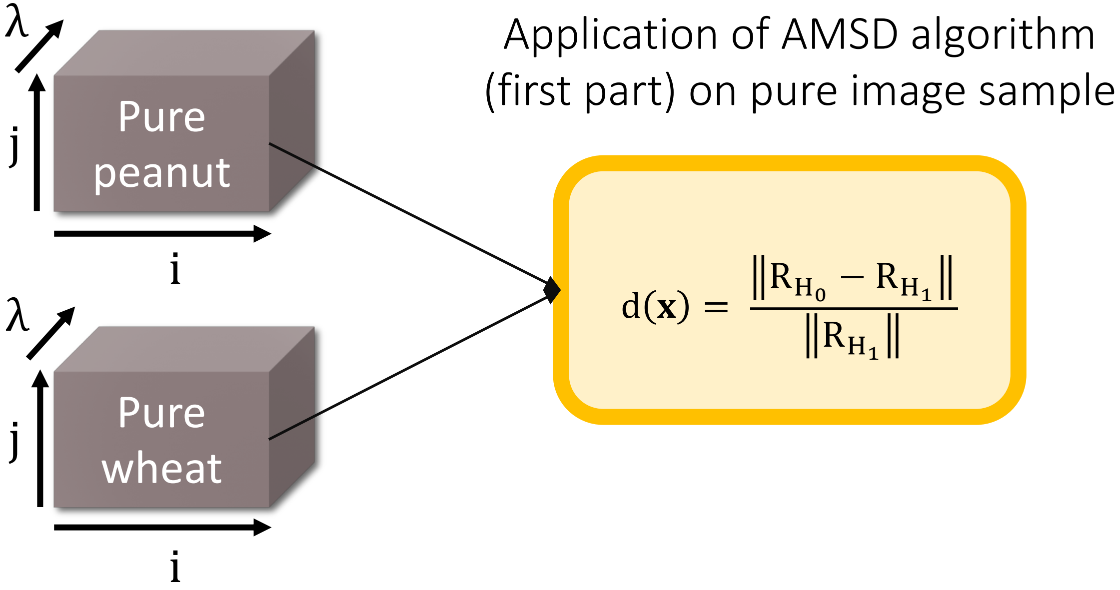
0: no detection

1: detection

AMSD algorithm (complete)

Threshold choice (θ)

- For pure images, true detection results are known → can be used to find the threshold for the detector.

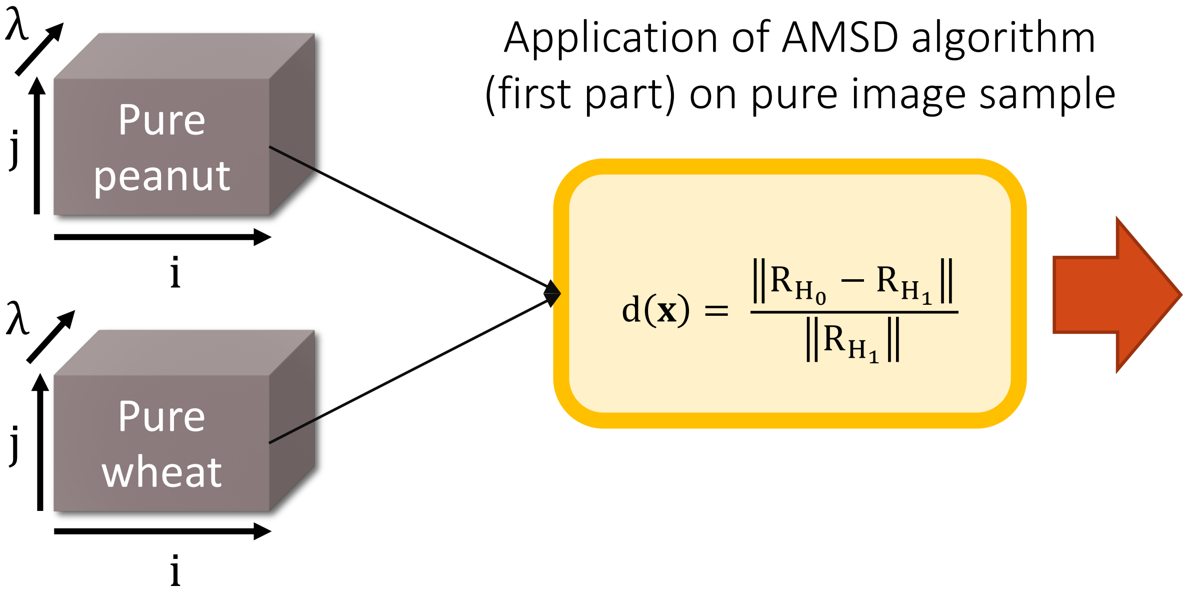


Application of AMSD algorithm (first part) on pure image sample

$$d(\mathbf{x}) = \frac{\left\| R_{H_0} - R_{H_1} \right\|}{\left\| R_{H_1} \right\|}$$

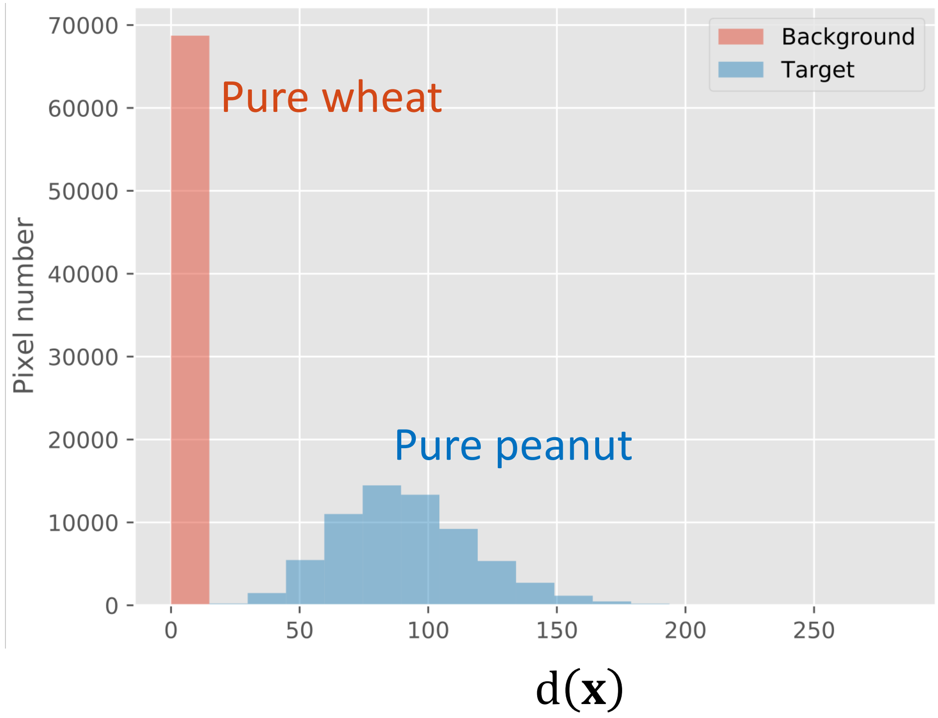
Threshold choice (θ)

- For pure images, true detection results are known → can be used to find the threshold for the detector

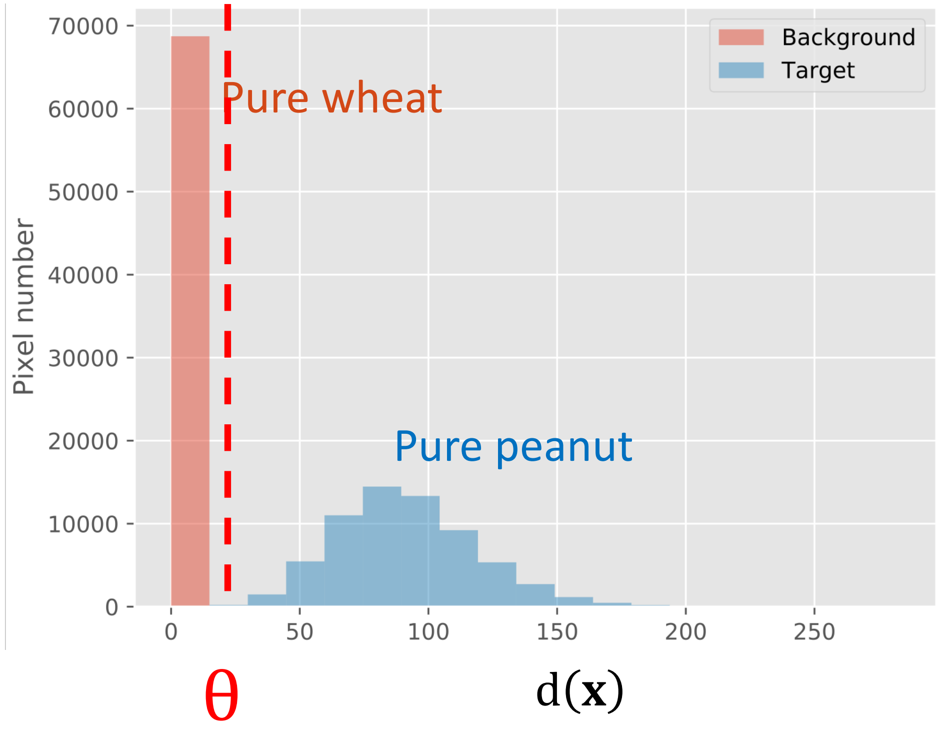


Application of AMSD algorithm (first part) on pure image sample

$$d(\mathbf{x}) = \frac{\left\| R_{H_0} - R_{H_1} \right\|}{\left\| R_{H_1} \right\|}$$

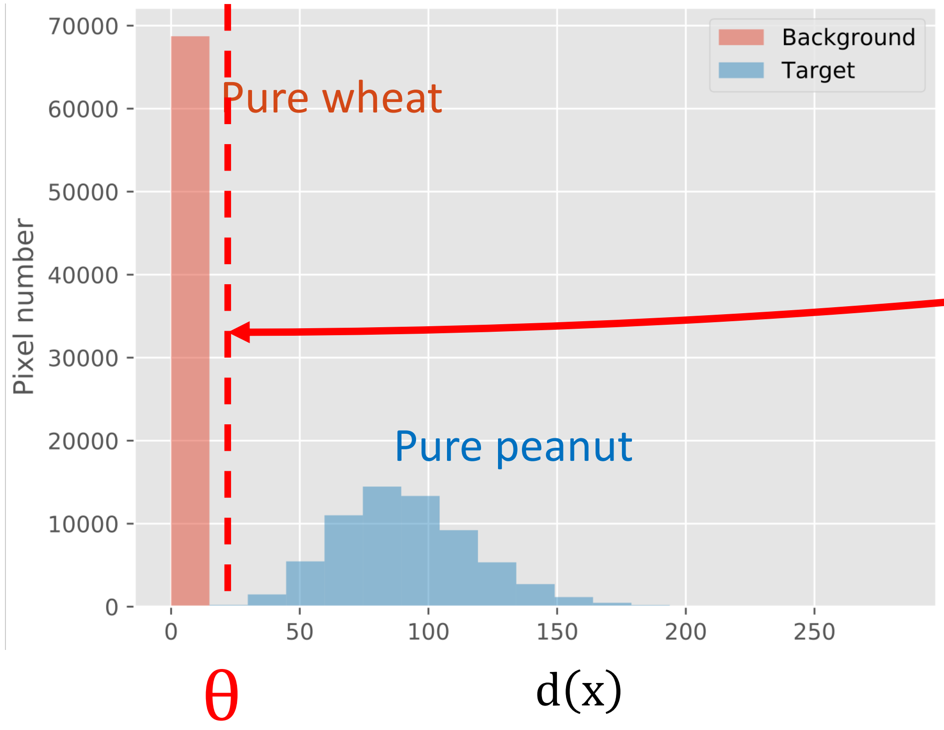


Threshold choice (θ)



P_D: detection probability
P_{FA}: false alarm probability

Threshold choice (θ)



$$\max P_D \text{ subject to } P_{FA} = 0 \Rightarrow \theta = \max(d(\mathbf{X}_{\text{wheat}}))$$

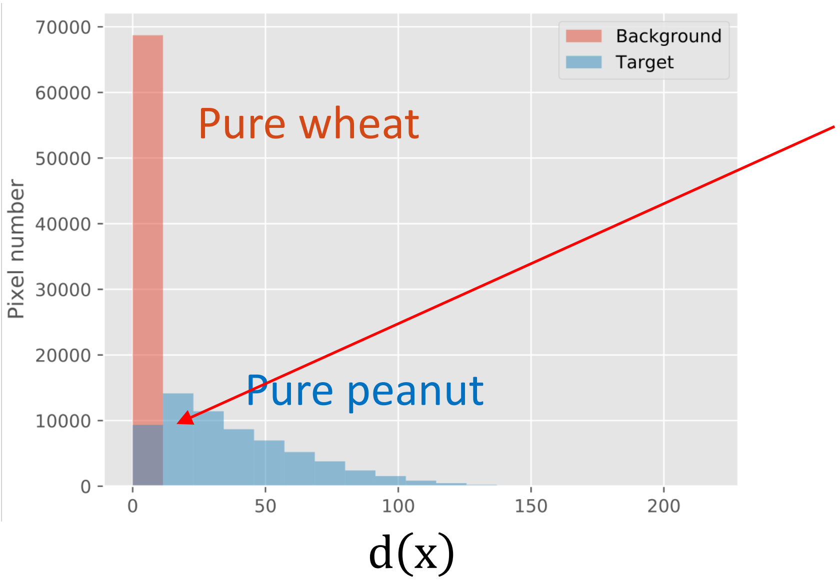
P_D: detection probability

P_{FA}: false alarm probability

Threshold dependence on subspace dimensionality

- The histogram of d depends on the choice of n
- Histograms below show that $\mathrm{n} = 2$ is a better choice than $\mathrm{n} = 1$ because the detection is not optimal on reference images

Histogram for references with $\mathrm{n} = 1$



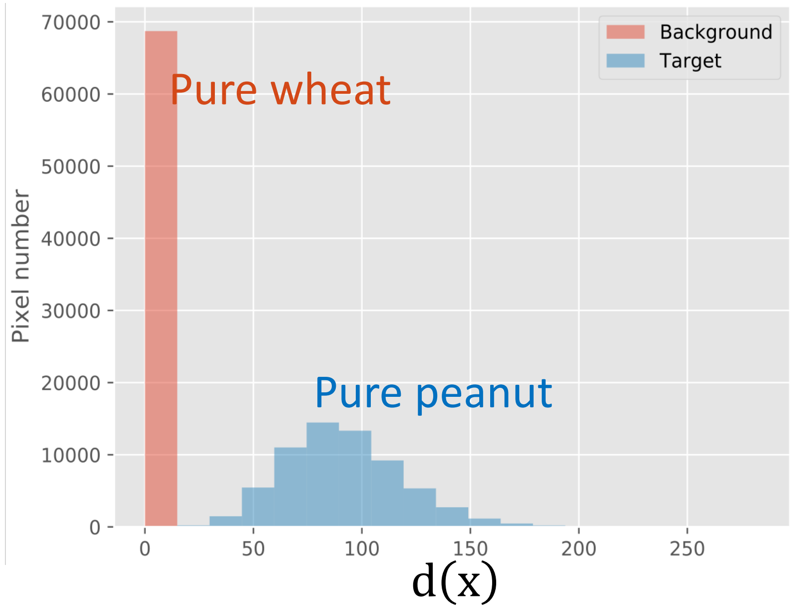
$\mathrm{d(x)}$

Pure wheat

Pure peanut

Overlap !

Histogram for references with $\mathrm{n} = 2$



$\mathrm{d(x)}$

Pure wheat

Pure peanut

Adaptive Matched Subspace Detector (AMSD)

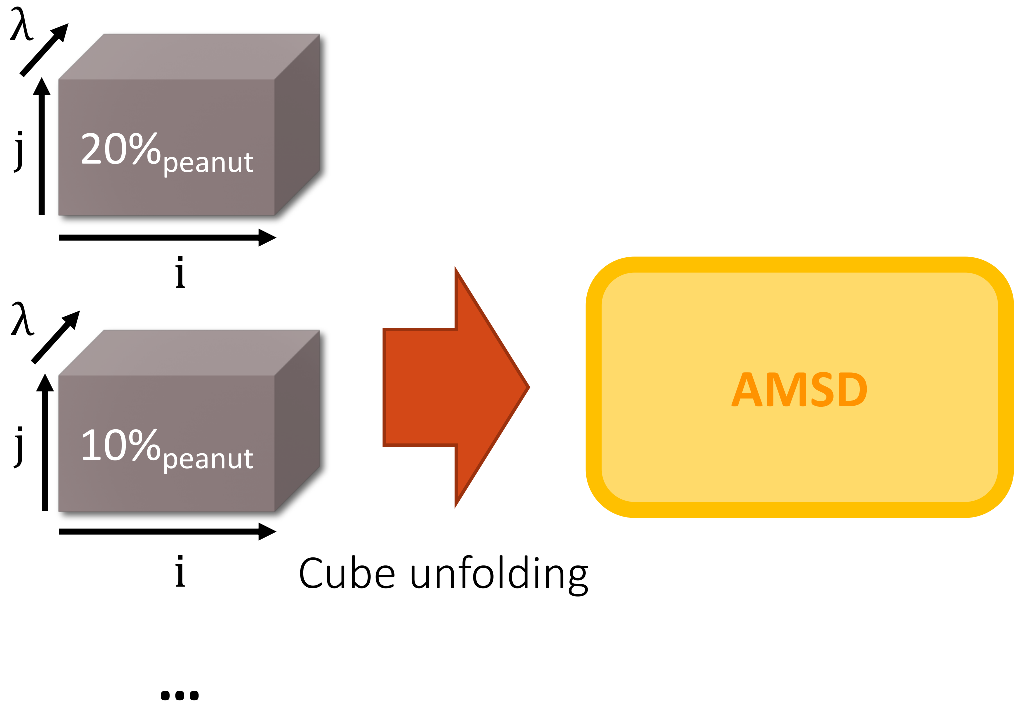
- AMSD approach is performed in 3 steps:

1. Construction of the **target** (S_b) and the **background** (S_t) **subspaces**

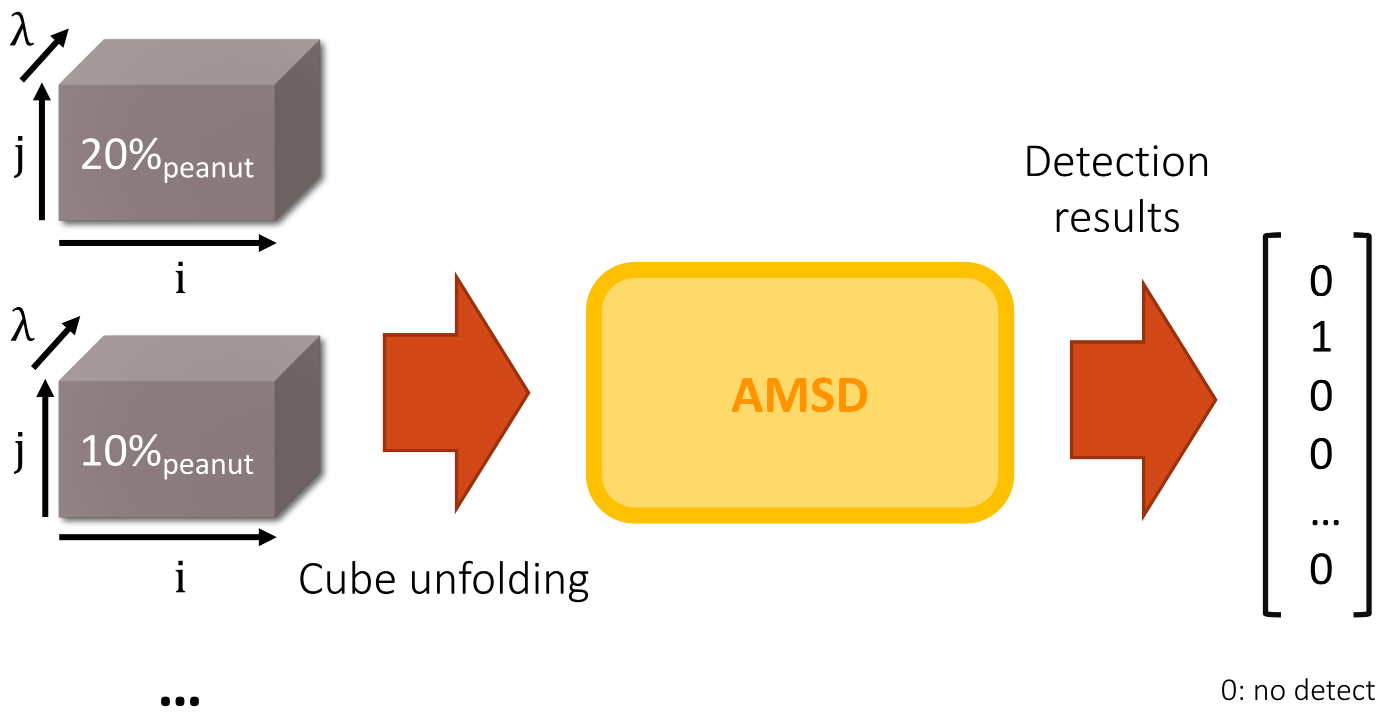
2. Design of the detector:
   - n stands for the target and background **subspace dimensionality**
   - θ stands for the **detection threshold**

3. **Application** on test images

AMSD application on test images



AMSD application on test images



Cube unfolding

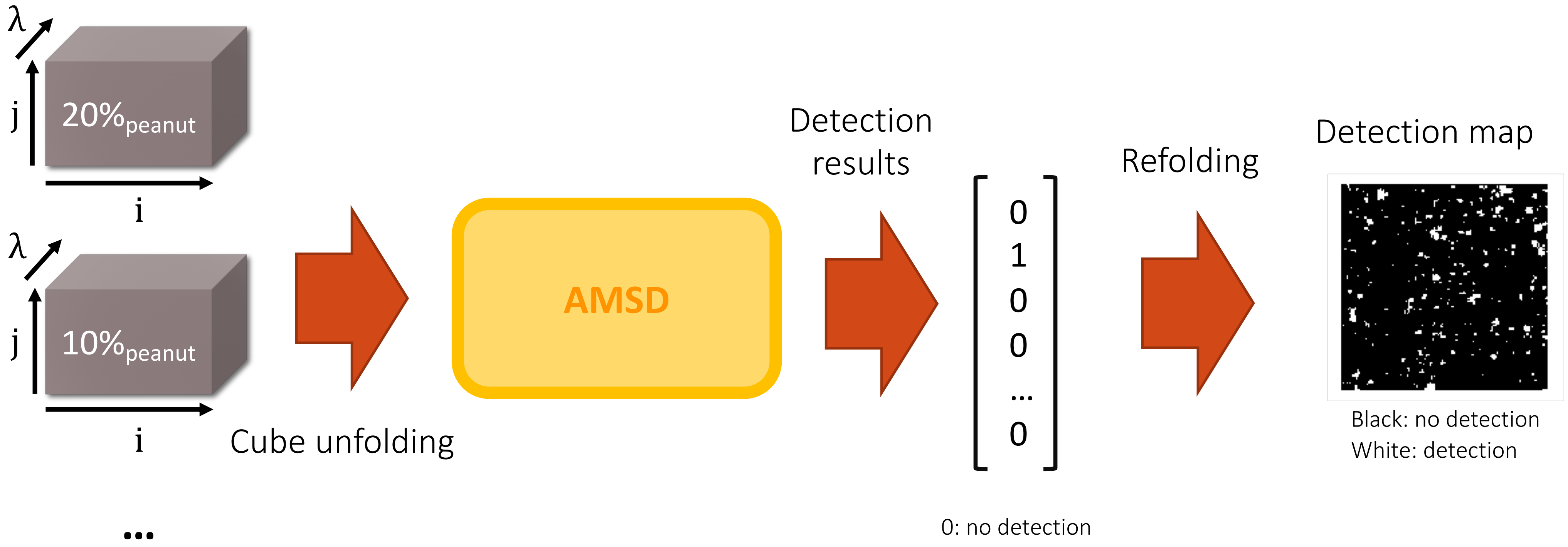
AMSD

Detection results

$$\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

0: no detection
1: detection

AMSD application on test images



$20\%_{peanut}$

$10\%_{peanut}$

…

Cube unfolding

AMSD

Detection results

$$\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

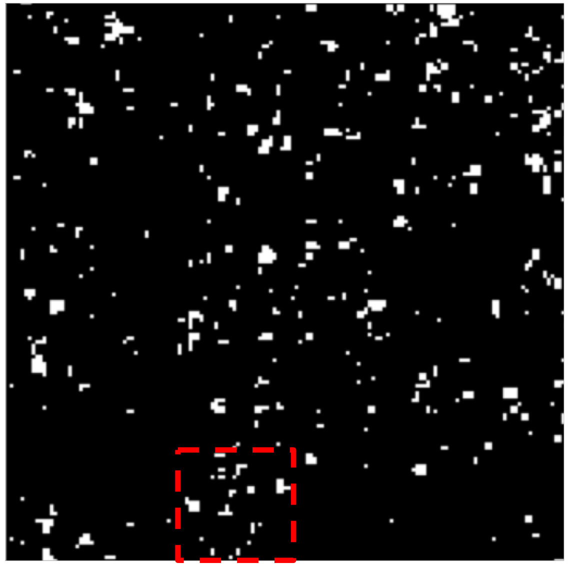
0: no detection
1: detection

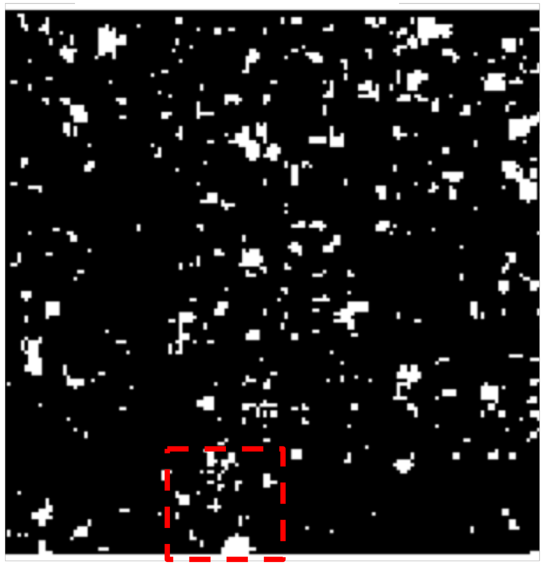
Refolding

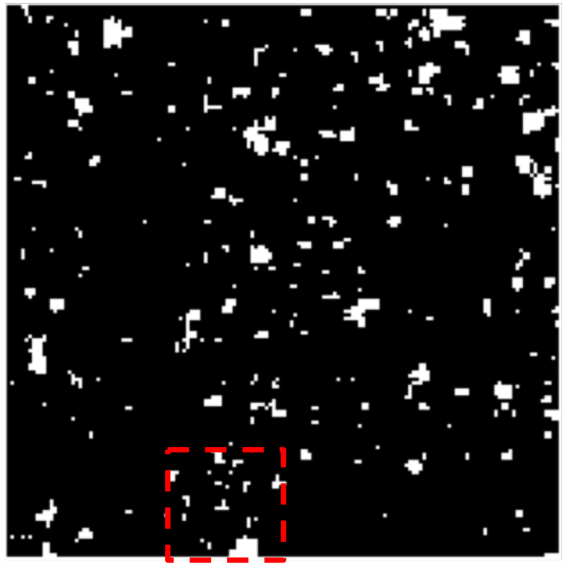
Detection map

Black: no detection
White: detection

AMSD application on test images – influence of n


10% - $n = 1$

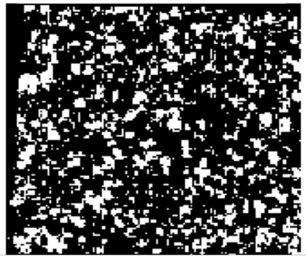

10% - $n = 2$


10% - $n = 3$

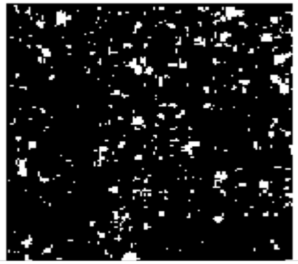
- AMSD with $n = 2$ tends to provide more detected pixels than AMSD with $n = 1$ (underfitting) and $n = 3$ (overfitting ?)

AMSD application on test images

20% - $n = 2$



10% - $n = 2$

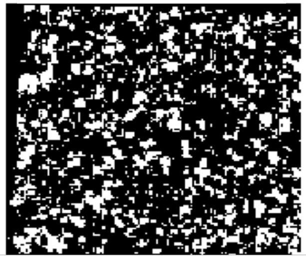


Decreasing global concentration

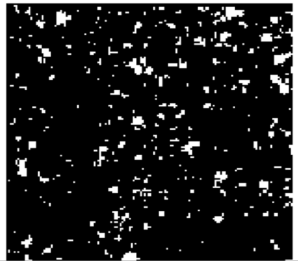
5% - $n = 2$

AMSD application on test images

20% - $n = 2$



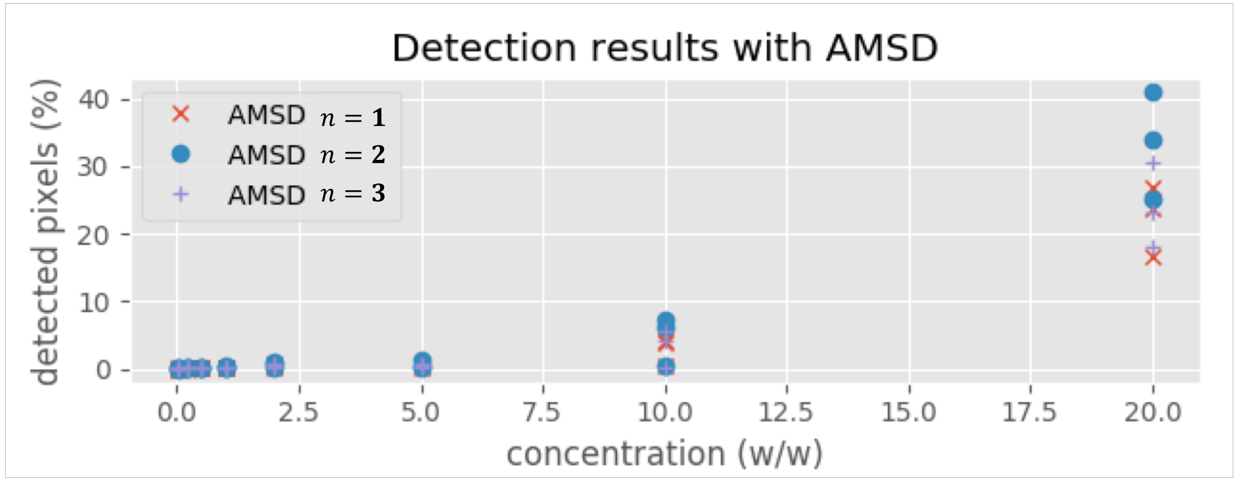
10% - $n = 2$



5% - $n = 2$

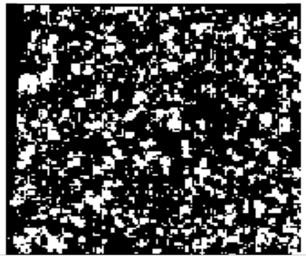
Decreasing global concentration

Detected pixels counting

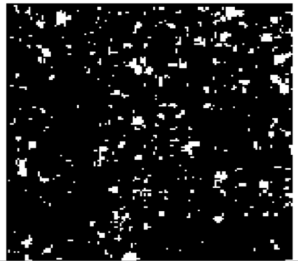


AMSD application on test images

20% - $n = 2$



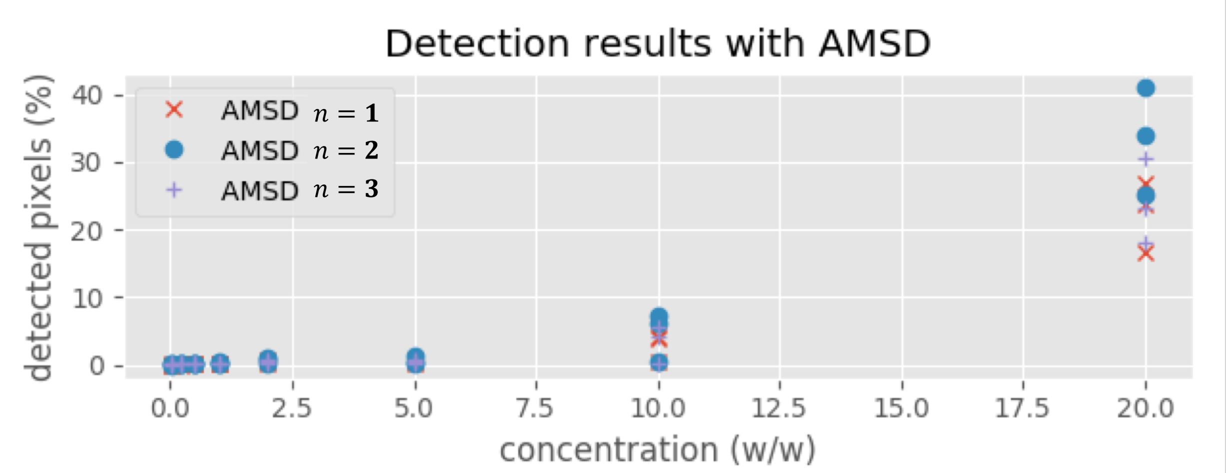
10% - $n = 2$



5% - $n = 2$

Decreasing global concentration

Detected pixels counting



Correlation between the number of detected pixels (%) and the global sample concentration

	$n = 1$	$n = 2$	$n = 3$
Pearson coefficient	0.901	0.902	0.897

Conclusions and perspectives

- **AMSD provides an appropriate metric for subpixel detection with close spectral endmembers**

- Detection map provides **convincing results** :
    - **High correlation score** with sample concentration (≥ 0.9)
    - **Stable detection map** with respect to the dimension of the subspace

- Detector sensitivity seems promising as it detects particles in samples with a **200 ppm concentration** (repeatable on three replica)

- Pending question: **detection limit/sensitivity** of the detector at the **pixel scale**: difficult as we don’t have any reference value at the pixel level…

References

Peanut allergy

- W. Burks, “Peanut allergy,” Lancet, vol. 371, no. 9623, pp. 1538–1546, 2008.
- E. W. Lusas, “Food uses of peanut protein,” J. Am. Oil Chem. Soc., vol. 56, no. 3, pp. 425–430, 1979.
- C. Hadley, “Food allergies on the rise?,” EMBO Rep., vol. 7, no. 11, pp. 1080–1083, 2006.
- Oqali, “Etude des allergènes dans les produits transformés disponibles sur le marché français entre 2008 et 2012,” Paris, 2015.

Peanut detection

- P. Mishra, A. Herrero-Langreo, P. Barreiro, and J. M. Roger, “Detection and quantification of peanut traces in wheat flour by near infrared hyperspectral imaging spectroscopy using principal-component analysis,” J. Near Infrared Spectrosc., vol. 23, no. 1, pp. 15–22, 2015.
- P. Mishra, C. B. Y. Cordella, D. N. Rutledge, P. Barreiro, J. M. Roger, and B. Diezma, “Application of independent components analysis with the JADE algorithm and NIR hyperspectral imaging for revealing food adulteration,” J. Food Eng., vol. 168, pp. 7–15, 2016.

AMSD

- D. Manolakis, C. Siracusa, and G. Shaw, “Hyperspectral subpixel target detection using the linear mixing model,” IEEE Trans. Geosci. Remote Sens., vol. 39, no. 7, pp. 1392–1409, 2001.

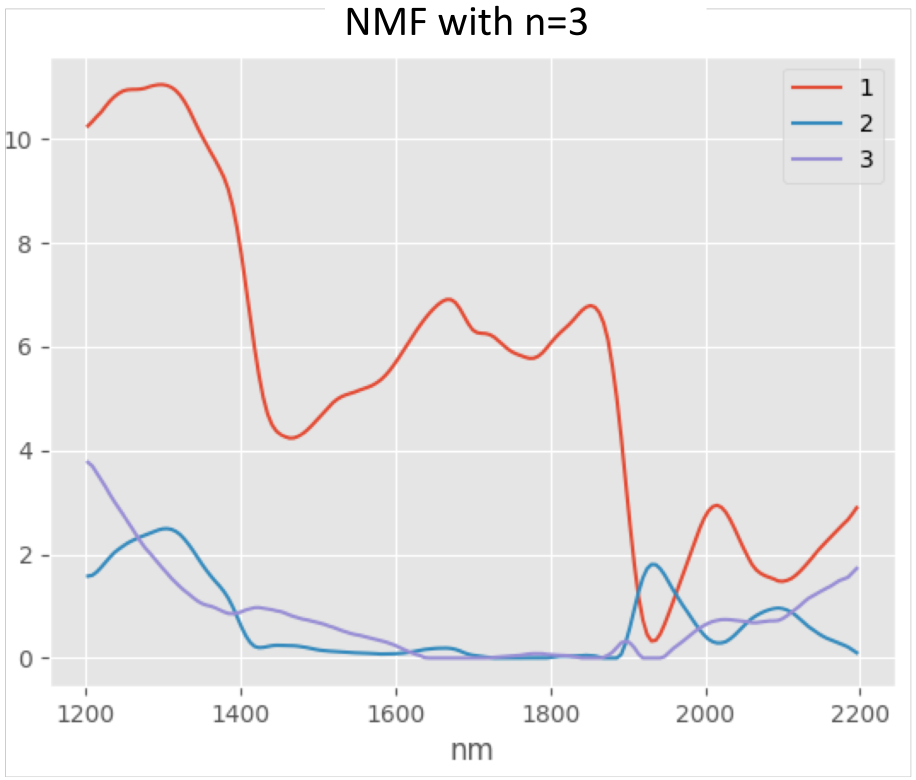
Thank your for your attention

Antoine Laborde[1], Benoît Jaillais[2], Delphine Jouan-Rimbaud Bouveresse[3], Anthony Boulanger[1], Christophe B.Y. Cordella[3]
[1]Greentropism, 75116 Paris, France.
[2]StatSC INRA/ONIRIS, 44322 Nantes, France.
[3]UMR914 PNCA INRA/AgroParisTech, 75231 Paris, France.

Contact: antoine.laborde@greentropism.com

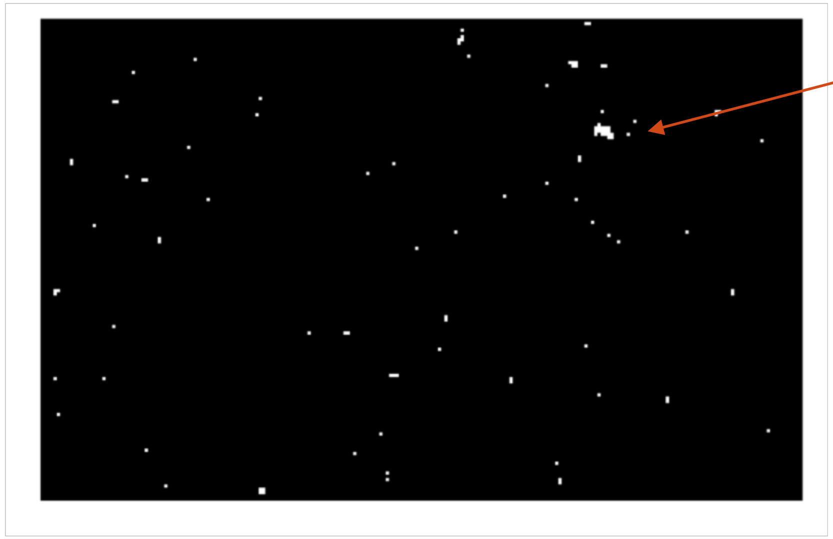
Additional slides: NMF components

NMF components obtained on wheat flour


NMF with n=3

- First component is clearly similar to the average spectrum of wheat flour
- NMF is not a nested method like PCA: components shape changes according to the number of requested components
- NMF components are constrained to be positive

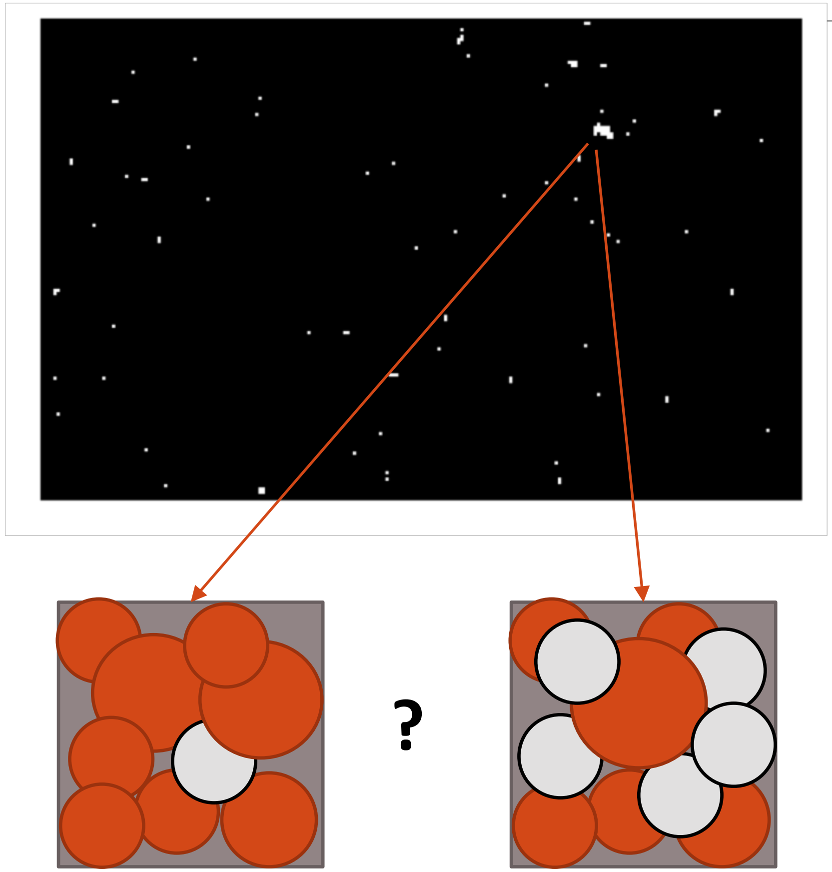
Additional slide: homogeneity assumption



Clusters of particles are detected

- Clusters of detected particles are detected on the maps.
- The assumption that mixtures are homogeneous is clearly wrong
- Very difficult to mix small particles (electrostatic effects that counteract blending when particle size is $\leq 100\ \mu\text{m}$)

Additional slide : detection linearity

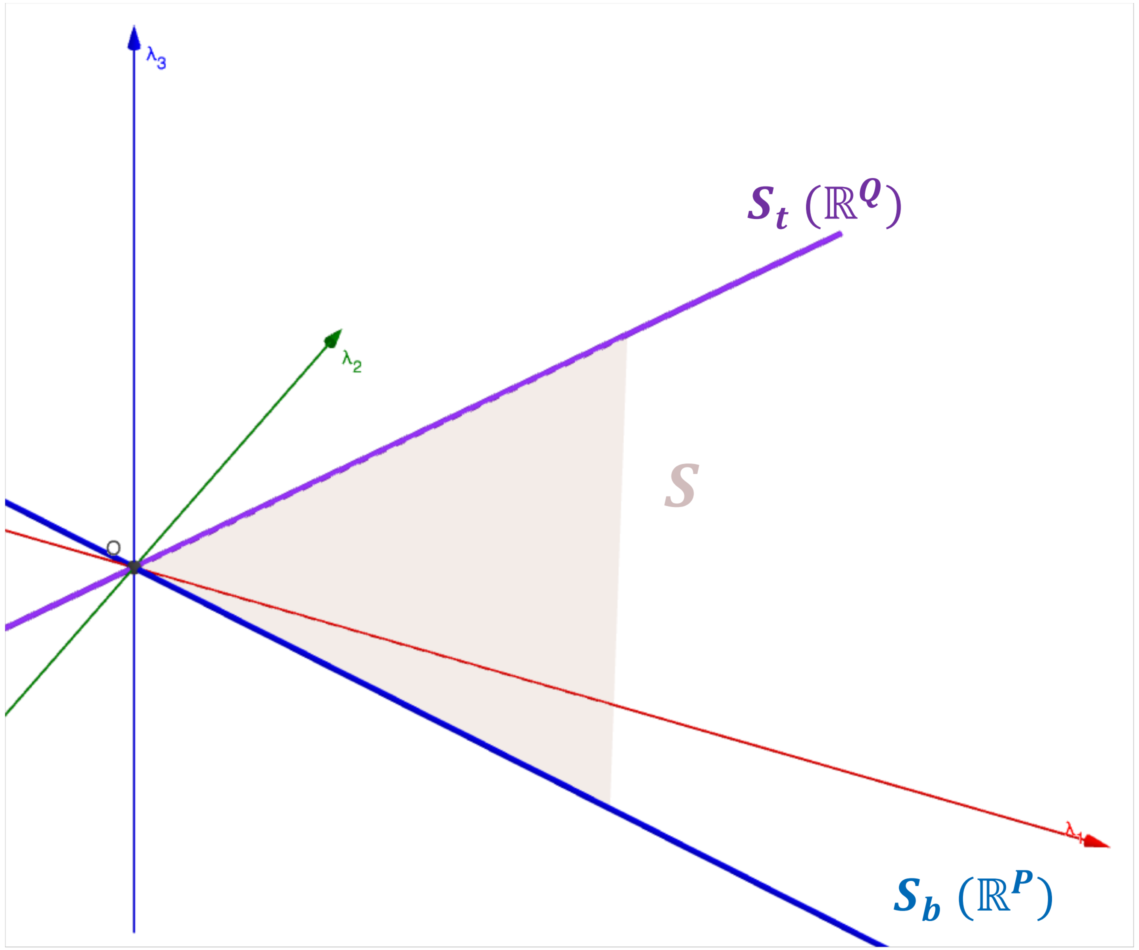


Both count for 1 detected pixel but they have ≠ concentration levels

Several reasons why the linearity between number of detected pixels and sample concentration can be broken:

- We do not perform quantification at the pixel level
- Intimate powder mixing creates nonlinear spectral mixing
- A white pixel (detection) only means the target spectral abundance has crossed the threshold

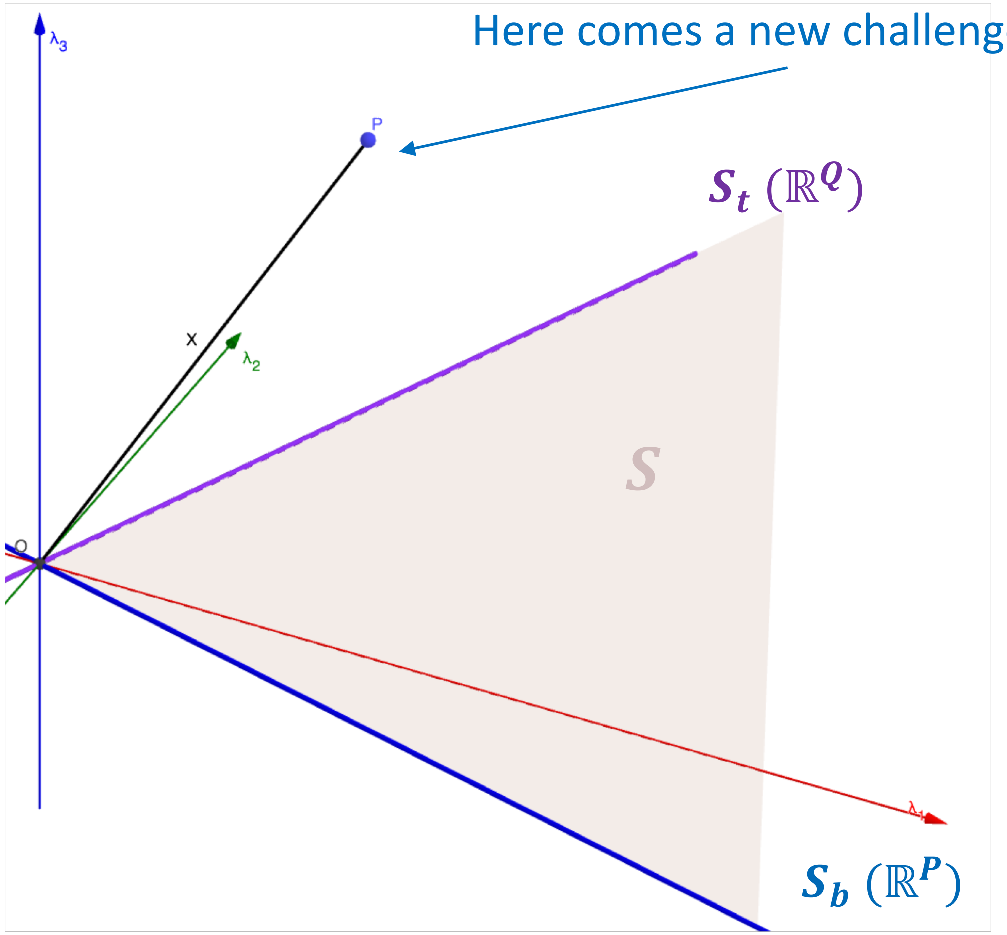
Additional slide: geometric interpretation



- Assume a spectrum lies in $\mathbb{R}^3$ $(\lambda_1, \lambda_2, \lambda_3)$ for simplicity.
- Assume the background ($\boldsymbol{S_b}$) and the target ($\boldsymbol{S_t}$) spaces lie in a 1 dimensional subspace of $\mathbb{R}^3$

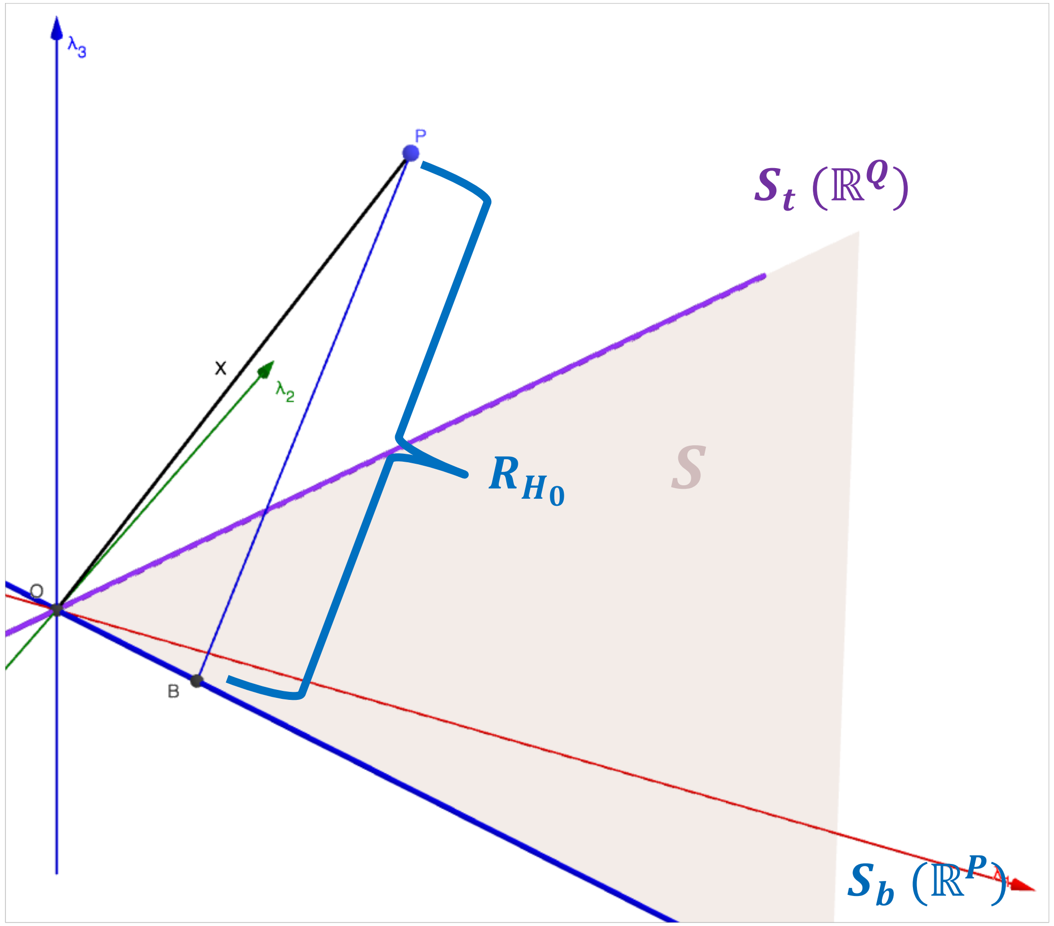
- The LMM subspace for modelling is $\boldsymbol{S} = \boldsymbol{S_t} \cup \boldsymbol{S_b}$

Additional slide: geometric interpretation



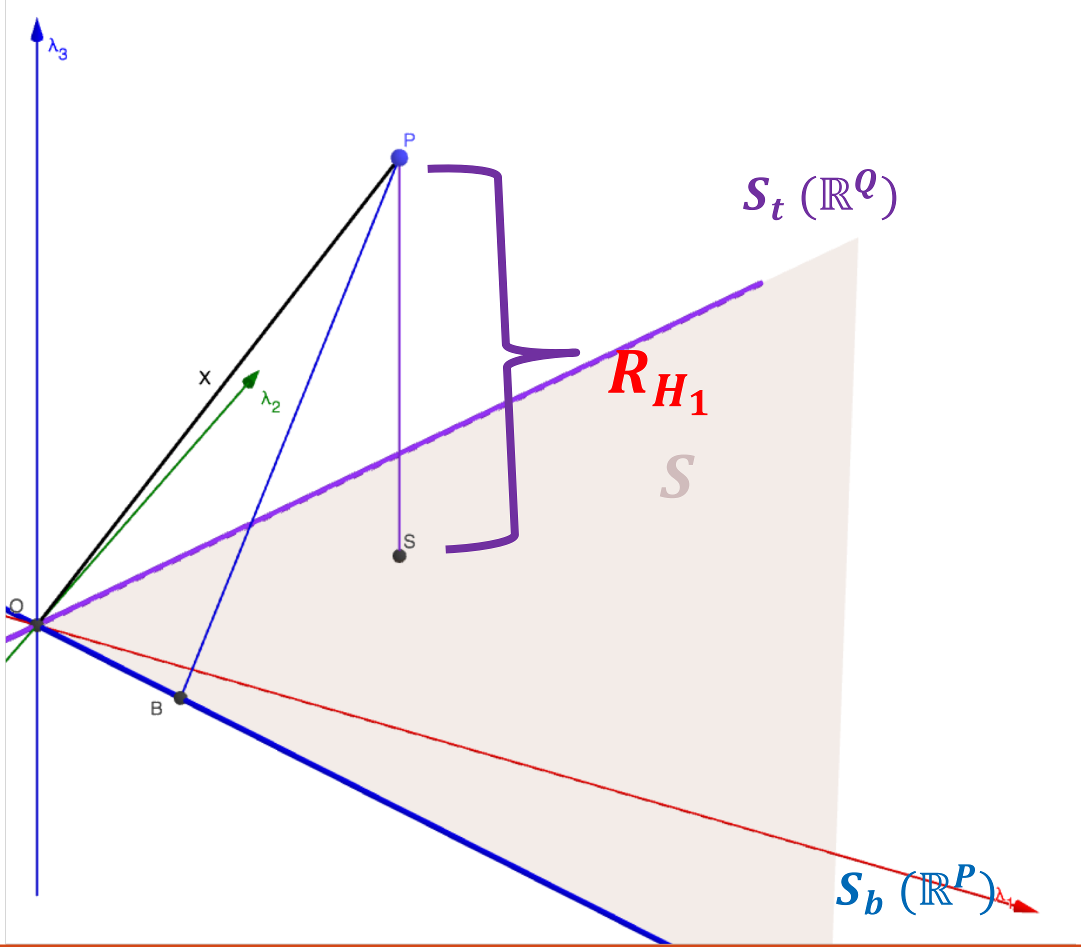
- x is a test spectrum

Additional slide: geometric interpretation



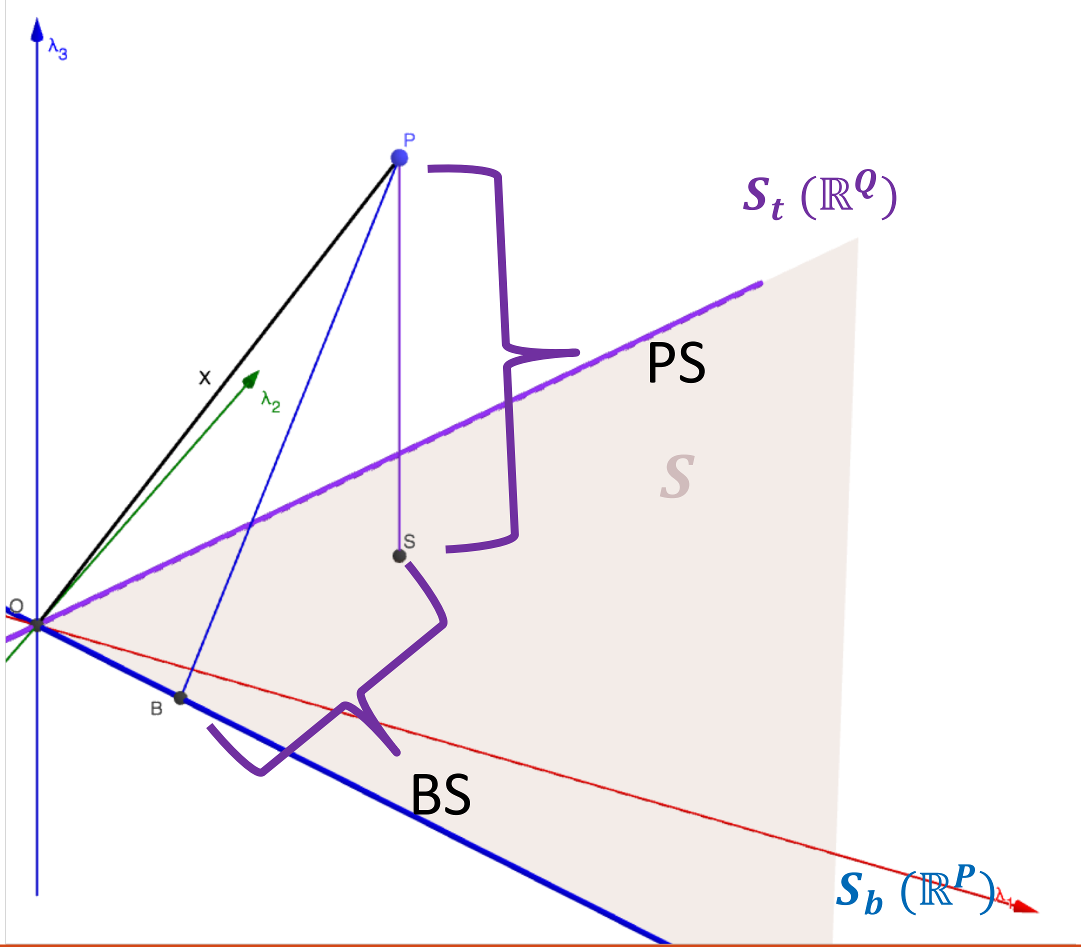
- x is a test spectrum
- x is projected onto $\boldsymbol{S_b}$ according to $\boldsymbol{H_0}$: $\boldsymbol{x = S_b a_b + R_{H_0}}$ and the residual can be calculated

Additional slide: geometric interpretation



- $\mathbf{x}$ is a test spectrum
- $\mathbf{x}$ is projected onto $\mathbf{S_b}$ according to $\mathbf{H_0}: \mathbf{x} = \mathbf{S_b a_b} + \mathbf{R_{H_0}}$ and the residual can be calculated
- $\mathbf{x}$ is projected onto $\boldsymbol{S}$ according to $\mathbf{H_1}: \mathbf{x} = \mathbf{S_t a_t} + \mathbf{S_b a_b} + \mathbf{R_{H_1}}$ and the residual can be calculated

Additional slide: geometric interpretation



- $\mathbf{x}$ is a test spectrum
- $\mathbf{x}$ is projected onto $\mathbf{S}_\mathrm{b}$ according to H_0 and the residual can be calculated
- $\mathbf{x}$ is projected onto $\mathbf{S}$ according to $\mathrm{H}_1$and the residual can be calculated
- AMSD ratio is the comparison of PS with BS
  - PS → $\mathrm{R}_{\mathrm{H}_1}$
  - BS → $\mathrm{R}_{\mathrm{H}_0} - \mathrm{R}_{\mathrm{H}_1}$
- AMSD ratio evaluate vector norm $\mathrm{d}(\mathbf{x}) = \dfrac{\left\|\mathrm{R}_{\mathrm{H}_0} - \mathrm{R}_{\mathrm{H}_1}\right\|}{\left\|\mathrm{R}_{\mathrm{H}_1}\right\|}$