Silesian University of Technology

Fuzzy-inference system for isotopic envelope identification based on analysis of the spatial distribution of components in Mass Spectrometry Imaging

Anna Glodek

30.06.2022

European Funds
Knowledge Education Development

AIDA
APPLIED INTEGRATIVE DATA ANALYSIS

European Union
European Social Fund

Outline

Mass spectrometry, Mass Spectrometry Imaging (MSI)

Isotopic envelope identification – problem statement

Algorithm details

Fuzzy C-means clustering

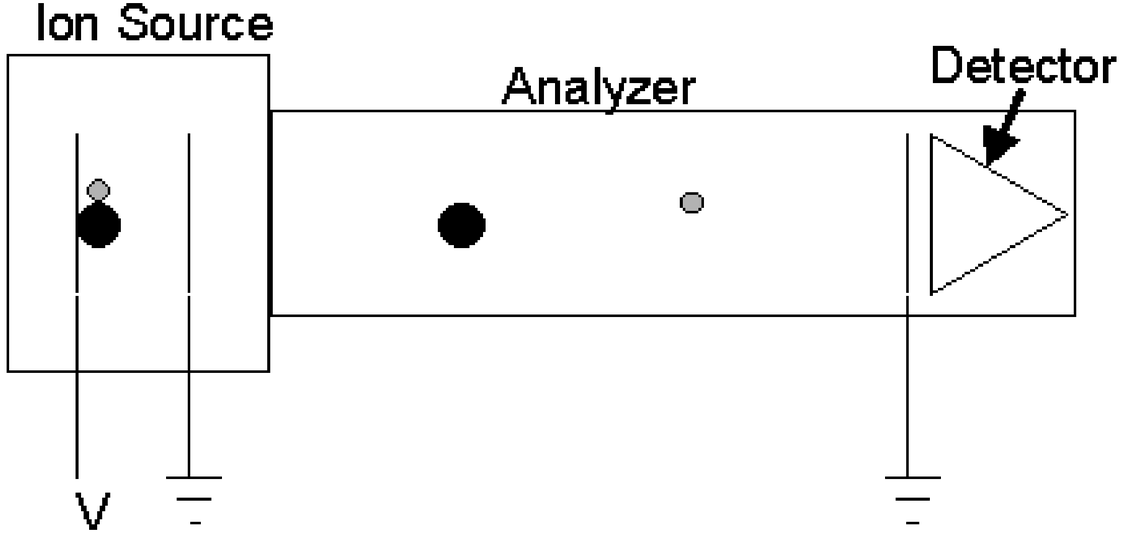
Summary

MALDI-ToF mass spectrometry & MSI

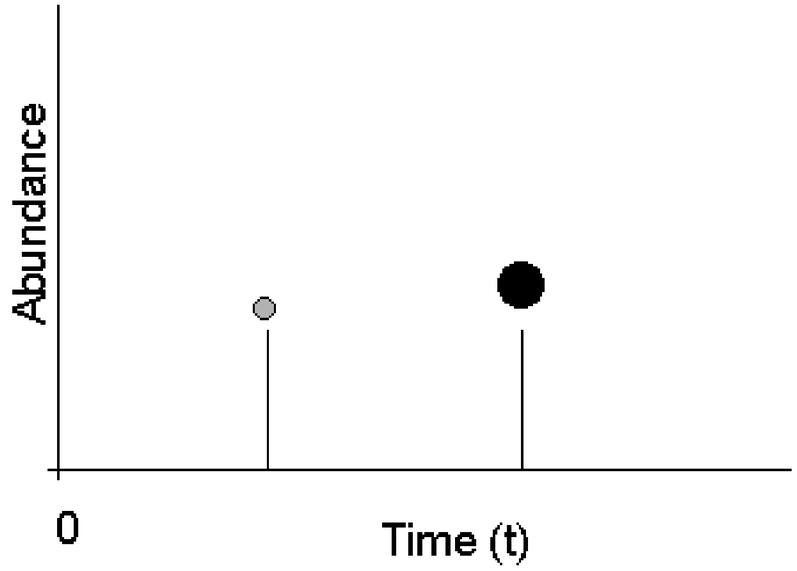
- **Matrix Assisted Laser Desorption / Ionisation – Time of Flight**
- a sample is mixed with a matrix and applied to a metal plate
- a laser irradiates the sample and analyte molecules are ionized **(+1/-1)**, rarely **+2**
- **enables the registration of isotopic envelopes**
- masses are analysed within the following range: **500 – 1 000 000 Da**

MSI

- mass spectrometry is performed directly on the sample surface
- used for gathering information on the molecules distribution (lipids, peptides, proteins, biomarkers) **in the sample**

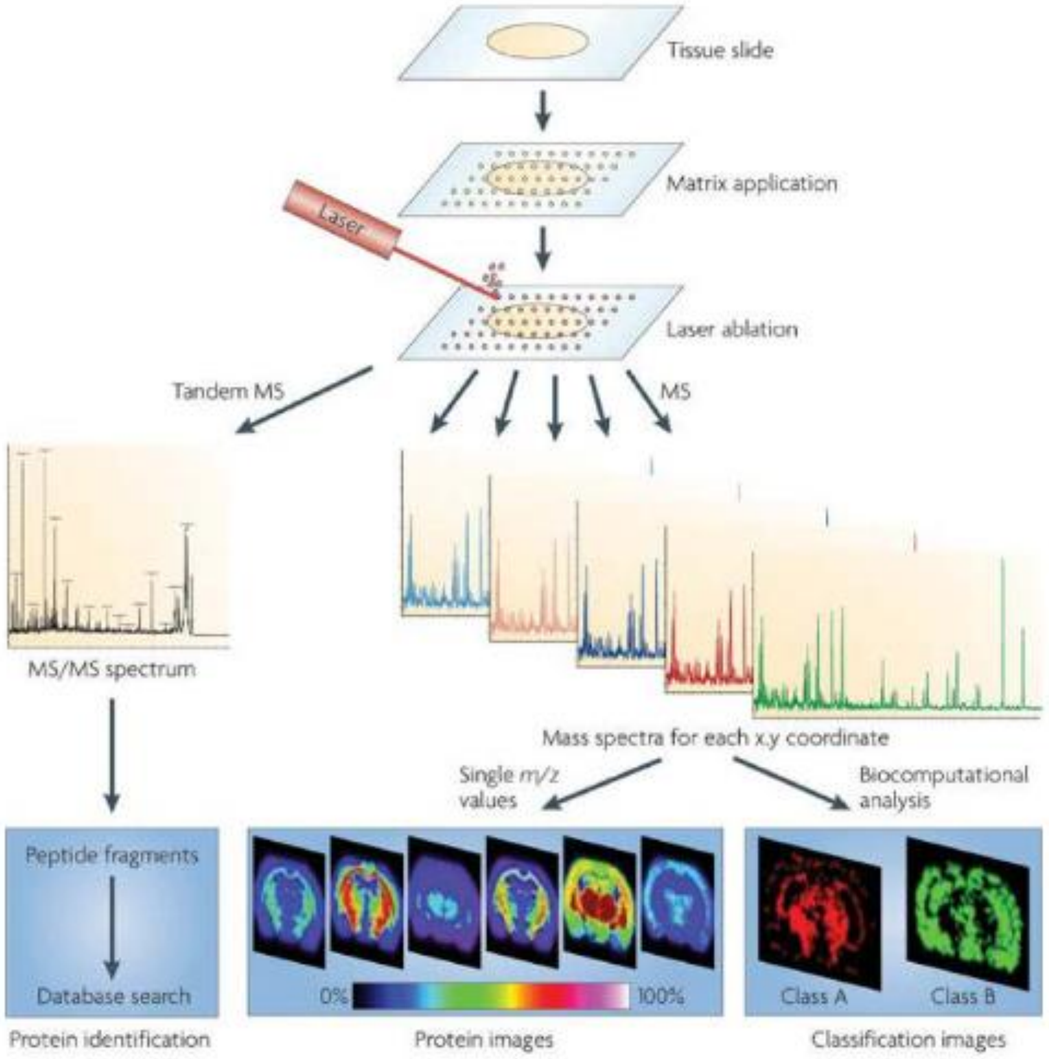


Mass spectrometer – ions are separated according to their mass-to-charge ratio in an analyzer



Mass spectrum

MALDI-MSI



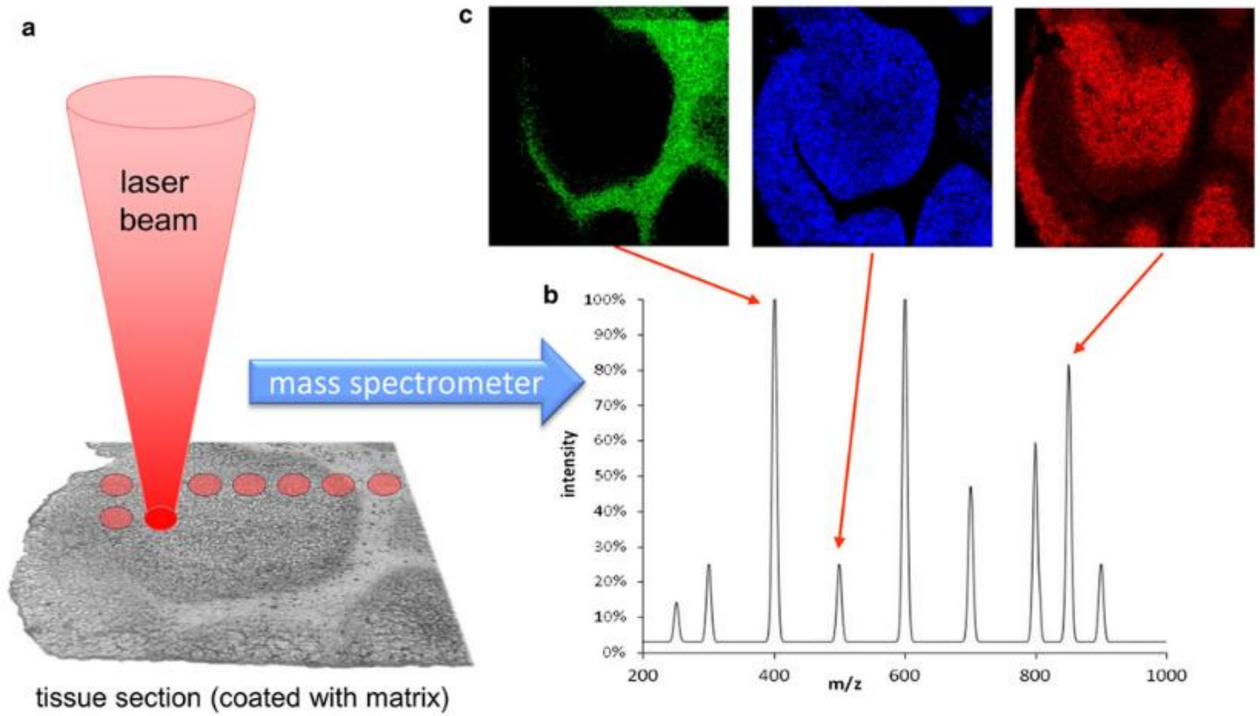


Fig. 1 Scheme of the mass spectrometry imaging process. **a** The tissue section is covered with matrix and irradiated by a pulsed laser beam. **b** Mass spectrum acquired from the tissue section. **c** MS images of different *m/z* peaks

Source: Norris JL, Caprioli RM. Analysis of tissue specimens by matrix-assisted laser desorption/ionization imaging mass spectrometry in biological and clinical research. Chem Rev 2013;113:2309–2342.

Mass spectrum

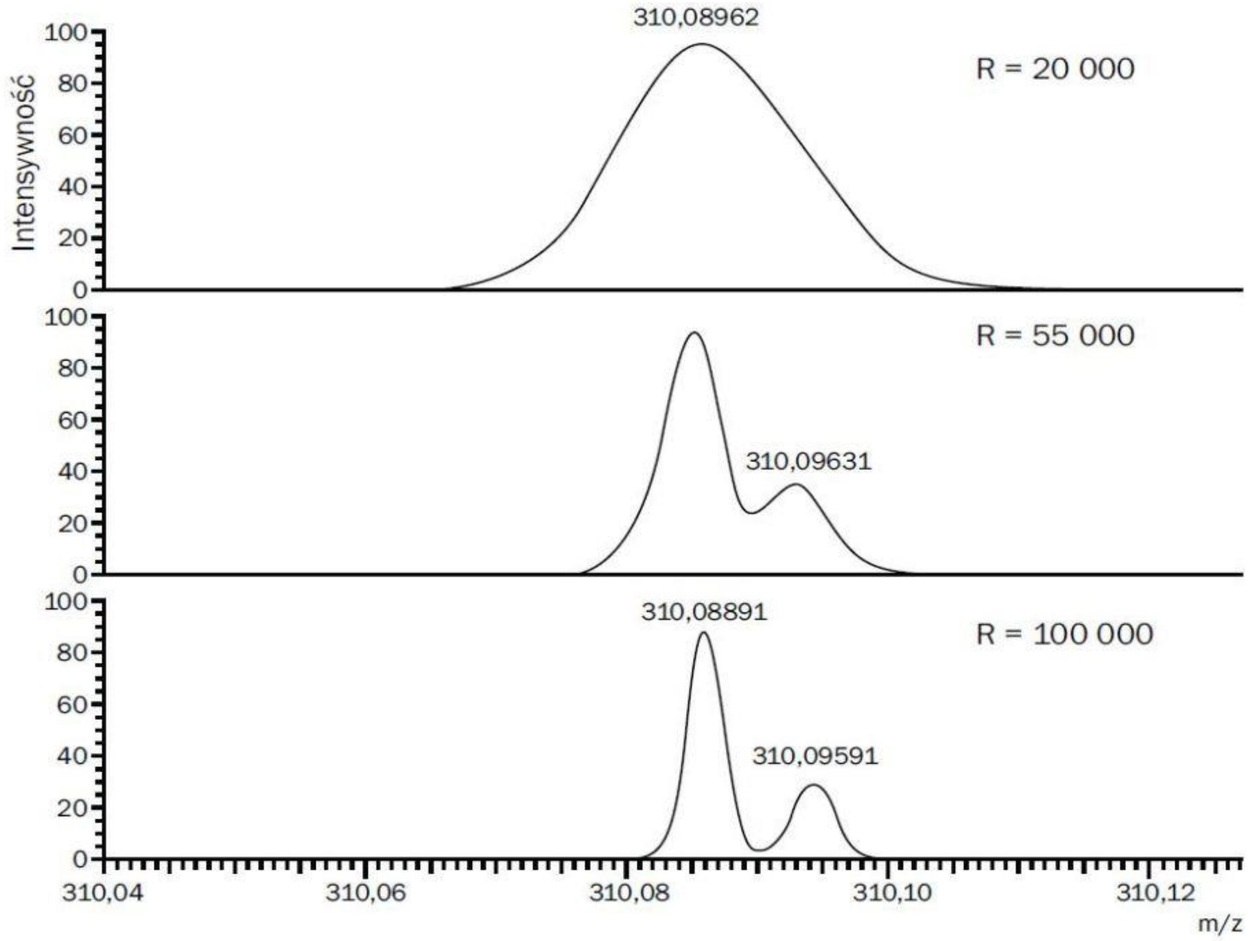
Two-dimensional representation of the signal intensity (abundance) vs mass-to-charge ratio (m/z)

Reference point: **C-12**

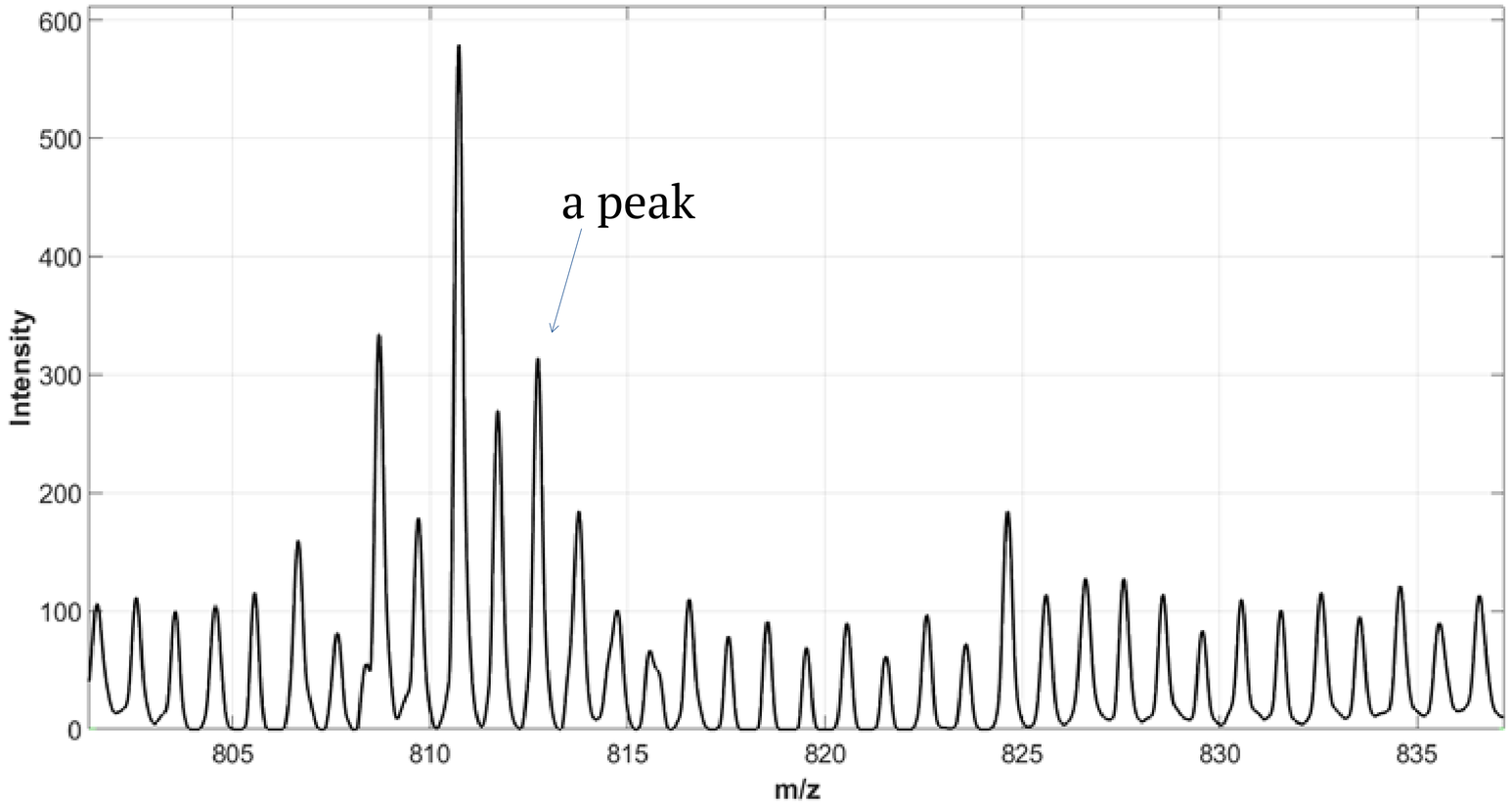
Unit: **Dalton** [Da]

$$1\ \text{Da} = \frac{1}{12}\ \text{mass of C-12}$$

Resolution of mass spectrometer

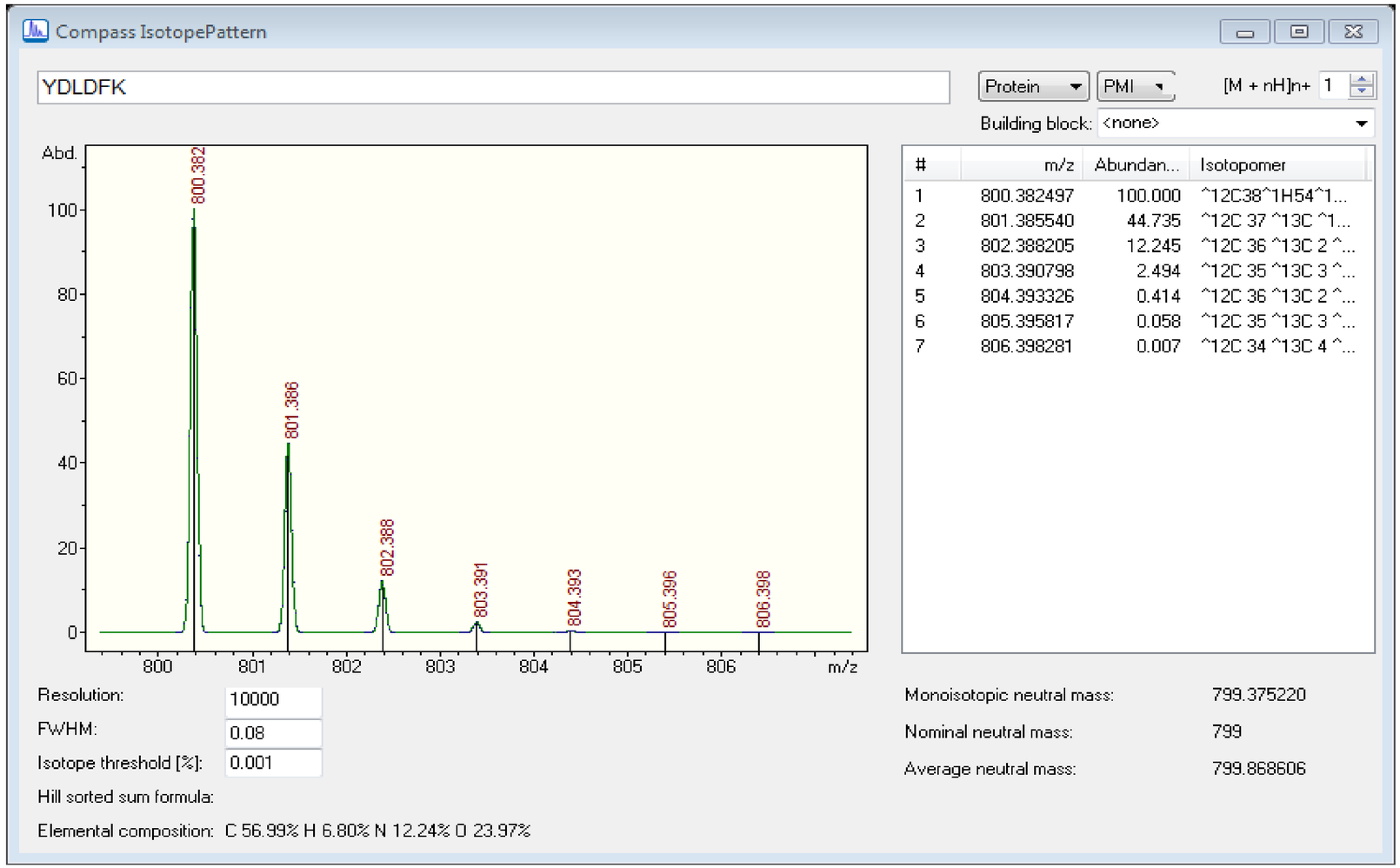


Mass spectrum



The mass spectrum can be considered as a **set of peaks**

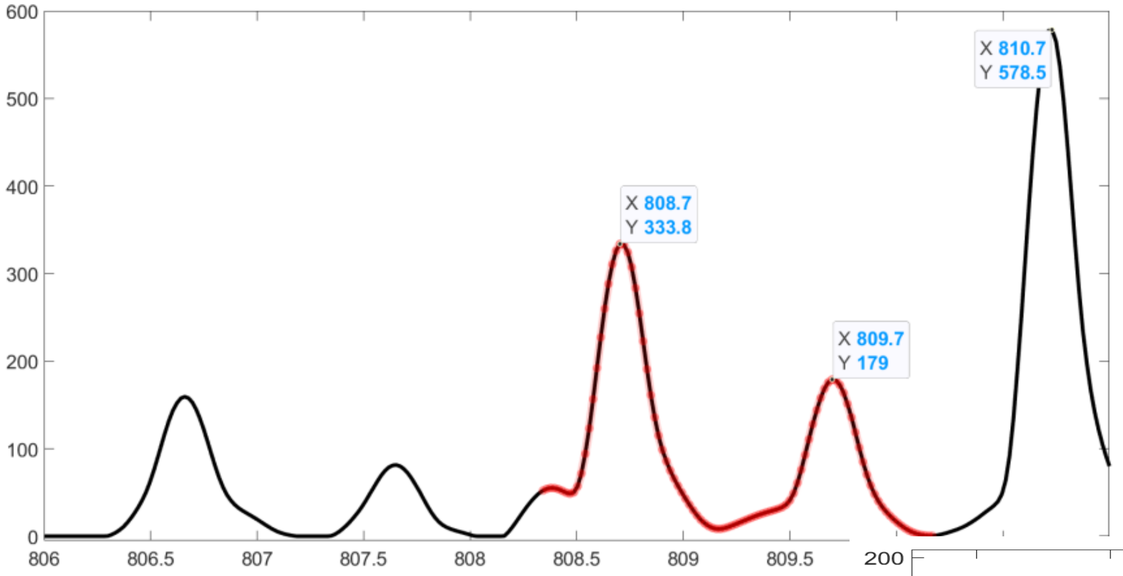
Isotopic envelope



Exemplary isotopic envelope of *YDLDFK* peptide

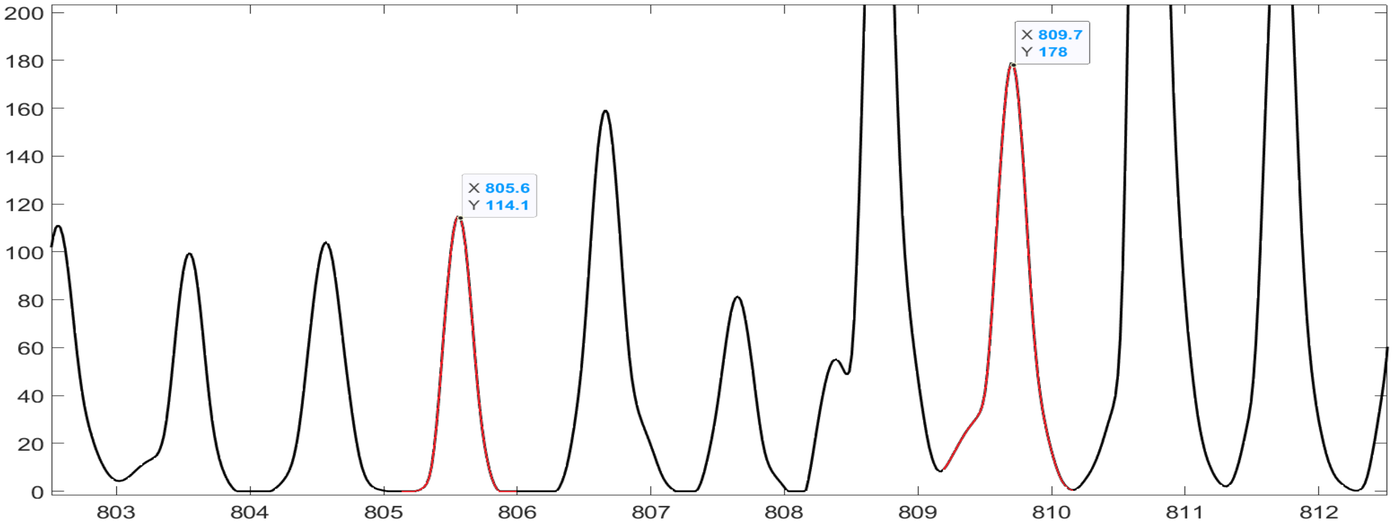
Isotopic envelope consists of the **isotopes of one compound**

Isotopic envelope

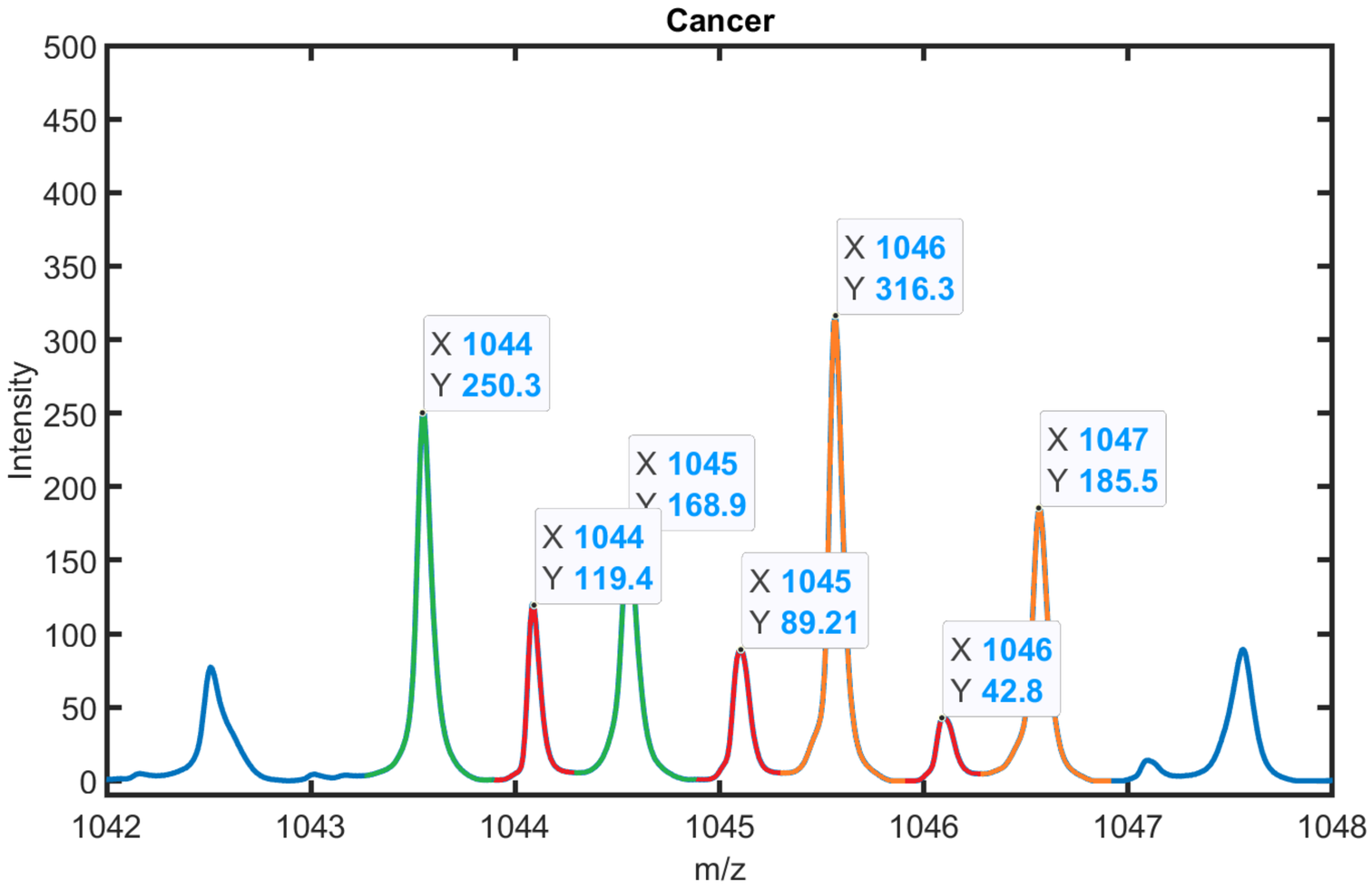


E (Envelope)

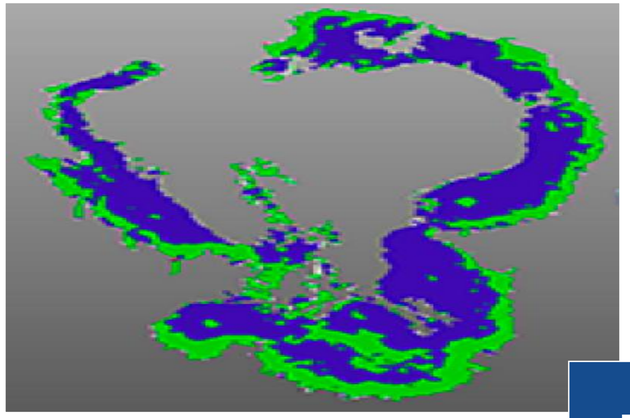
nE (non-Envelope)



Isotopic envelope



Data



Frozen tissue

Peptides

Head and neck cancer

9 492 averaged spectra with **109 568** **mass channels [m/z]**

2 435 peaks after pre-processing

(resampling, baseline removal, TIC normalisation, alignment to the average spectrum based on the Fast Fourier Transform, Gaussian Mixture Model)

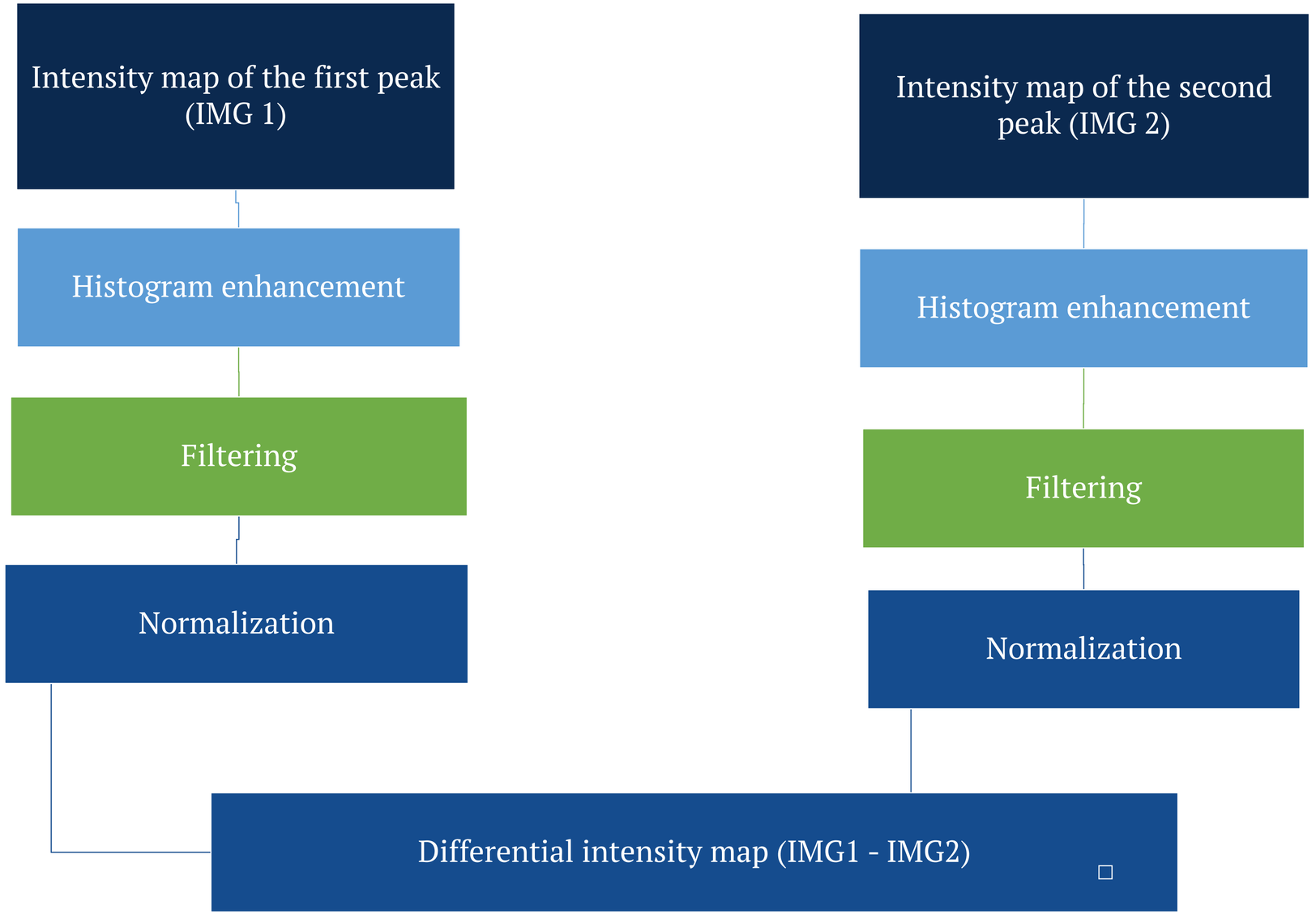
Pipeline

Mamdani-Assilan fuzzy-inference system

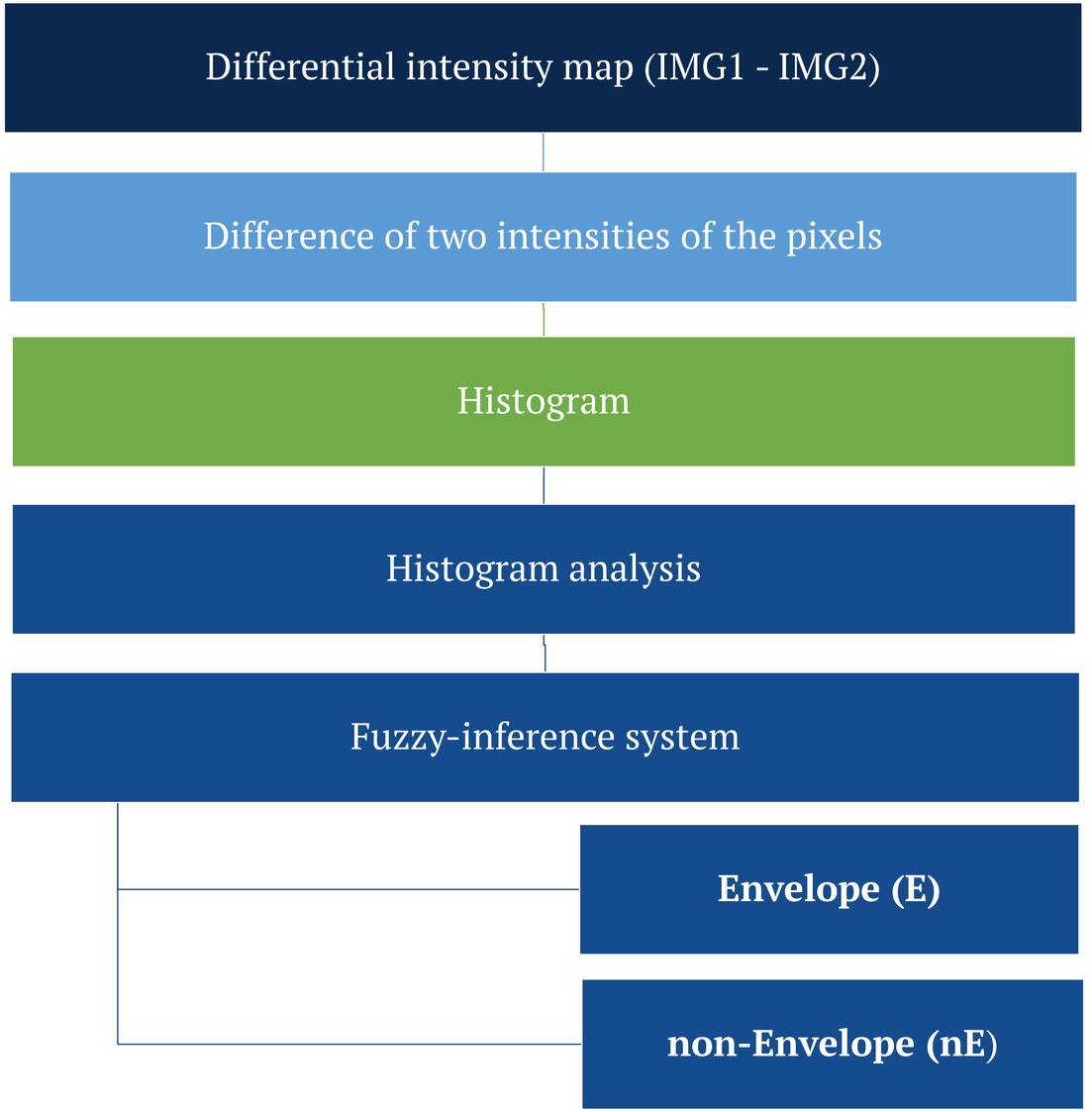
Algorithm based on the spatial distribution of peaks

Decision making process based on Sugeno fuzzy-inference system

Pipeline


Intensity map of the first peak (IMG 1)
Histogram enhancement
Filtering
Normalization
Intensity map of the second peak (IMG 2)
Histogram enhancement
Filtering
Normalization
Differential intensity map (IMG1 - IMG2)

Pipeline


Differential intensity map (IMG1 - IMG2)
Difference of two intensities of the pixels
Histogram
Histogram analysis
Fuzzy-inference system
Envelope (E)
non-Envelope (nE)

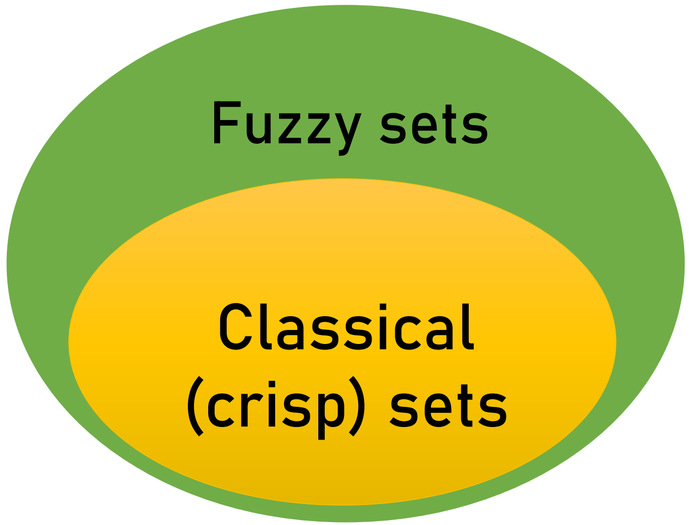
Pipeline: 1st step

Mamdani-Assilan fuzzy-inference system

A term ***fuzzy set*** was introduced by Professor **Lotfi Zadeh** in 1965.

A fuzzy set A in space $\mathbb{X}$ can be described by a function $\mu A(x)$ or by a set of ordered pairs $(x, \mu A(x))$, where $\mu A(x)$ represents a degree of membership of an object x to the fuzzy set A:

$$\mathbf{A = \{ (x, \mu A (x))| \ x \quad [0, 1]\}.}$$

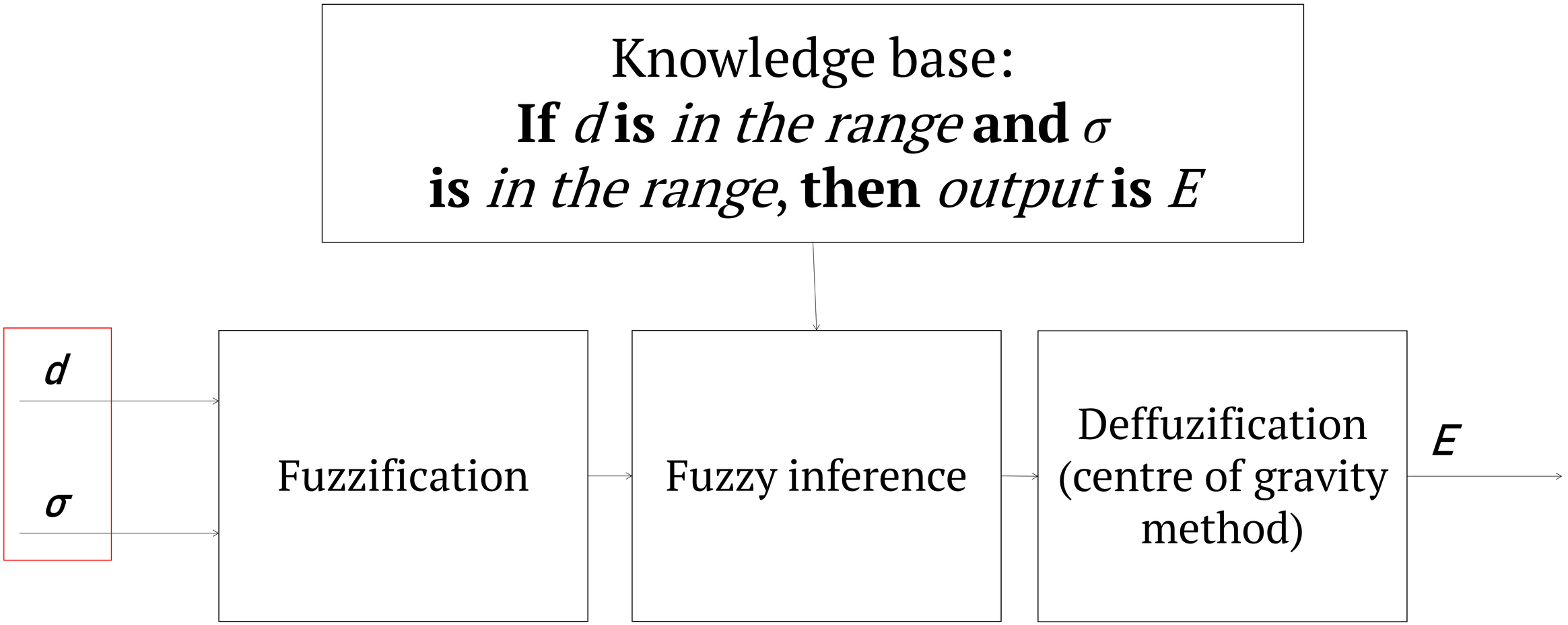


An element can be included in a fuzzy set in the following ways:

1. not included: **$\mu A (x) = 0$**
2. partially included: **$0 < \mu A (x) < 1$**
3. fully included: **$\mu A (x) = 1$.**

Pipeline: 1st step

Mamdani-Assilan fuzzy-inference system



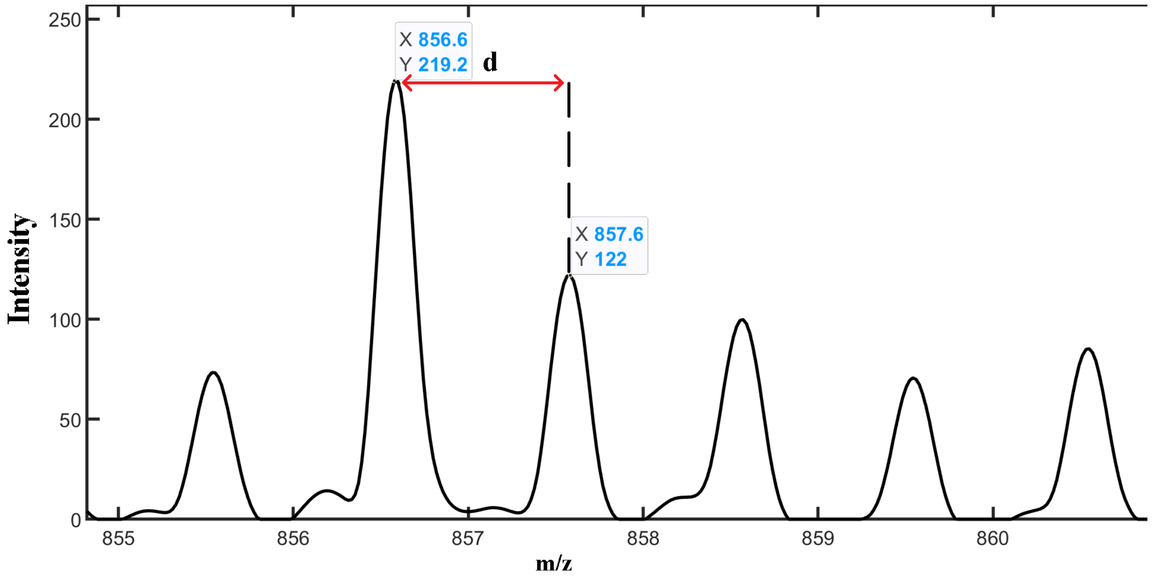
Pipeline: 1st step

Mamdani-Assilan fuzzy-inference system

The distance between two neighbouring peaks is approximately equal to 1.003 Da

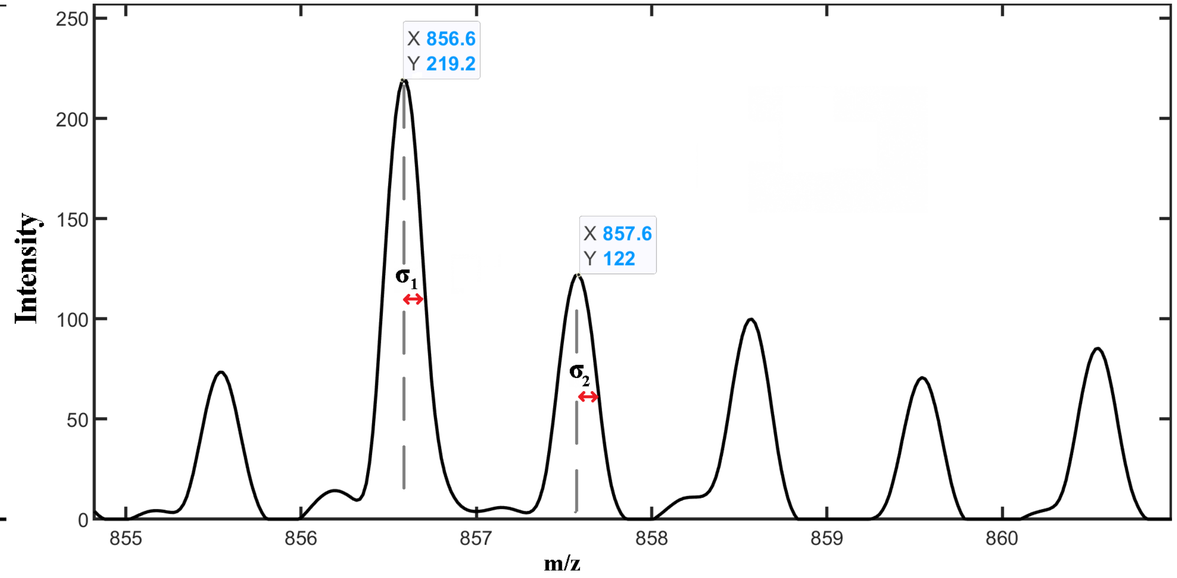
$$\mathrm{d} = \frac{1.003}{z} = 1.003$$

z – ion charge (in MALDI ~= 1.003 Da)



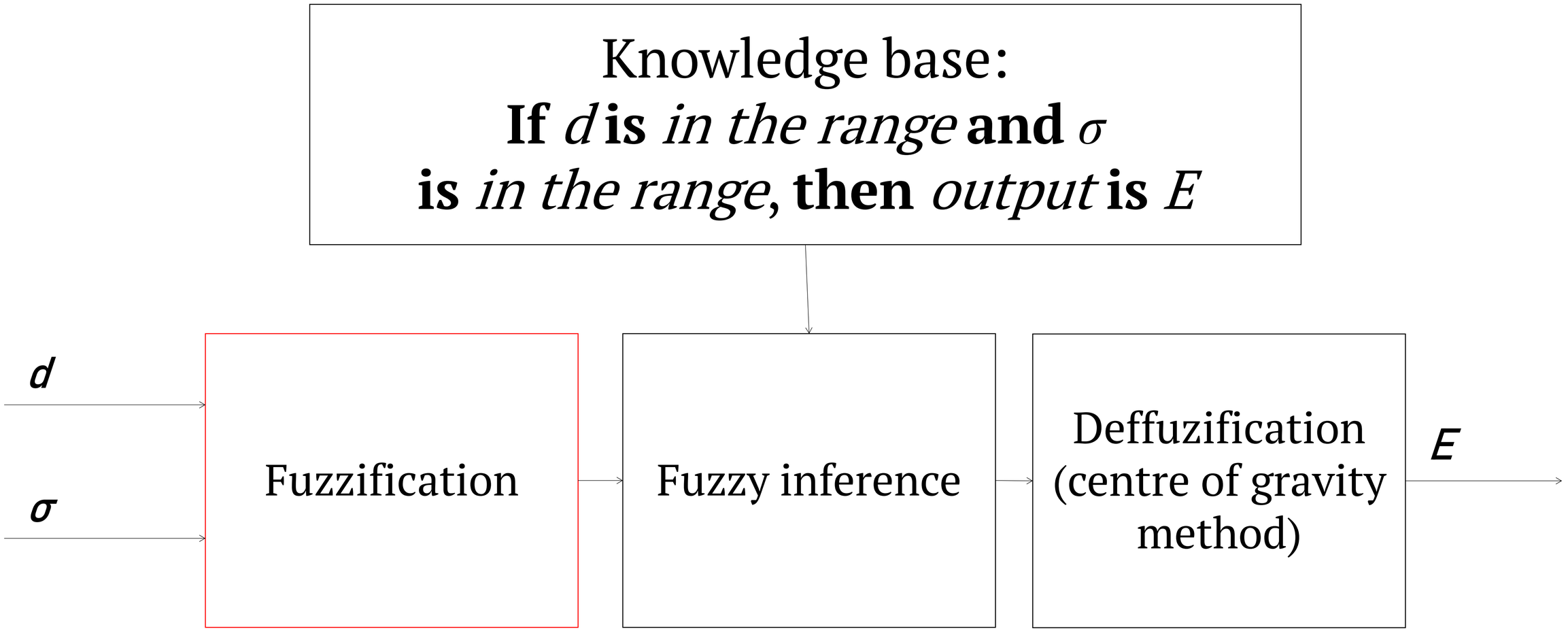
The variance ratio of two neighbouring peaks is approximately equal to 1

$$\sigma = \frac{\sigma_1}{\sigma_2} = 1$$



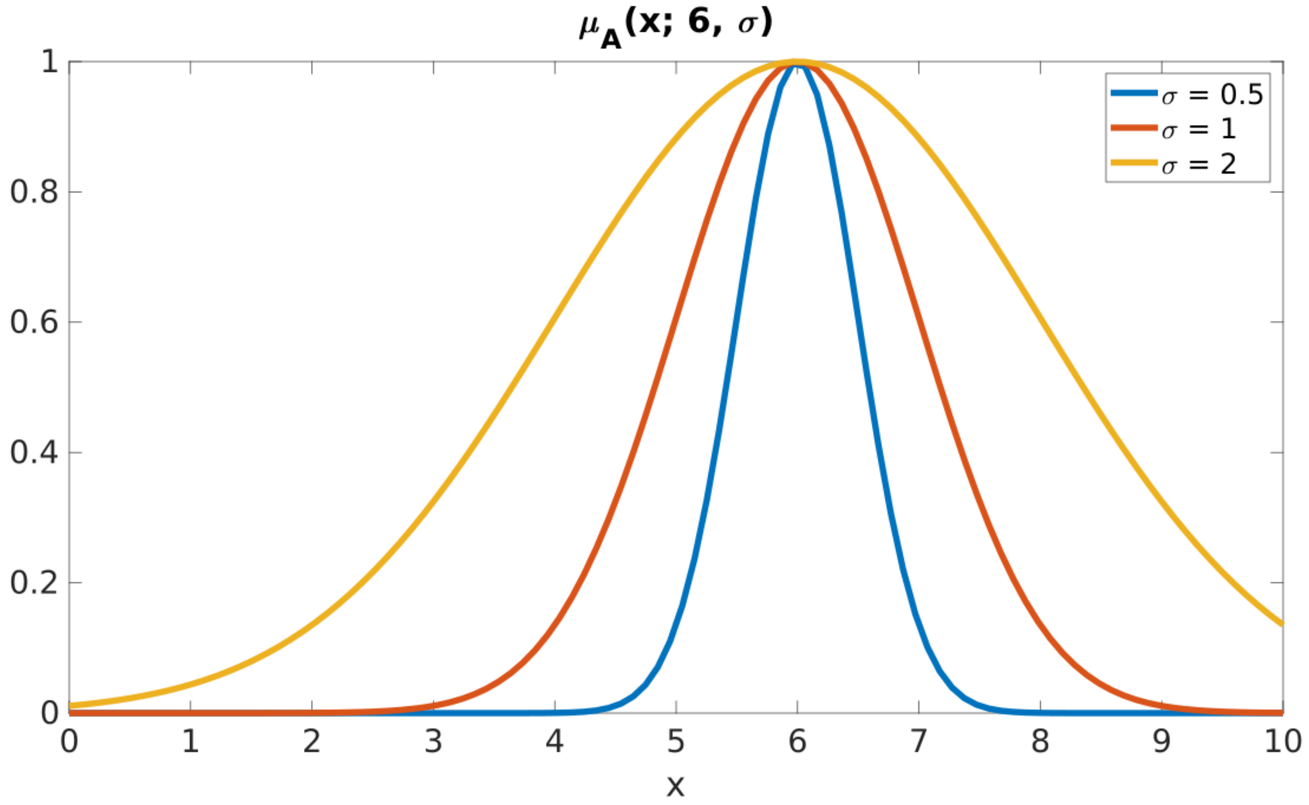
Pipeline: 1st step

Mamdani-Assilan fuzzy-inference system



Pipeline: 1st step

Mamdani-Assilan fuzzy-inference system



Combination of two **Gaussian membership functions**:

$$\mu_A(x; m, \sigma) = e^{\frac{-(x-m)^2}{2\sigma^2}}$$

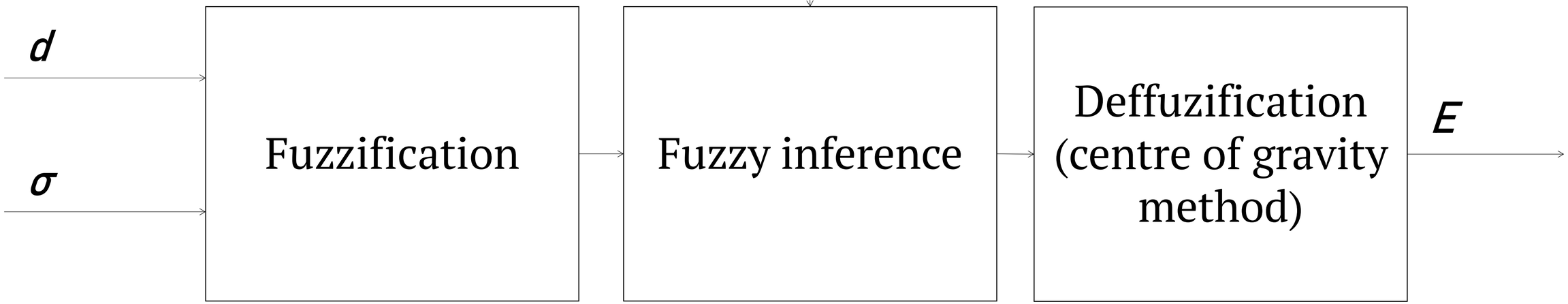
m, σ - parameters

	Input no. 1 (d)	Input no. 2 (σ)	Output
m	0.99 | 1.01	0.99 | 1.01	0.9405 | 1.06
σ	0.0637 | 0.0637	0.02 | 0.1	0.09216 | 0.0871

Pipeline: 1st step

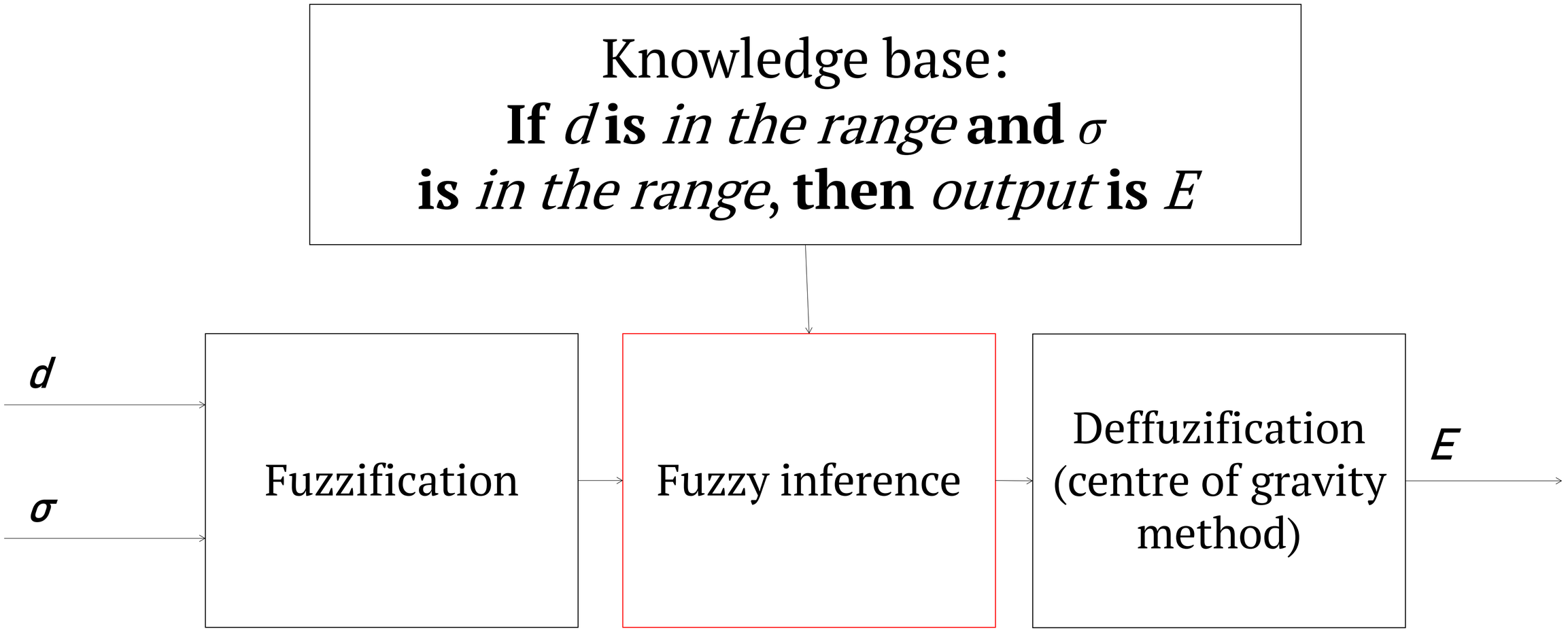
Mamdani-Assilan fuzzy-inference system

Knowledge base:
If *d* **is** *in the range* **and** σ
is *in the range*, **then** *output* **is** *E*



Pipeline: 1st step

Mamdani-Assilan fuzzy-inference system



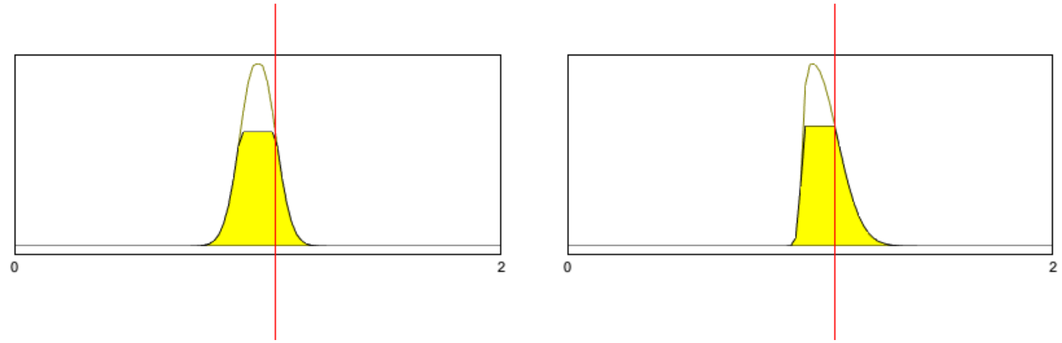
Pipeline: 1st step

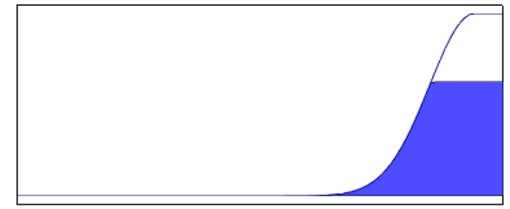
Mamdani-Assilan fuzzy-inference system

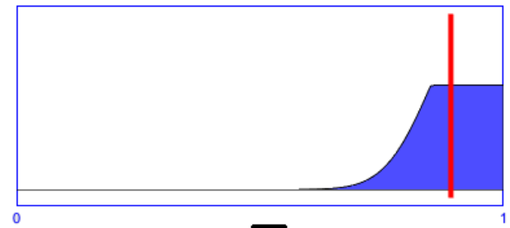
E. H. Mamdani and S. Assilan used the following operations in their system:

- **minimum** as a *t*-norm for **„AND”** modelling
- **maximum** as an *s*-norm for **aggregation** of results for every rule

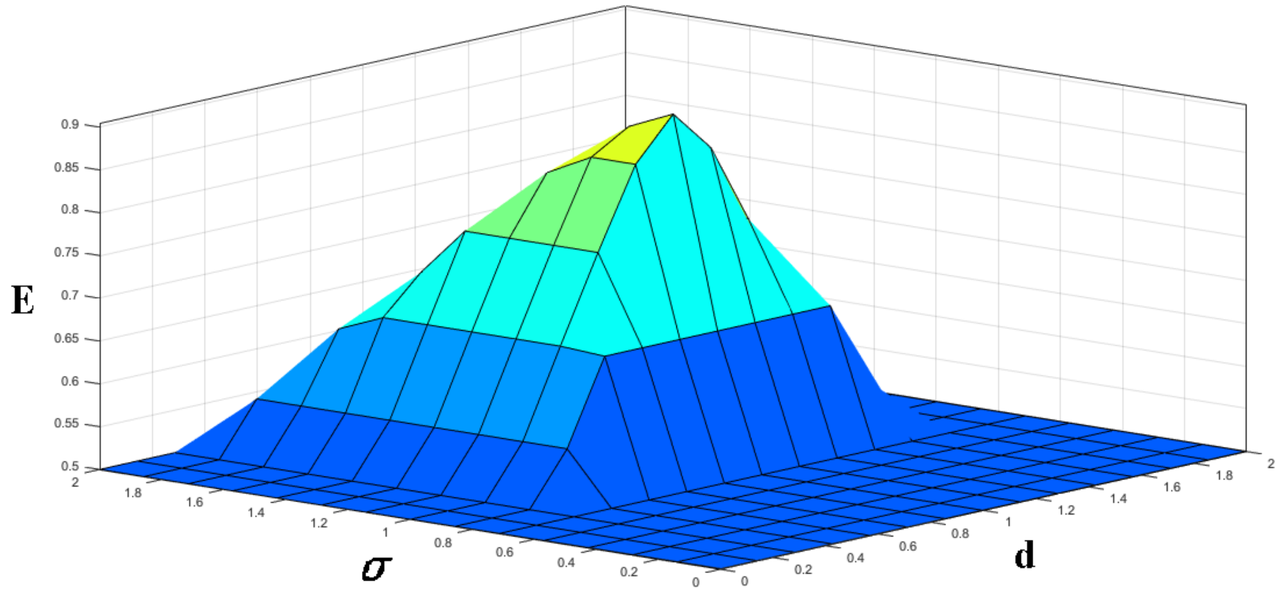
If d **is** in the range **and** σ **is** in the range, **then** the output **is** E.





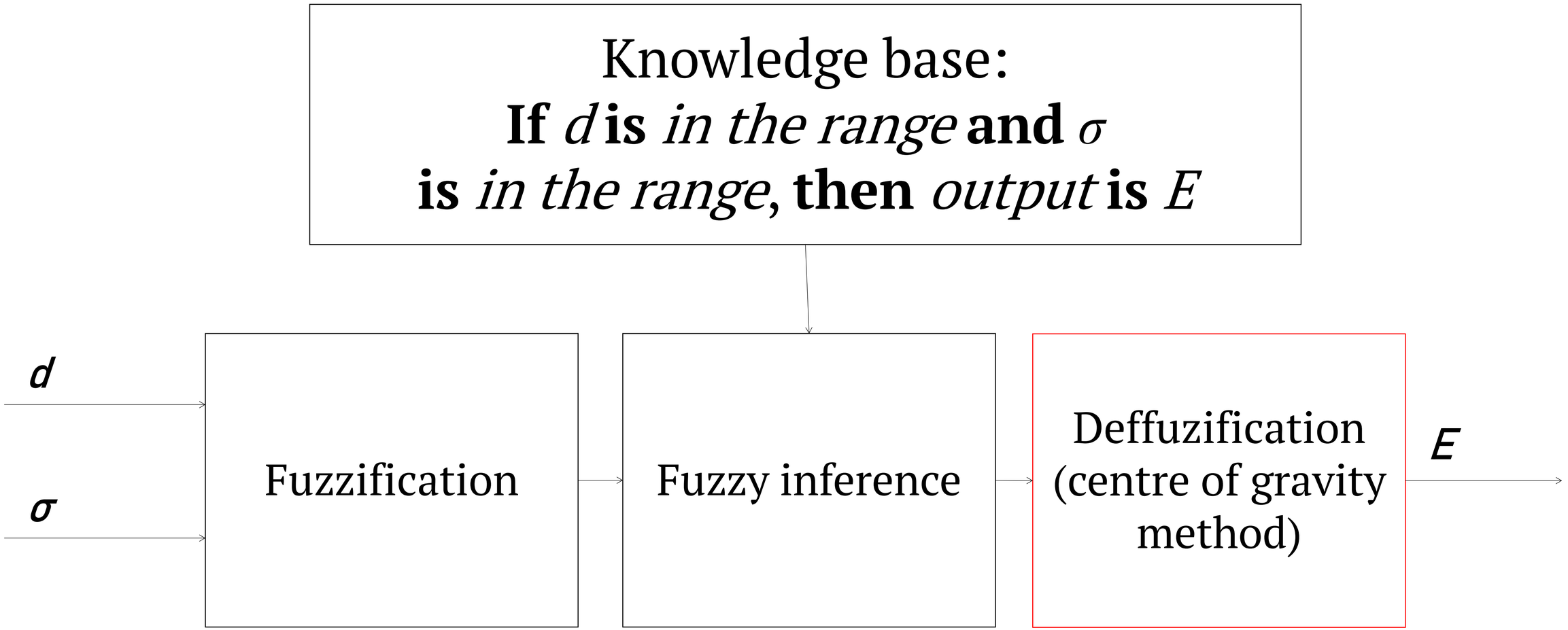


Envelope (E)



Pipeline: 1st step

Mamdani-Assilan fuzzy-inference system



Pipeline: 1st step

Mamdani-Assilan fuzzy-inference system

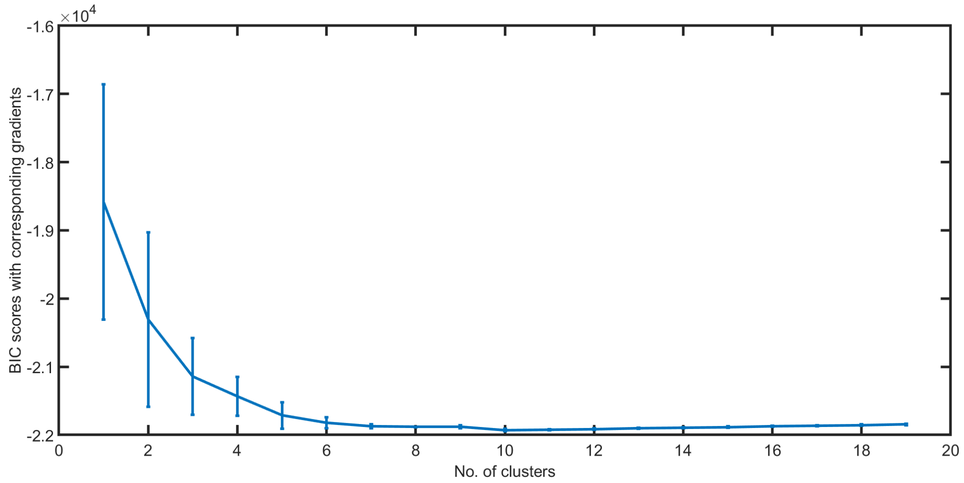
Centre of Gravity method:

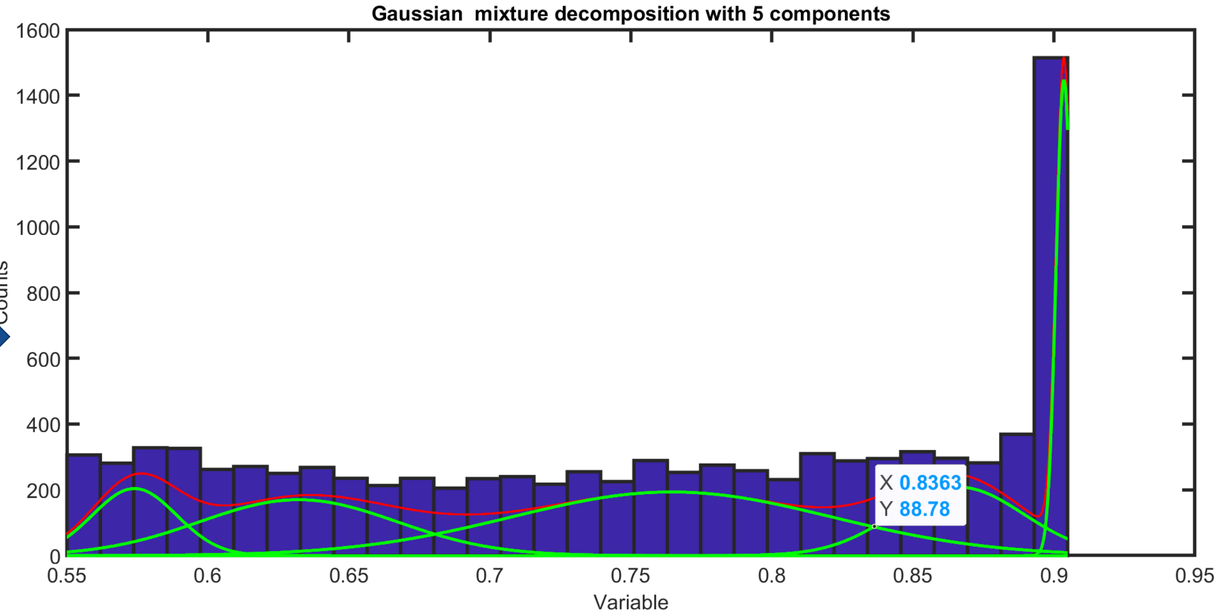
$$y_0 = \frac{\int_{\mathbb{Y}}^{\cdot} y \mu_{B`}(y) dy}{\int_{\mathbb{Y}}^{\cdot} \mu_{B`}(y) dy}$$

Pipeline: 1st step

Mamdani-Assilan fuzzy-inference system

In order to define the threshold, **GMM decomposition** was applied. The number of components was defined by **BIC** criterion.

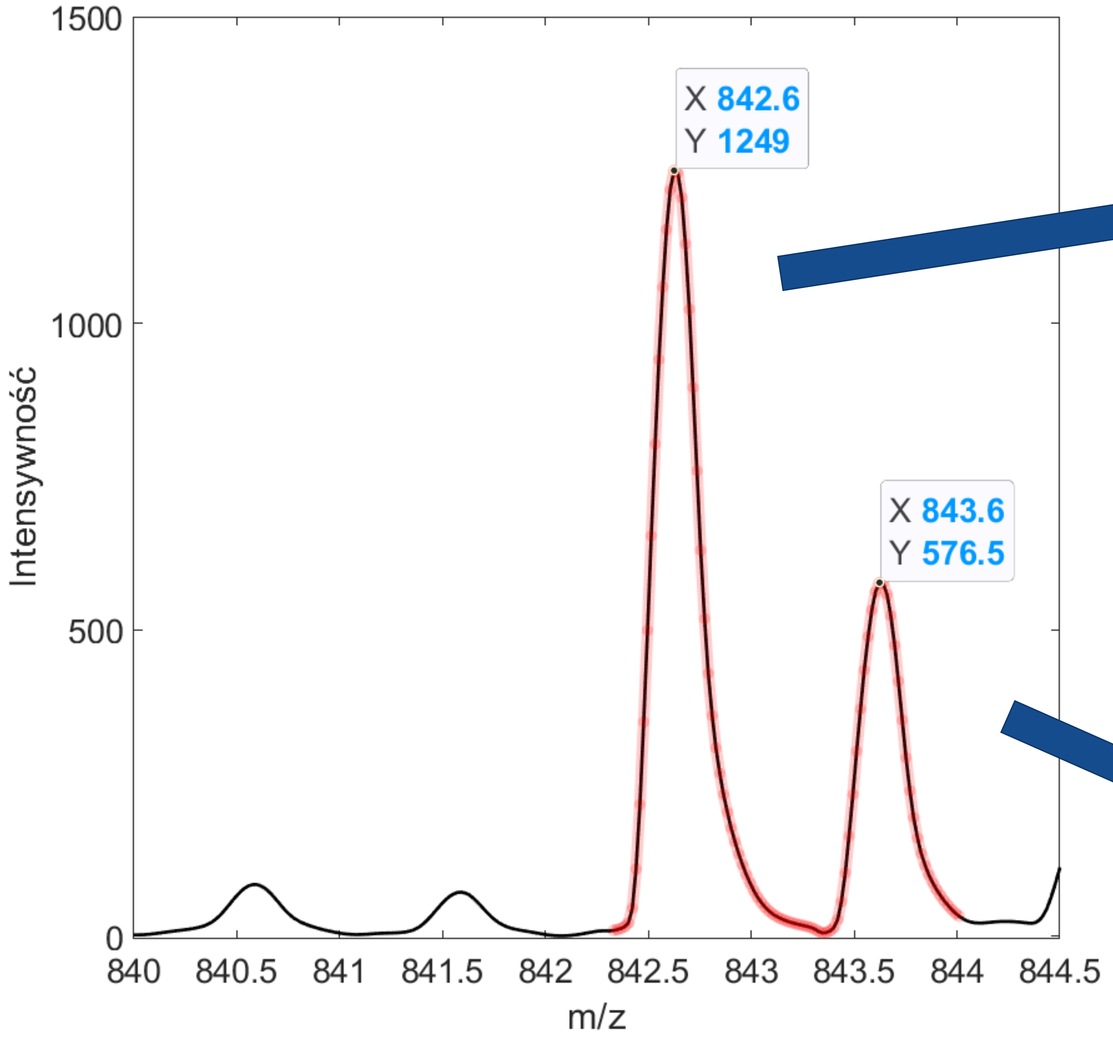


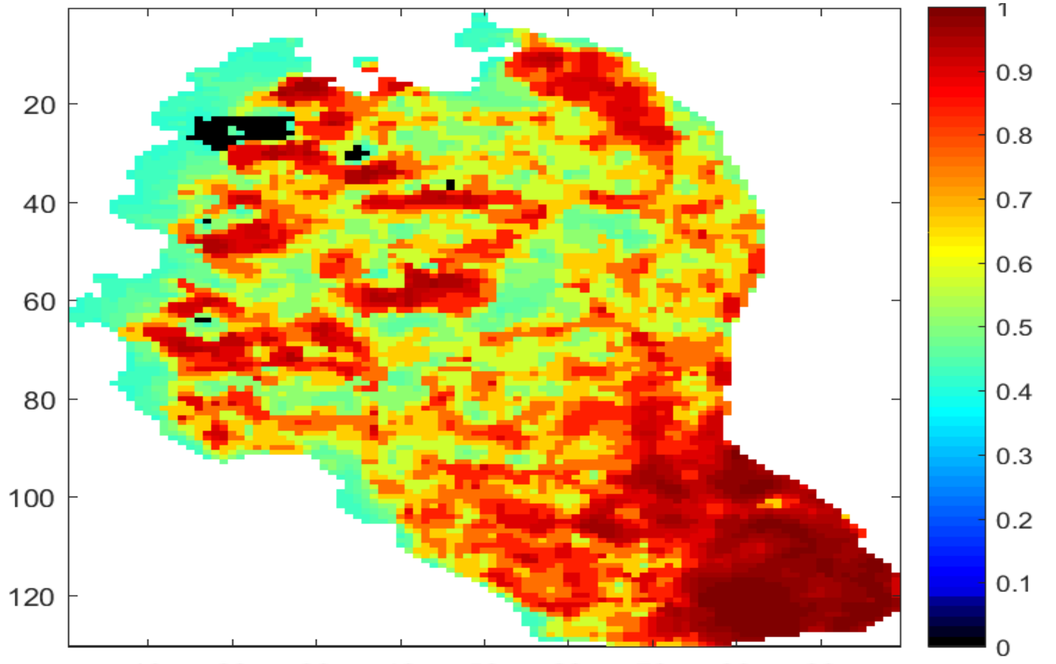


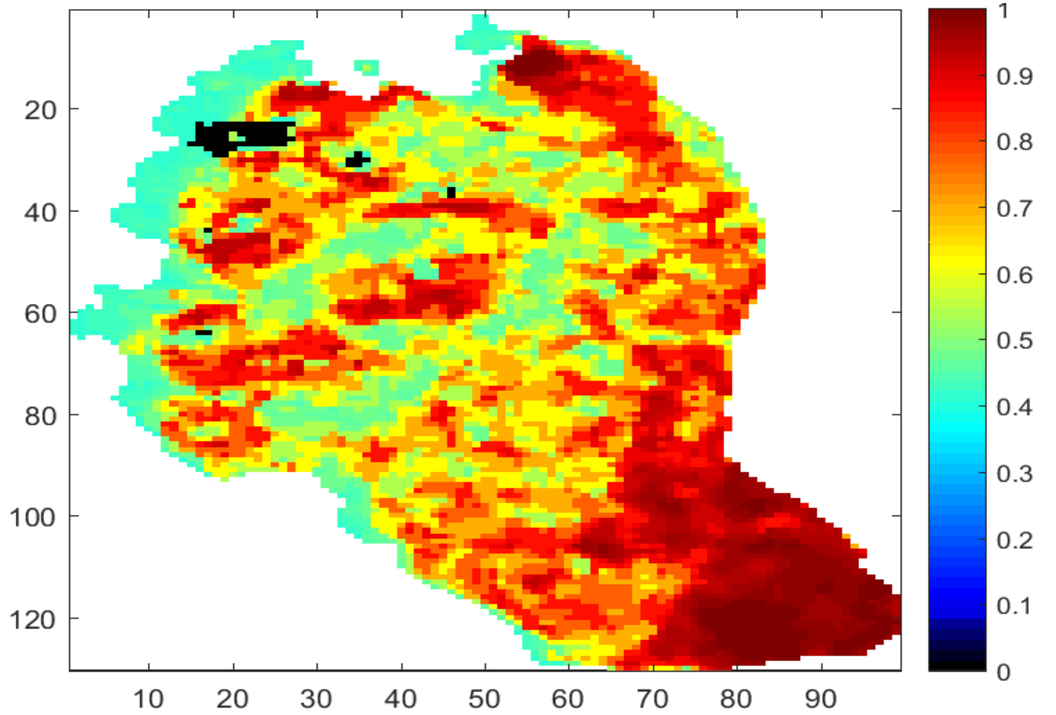
<u>Cutoff</u>: 0.8363

Pipeline: 2nd step

Spatial distribution of peaks

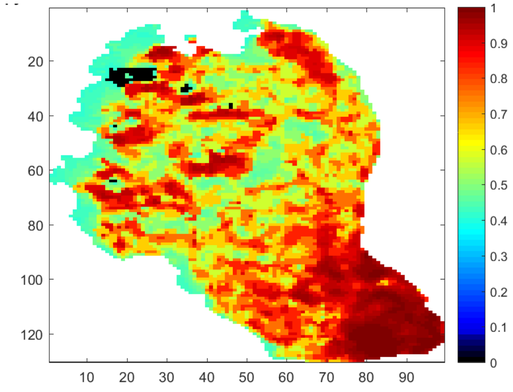
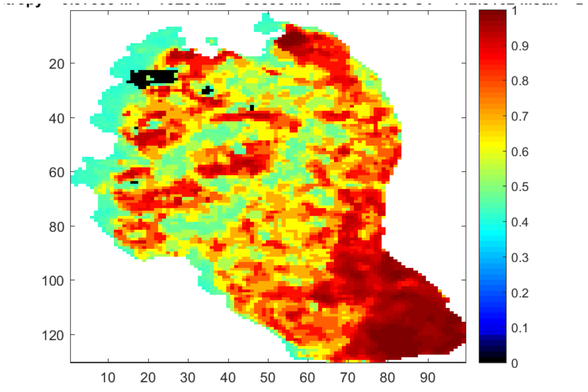
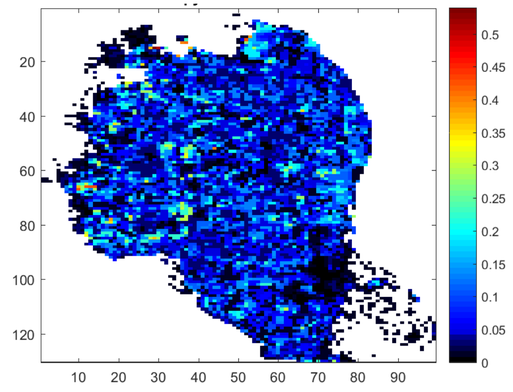
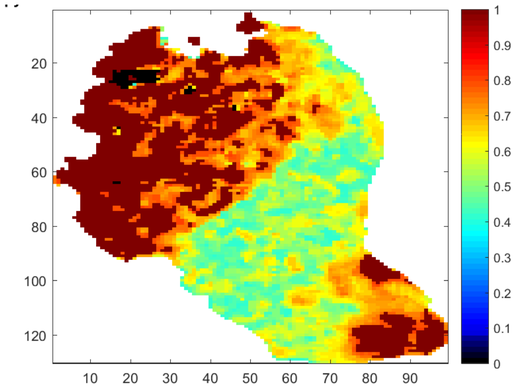
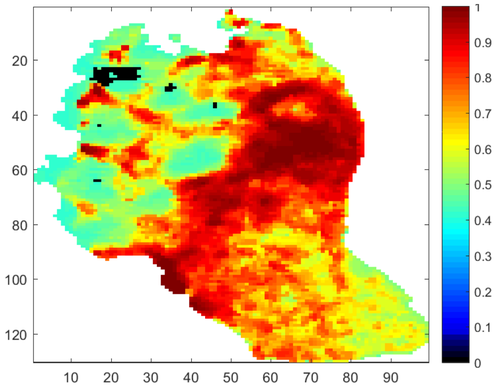
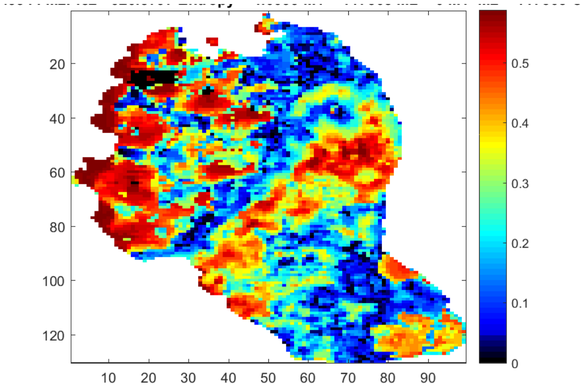






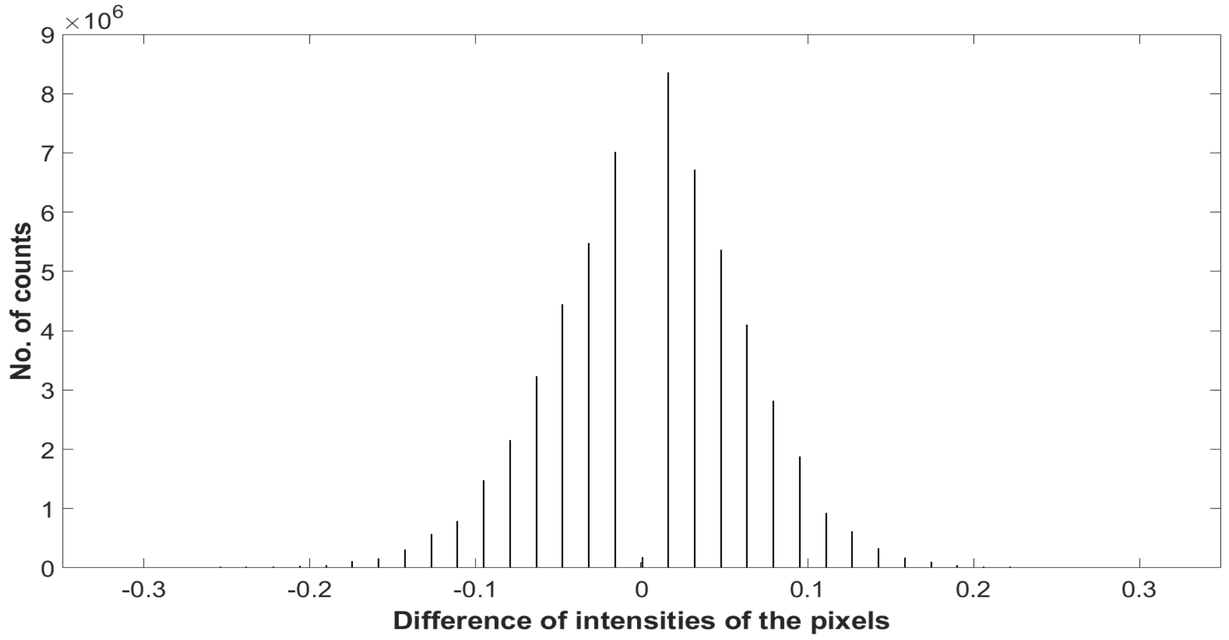
Pipeline: 2nd step

Spatial distribution of peaks – maps of intensities

	Intensity map 1 Peak 1	**Intensity map 2** Peak 2	**Differential intensity map** Peak 1 – Peak 2
E (Envelope)			
nE (non – Envelope)			

Pipeline: 3rd step

Contribution of peaks intensities in isotopic envelope defining



Envelope

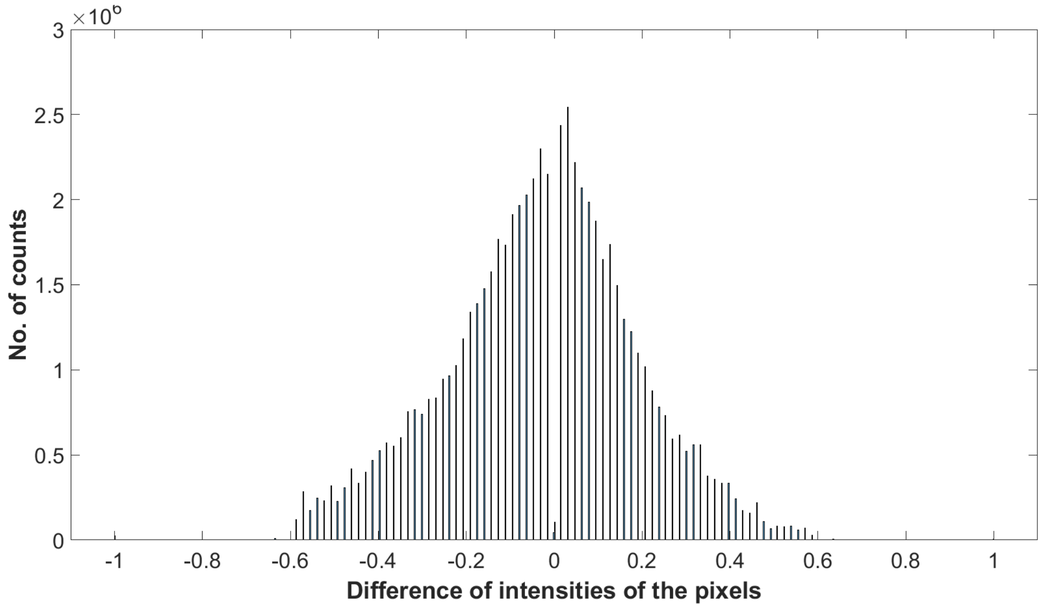
standard deviation of envelope peaks is lower than the standard deviation of peaks that are not members of an isotopic envelope

Based on that,

the **number of peaks included in the range <-0.2 ; 0.2> were calculated**.

Finally, the value [%] is an **input for the fuzzy-inference system.**

the **difference in intensities of the pixels was pairwise calculated**
and as a result,
the **histograms of that difference** were created:

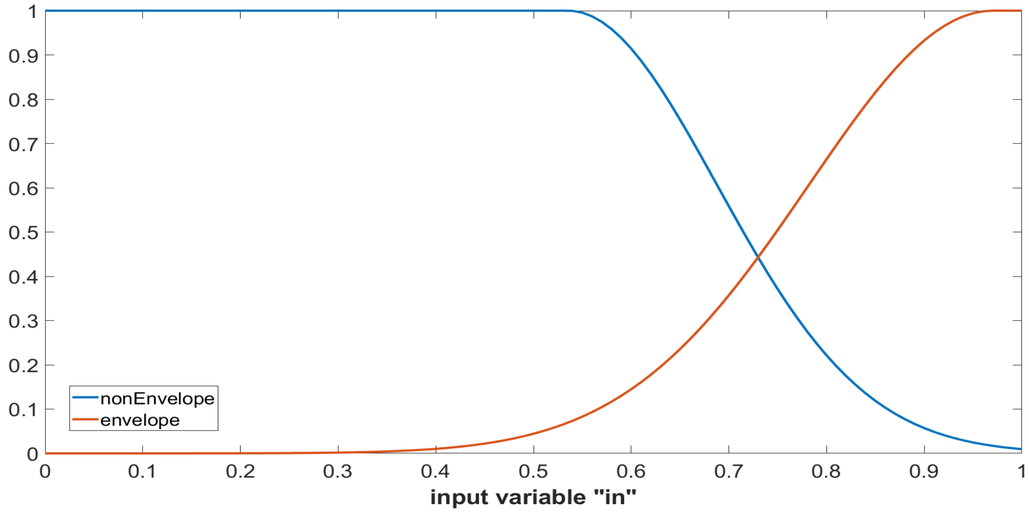


non-Envelope

Pipeline: 3rd step

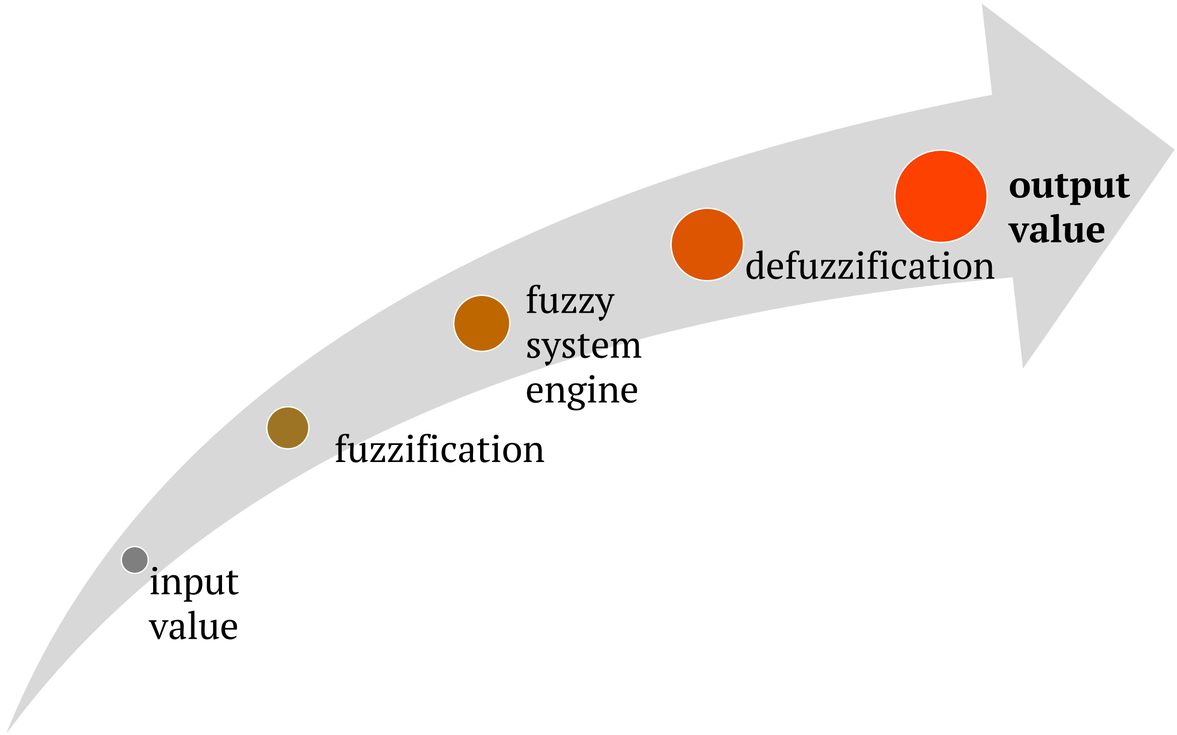
Decision making process based on Sugeno fuzzy-inference system

Input value → **Fuzzy-inference system** → Output value: **possibility of being a member of an isotopic envelope**



$$f(x;\ \sigma, c) = e^{\frac{-(x-c)^2}{2\sigma^2}},$$

where:
σ – standard deviation,
c – mean for each Gaussian function.



Results

Sugeno fuzzy-inference system

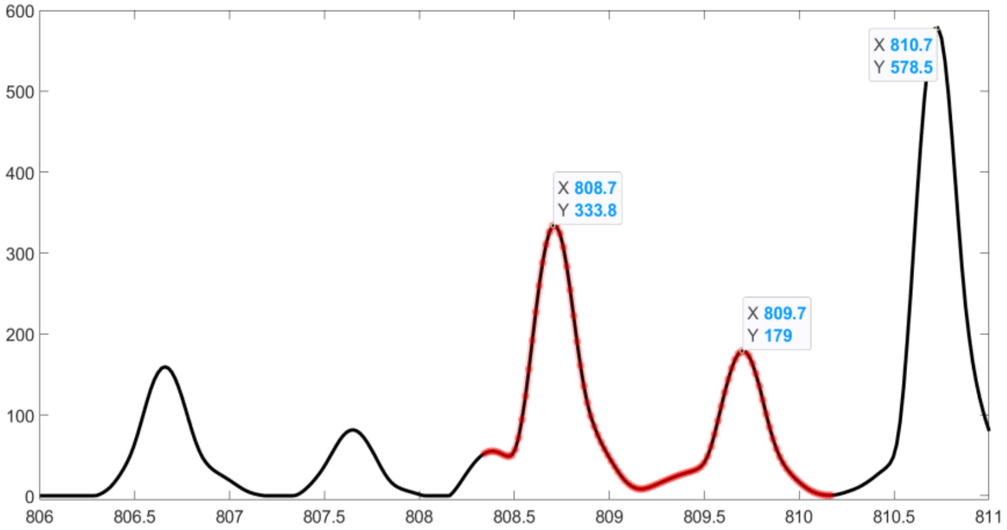
m/z_1	m/z_2	Possibility of isotopic envelope membership [%]
805.6	809.7	46 (Non-envelope)
808.7	809.7	74.7 (Envelope)
810.7	811.7	98.1 (Envelope)
810.8	897.6	15.3 (Non-envelope)
812.7	813.7	98.7 (Envelope)
812.7	897.6	25.1 (Non-envelope)
843.7	844.7	99 (Envelope)

- Members of an isotopic envelope are characterised by **possibility values bigger than 50%.**

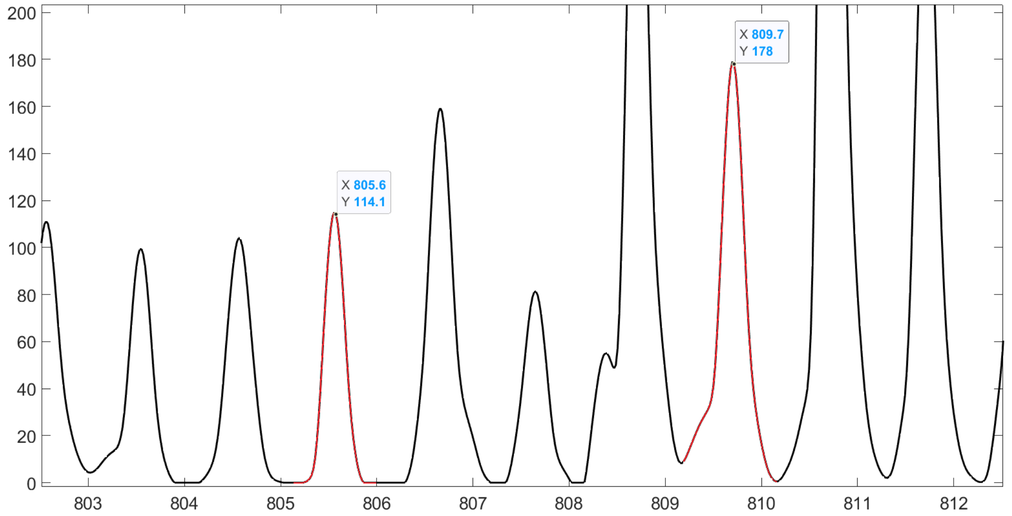
- Isotopic envelope members are characterized by the **lower number of peaks within the range <-0.2; 0.2>**
  Reason: peaks of one isotopic envelope in such a range of m/z values (~800 – ~1000 Da) follow such a pattern:
  the first peak has the highest intensity (monoisotopic peak), whereas the successive peaks represent ~45% and ~12% of the intensity of the first peak, respectively. According to that, the **intensity histogram of peaks included in one envelope is denser within the range <-0.2 ; 0.2>.**

Results

Sugeno fuzzy-inference system



Envelope

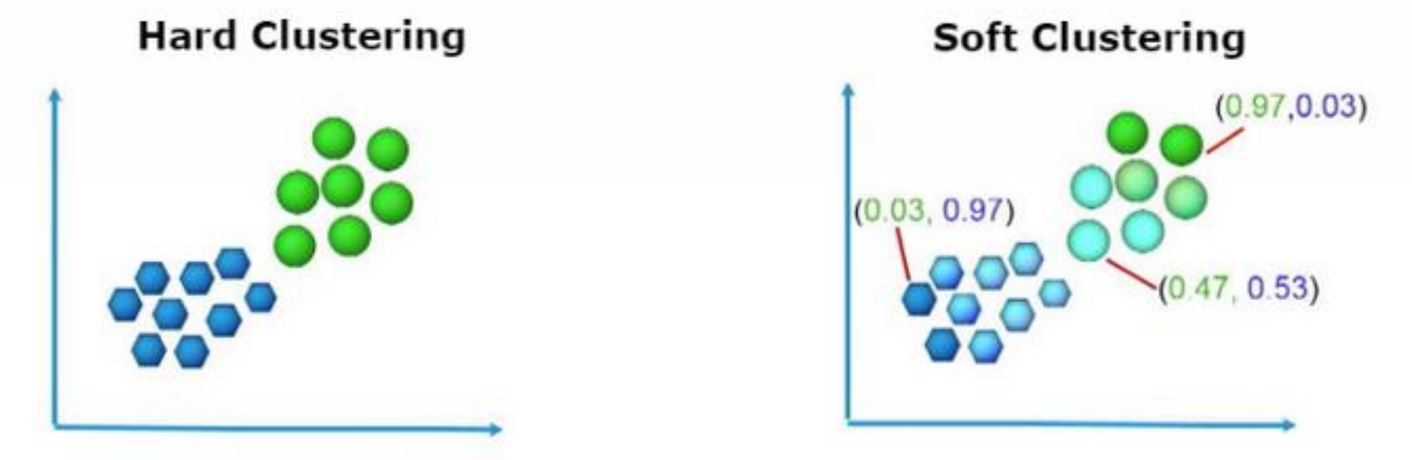


non-Envelope

The obtained results were compared to results of an analysis of an average MSI spectrum performed by **an experienced mass spectrometrist, who assessed whether a particular isotopic peak belonged to a given isotopic envelope based on the theoretical isotope pattern for a peptide with a given mass.** The theoretical isotopic pattern for a peptide was obtained using the *Compass IsotopePattern Calculator* (Bruker®) taking into account the peptide sequence obtained in an **LC-MALDI MSMS analysis of the tissue protein extract.**

Different approach
Fuzzy C-means clustering approach

- Data point can belong to **two or more clusters**
- **Soft Clustering method:** every data point **can belong to every cluster with a certain degree**: likelihood or probability score
- **Fuzzy C-means segmentation** was performed by converting an input differential image into two segments by the fuzzy C-means algorithm



Source: https://towardsdatascience.com/fuzzy-c-means-clustering-is-it-better-than-k-means-clustering-448a0aba1ee7

Results

Fuzzy C-means clustering approach

Final segmentation after fuzzy C-means clustering

Envelope:
no structure visible

non-Envelope:
clear structure visible

Envelope		Non-envelope	
Cluster center 1	Cluster center 2	Cluster center 1	Cluster center 2
2.6	32.4	7.0	71.7
2.1	23.9	7.9	86.1
2.2	23.6	7.7	81.0

cluster center:
arithmetic mean of all the data points that belong to the specific cluster

Envelope peaks are characterized by significantly **lower values** in comparison to the non-envelope ones

Summary

In this work we proposed an algorithm **for automatic identification of an isotopic envelope.**

There are a plethora of algorithms for deisotoping, but they are usually dedicated to a specific type of experimental platform (e.g. MS-Deconv, BPDA) or type of a molecule (lipids or peptides, e.g. YADA, BPDA).

The presented method can be used **for each kind of mass spectrum**, no matter what type of mass spectrometry experiment it comes from, and **various types of molecules,** as it takes into consideration only one aspect of a mass spectrum: **spatial distribution of the peaks.**

Limitations: the proposed method is **dedicated only to molecular imaging techniques** and cannot be used in other proteome studies.

Acknowledgements

Joanna Polańska, Professor, Silesian University of Technology, Poland

Marta Gawin, PhD, Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology Gliwice Branch

Jacek Łęski, Professor, Silesian University of Technology, Poland

Katarzyna Frątczak, MSc, Silesian University of Technology, Poland

This work was co-financed by European Union grant under the **European Social Fund, project no. POWR.03.02.00-00-I029.**

Thank you for your attention