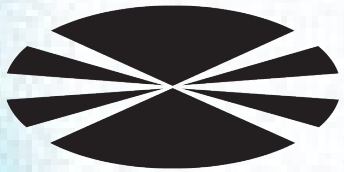
Gaia data analysis services on the SPACIOUS platform: GUASOM

Marco A. Álvarez



UNIVERSIDADE DA CORUÑA

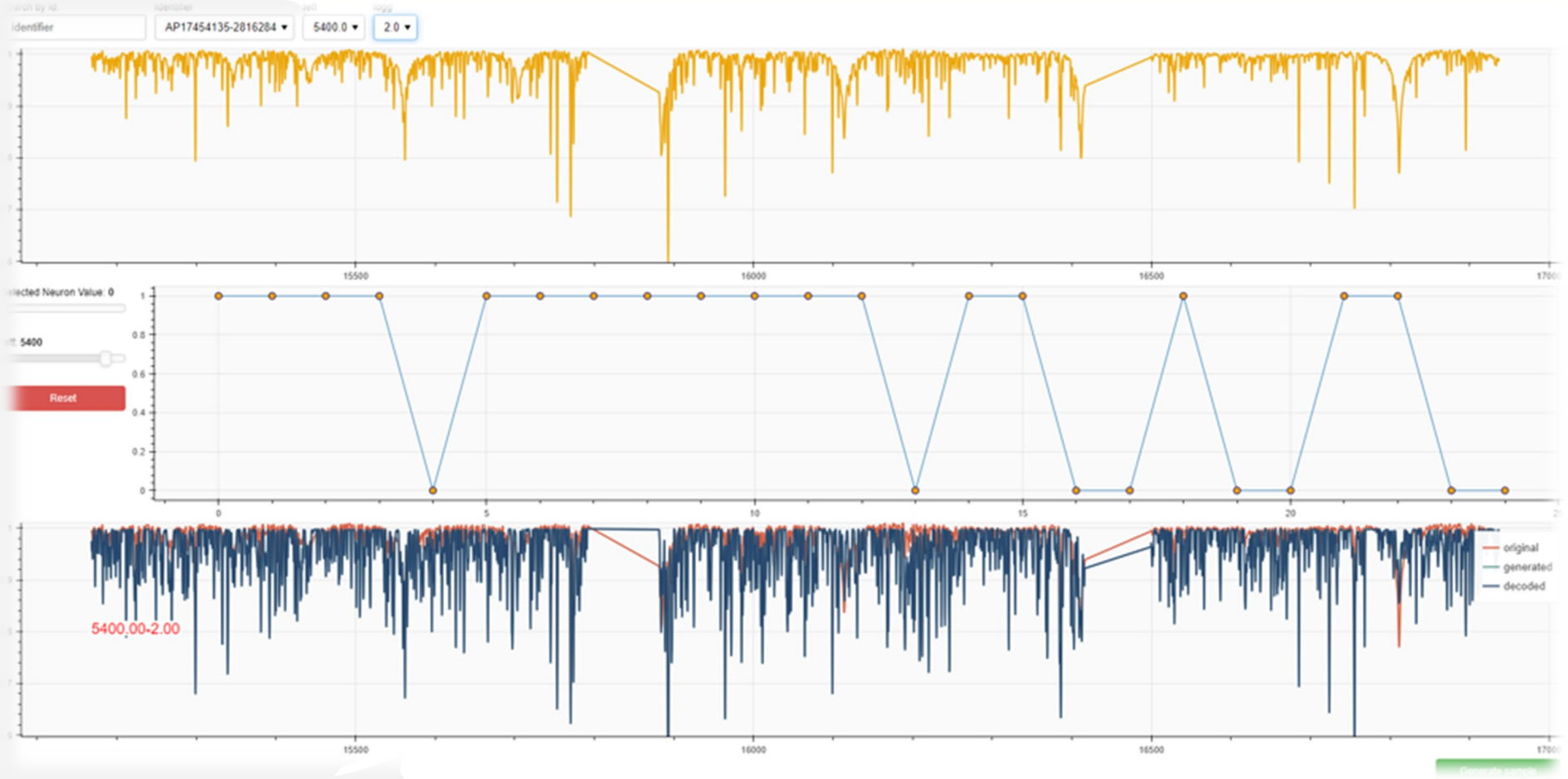
SPACIOUS

WP52 – Data Mining Tools and specific techniques

Domain specific techniques to perform a satisfactory analysis of the data.

GANDALF	Defining, training, testing and visualising disentangling models
SOM train	Efficient training Self-Organizing Maps (SOM)
GUASOM	Adaptive visualisation tool for SOMs

GANDALF



Framework to define, train and test disentangling models.

GANDALF

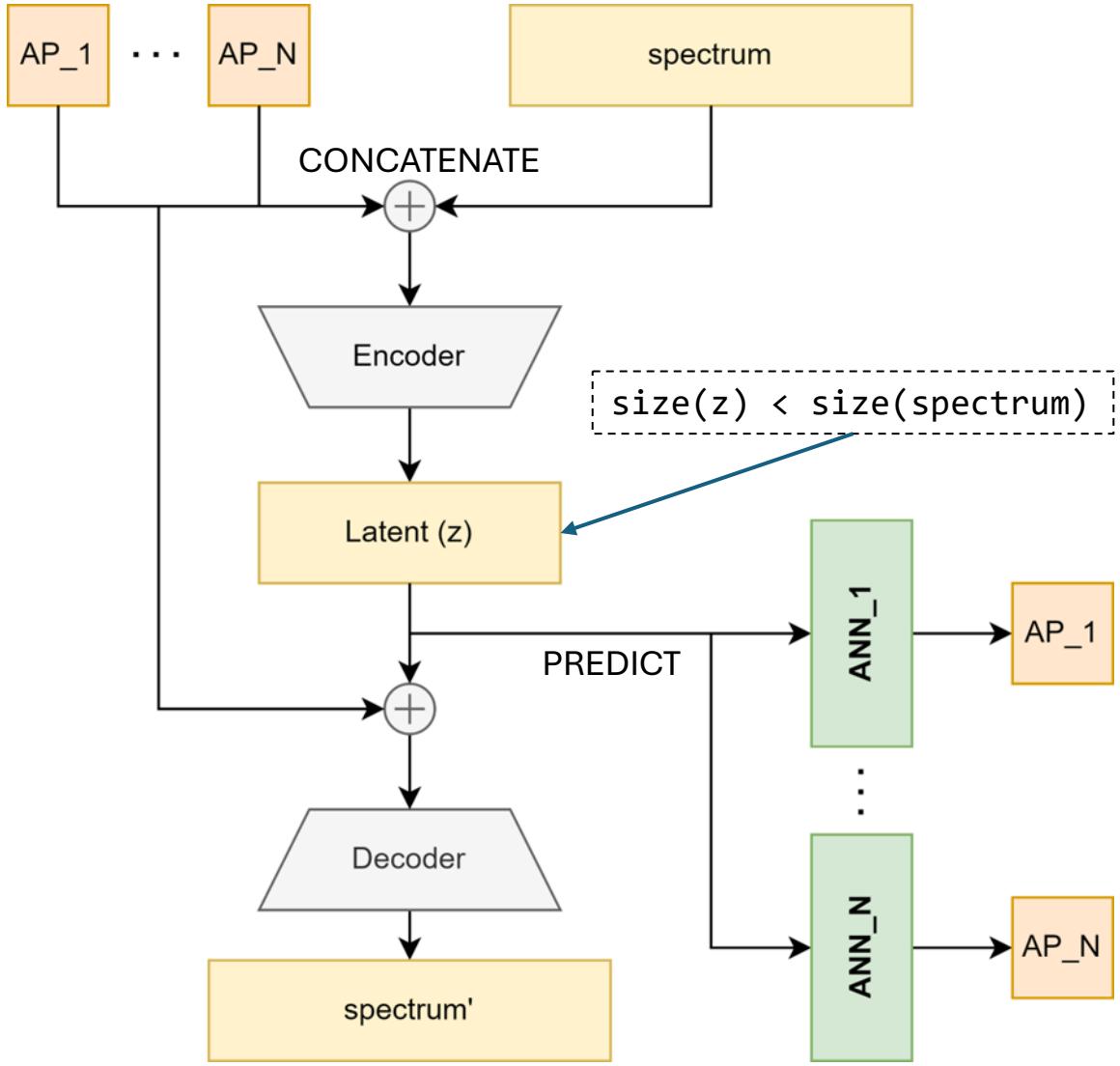
Combination of *autoencoders* and *adversarial training*.

- Remove parameters from the input data and create a new **disentangled representation**.
- **Generate** samples with new parameters.

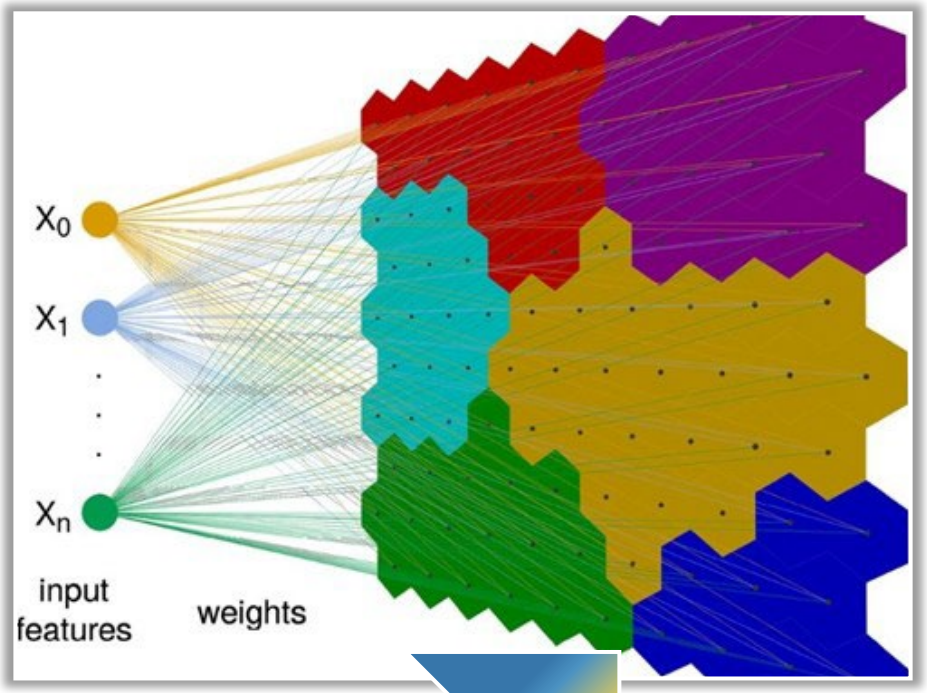
Autoencoder learns to reconstruct the original spectrum.

Discriminators (ANN) attempt to predict the parameters → their success acts as a penalisation for the autoencoder (adversarial).

The latent excludes the information from the parameters that want to be removed.



SOM Train



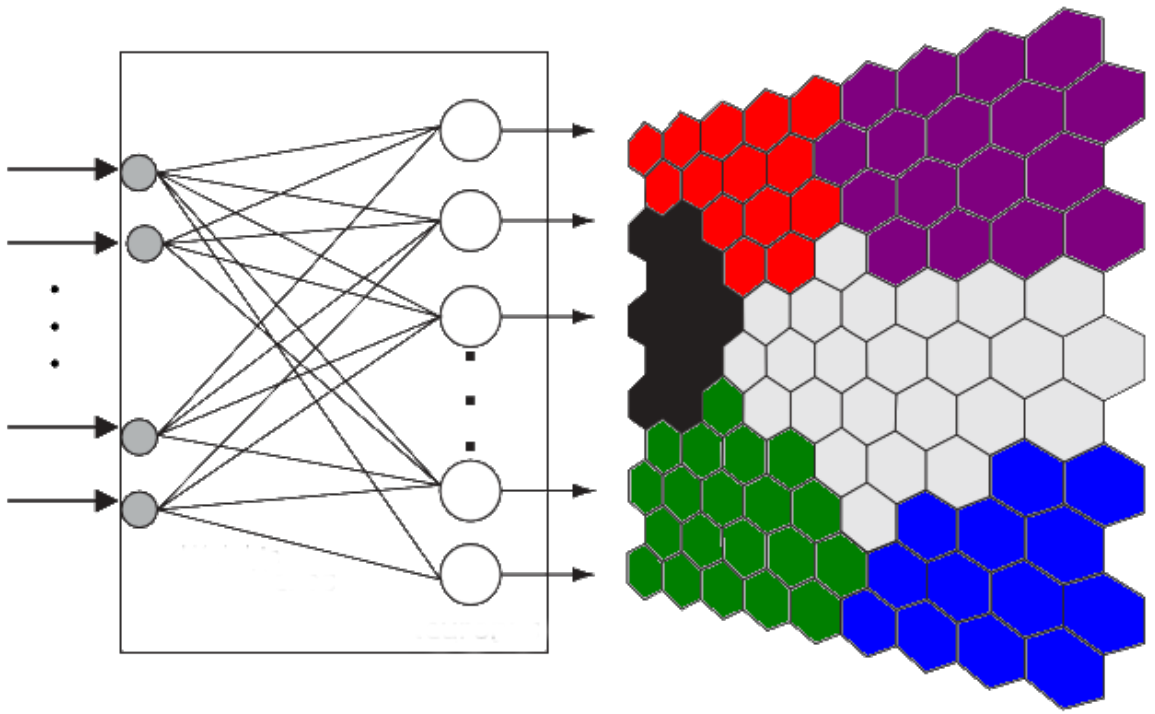
Python library designed to simplify the process of training, visualising, and evaluating Self-Organising Maps (SOMs)

SOM

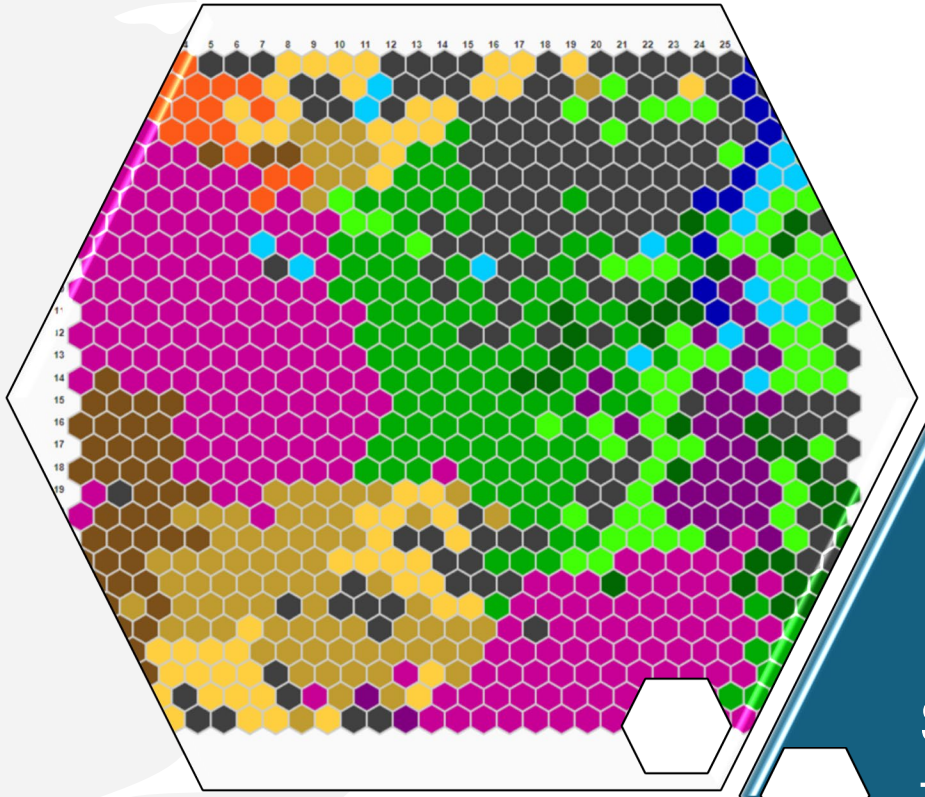
Unsupervised learning technique used to **project high-dimensional data onto a low-dimensional grid** while preserving the data's topological structure.

Each grid cell (neuron) has a **weight vector** (prototype) with the same dimension as the input.

For every data point, the SOM finds the neuron whose weights are most similar (**Best Matching Unit**) and then **updates the BMU and its neighbouring neurons** to move closer to that input.



GUASOM



spring®

Specific purpose tool for visualising Self-Organizing Maps.

GUASOM

GUASOM is a web visualisation tool built on a client-server model using Angular and Spring.

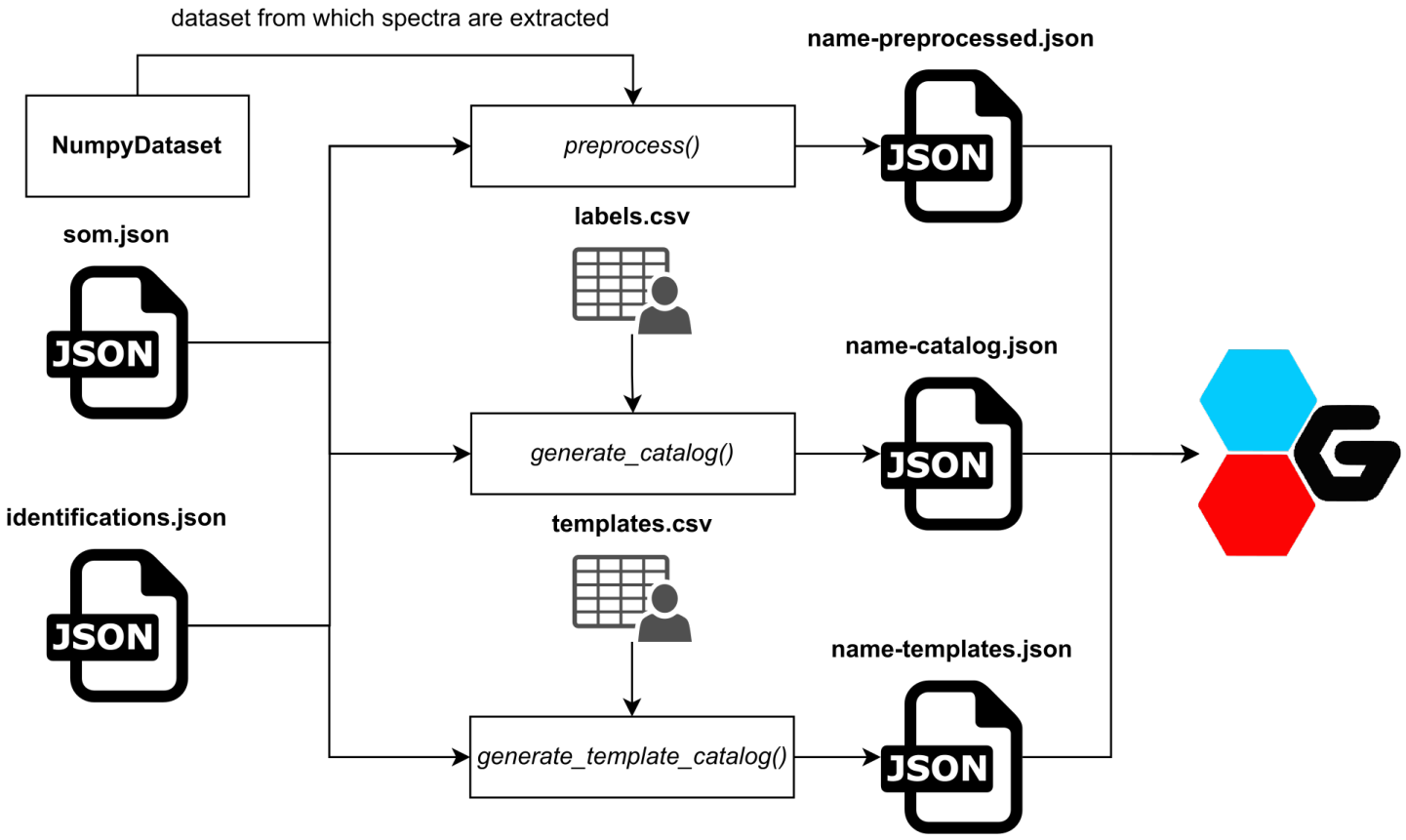
Users can deploy their maps trained with the *SOM Train* library for analysis through multiple visualisations.

The main advantage of this application is its ability to analyse large volumes of data with minimal time and resource consumption.

GUASOM Preprocess

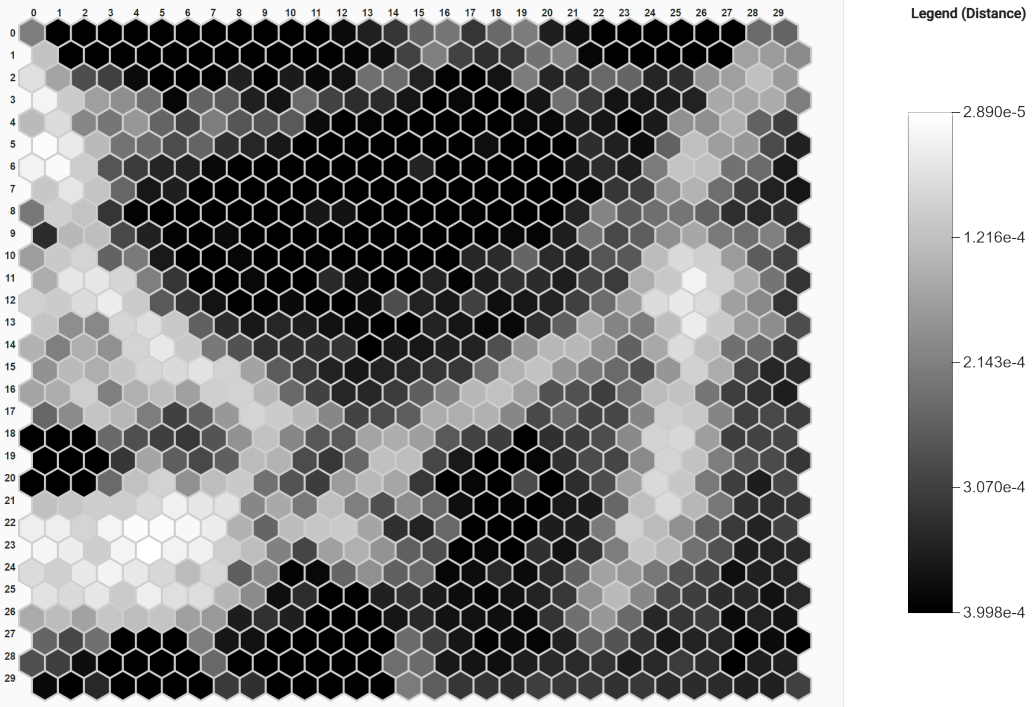
The user can add more information of interest to the objects represented on the map.

Using the Python GUASOM Preprocess library, the user can prepare the visualisations that will then be rendered in GUASOM.

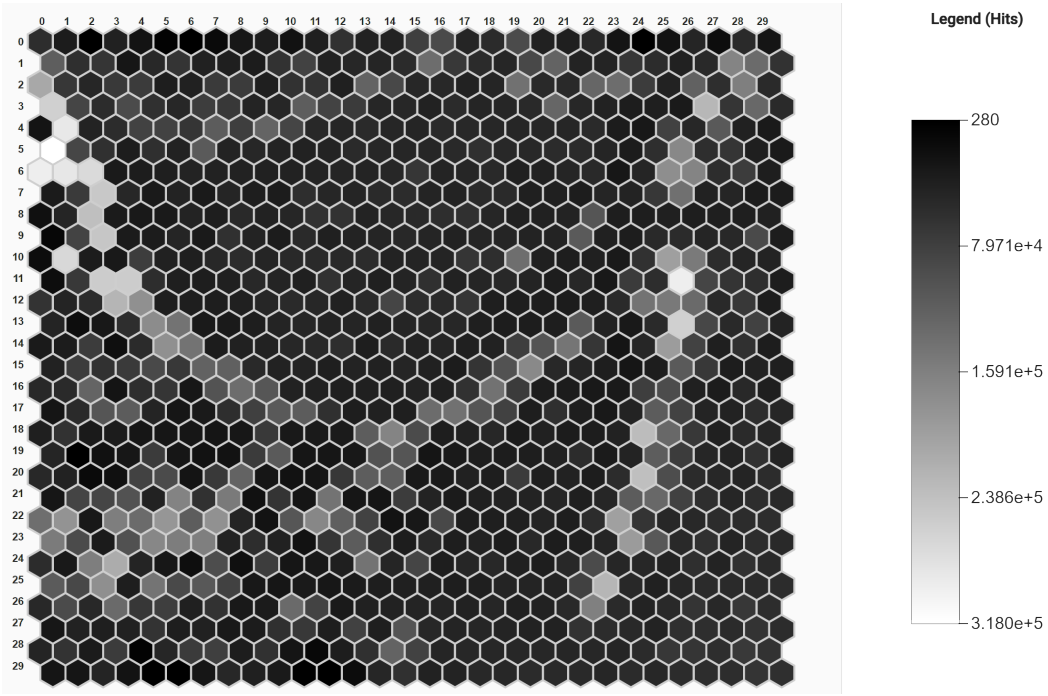


GUASOM Visualisations

Explore the inner structure of the SOM through classic visualisations, like U-Matrix or Hits.



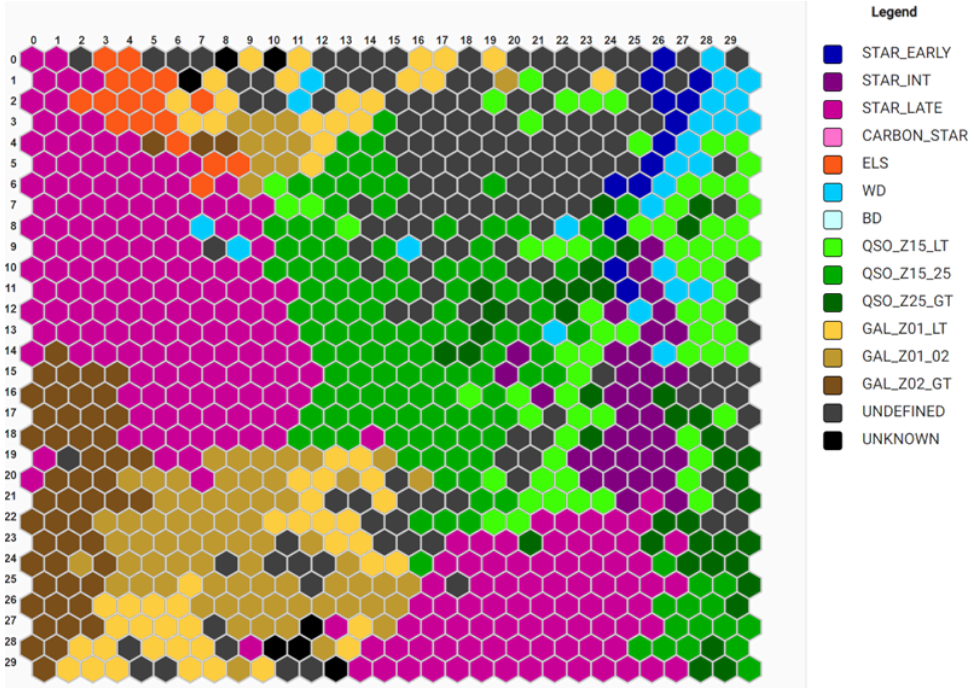
U-Matrix: It shows the distance between the neurons. As higher the distance, lesser the similarity between the prototypes of the neurons.



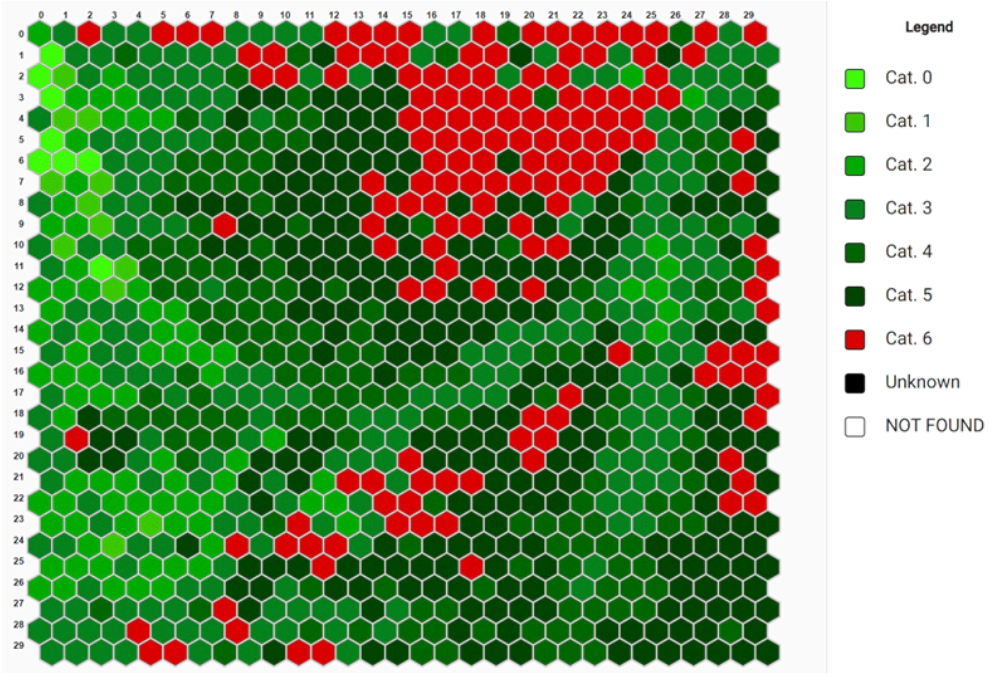
Hits: It shows a density map of the objects that populate the neurons.

GUASOM Visualisations

Create domain-specific graphs to better understand the nature and characteristics of the data.



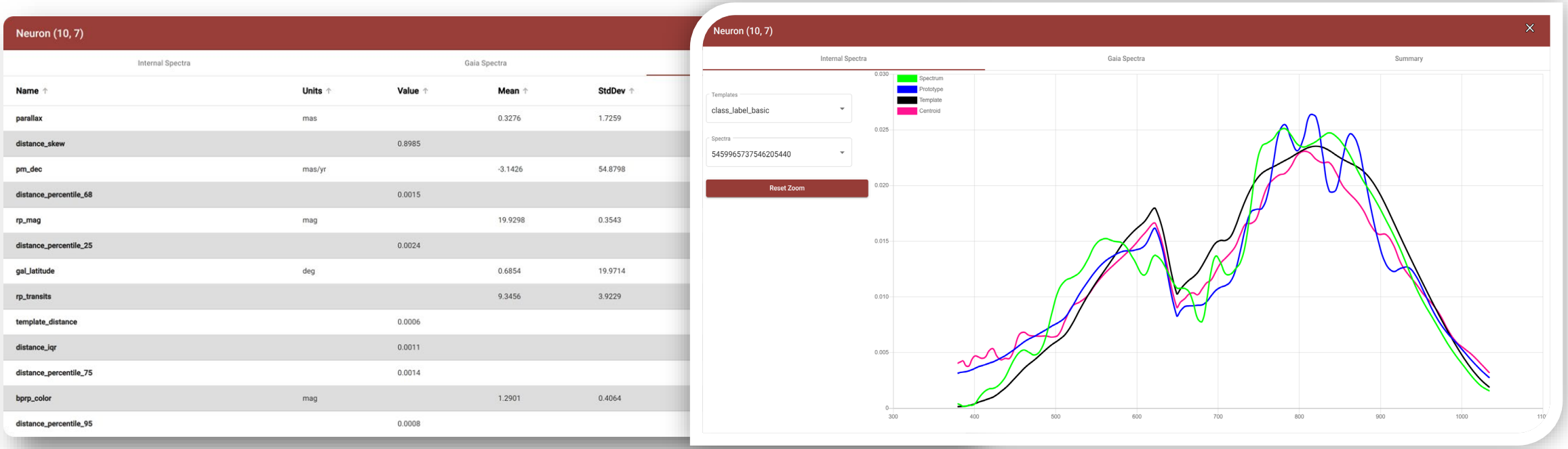
Templates: Identify the **class of the objects** that belong to each neuron using a reference dataset that represents the different astronomical classes (templates).



Catalogue: It shows the representative class of each neuron using data obtained from catalogue cross-referencing, and you can even visualise the quality of the neurons after applying statistical techniques.

GUASOM Visualisations

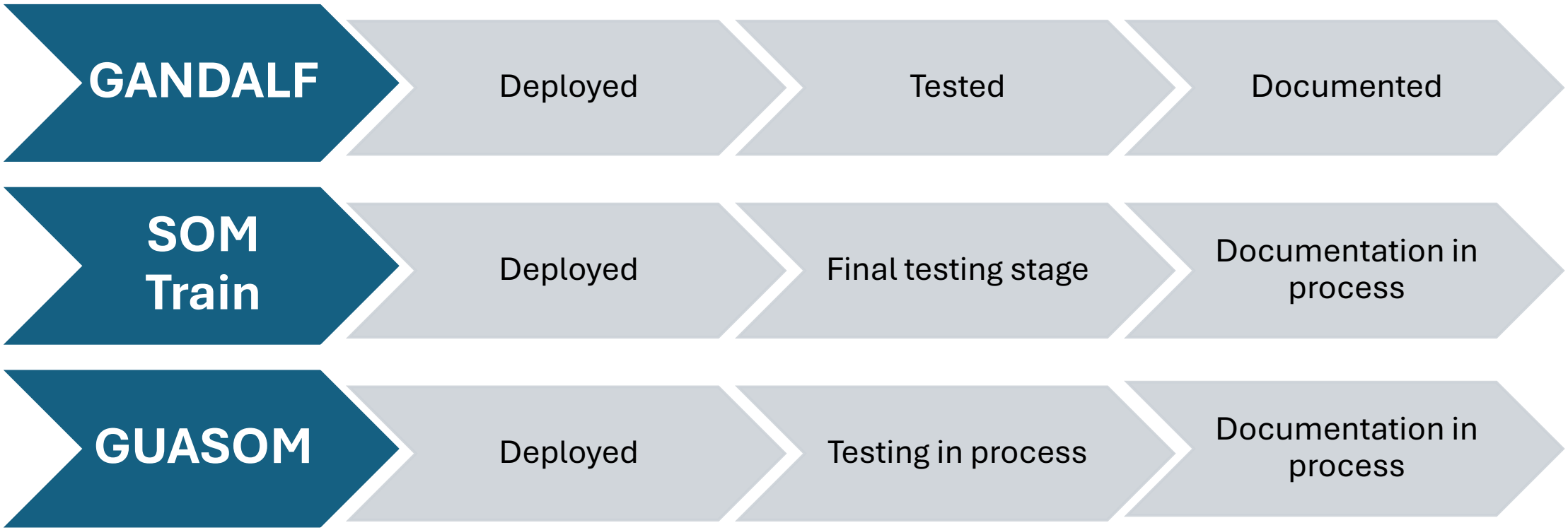
Analyse the content of each neuron and its belonging objects.



Parameters: It offers the statistical report of your favourite parameters used to characterise the objects represented in the neuron.

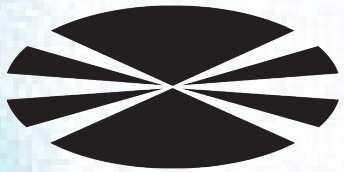
Features: Displays the feature vector of the neuron's objects, the prototype of that neuron, the centroid, and the template that best matches (if it exists).

Status


GANDALF
Deployed
Tested
Documented
SOM Train
Deployed
Final testing stage
Documentation in process
GUASOM
Deployed
Testing in process
Documentation in process

Gaia data analysis services on the SPACIOUS platform: GUASOM

Marco A. Álvarez



UNIVERSIDADE DA CORUÑA

SPACIOUS