

---

# Dimensionality Reduction

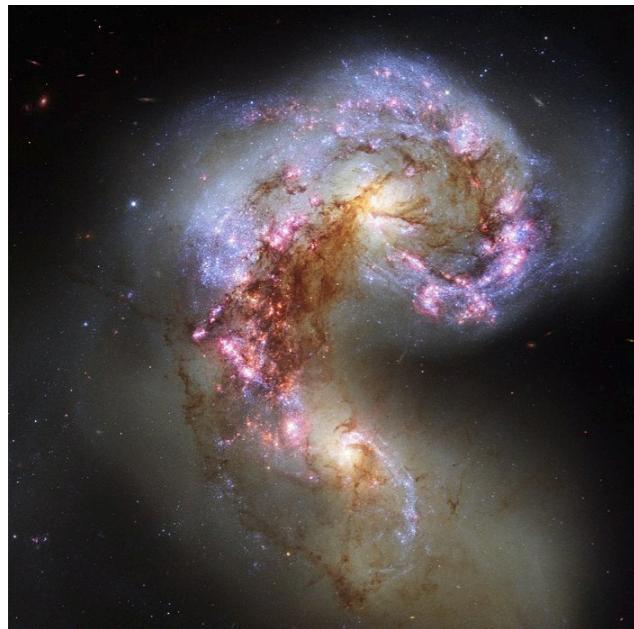
## Part 3: The Sequencer

Dalya Baron  
Carnegie Observatories

---

*Vatican Observatory Summer School on Big Data and  
Machine Learning 2023 (VOSS-2023)*

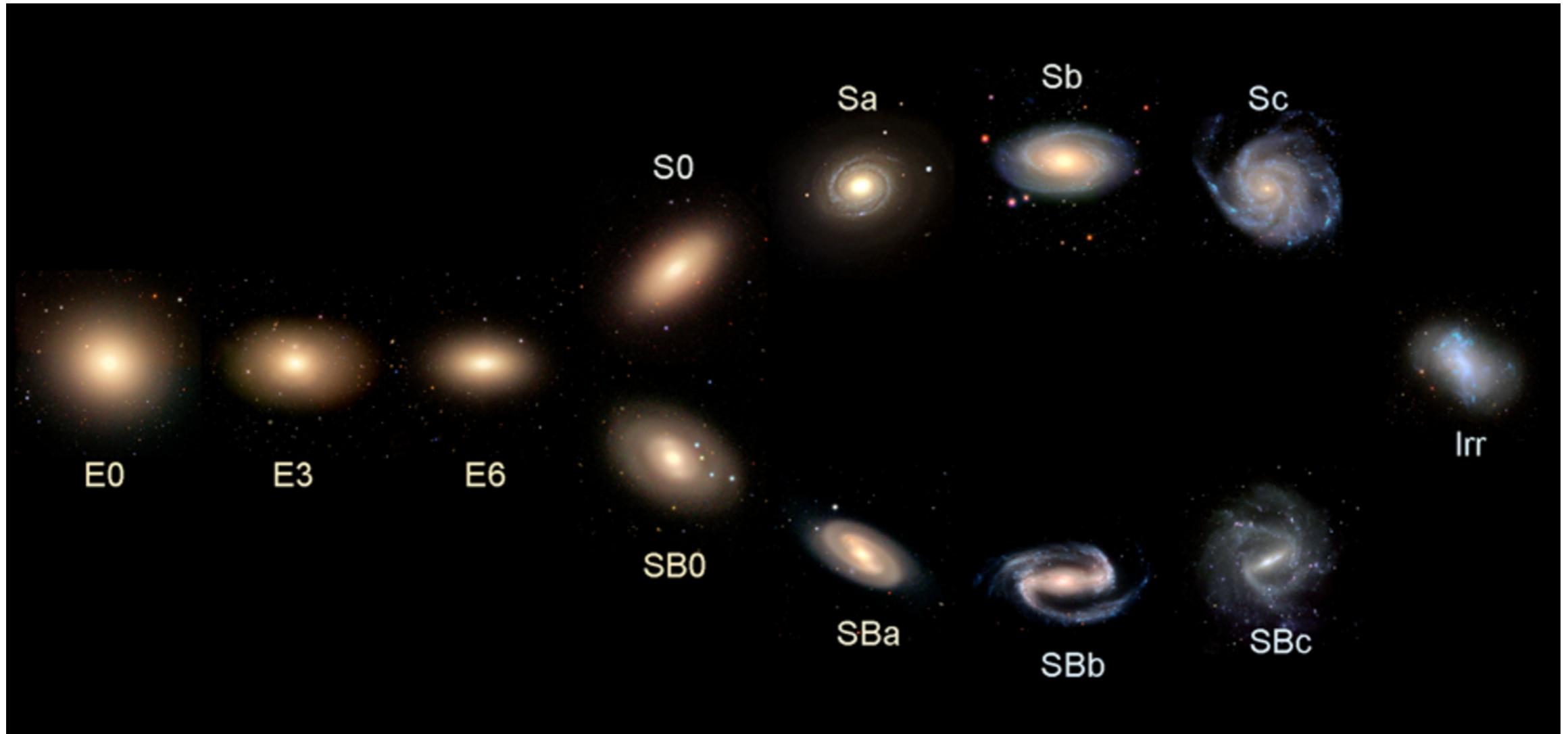
# Science is about compression



$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$$

How can we extract simplicity from the observed complex world?

# Simple structures: classes

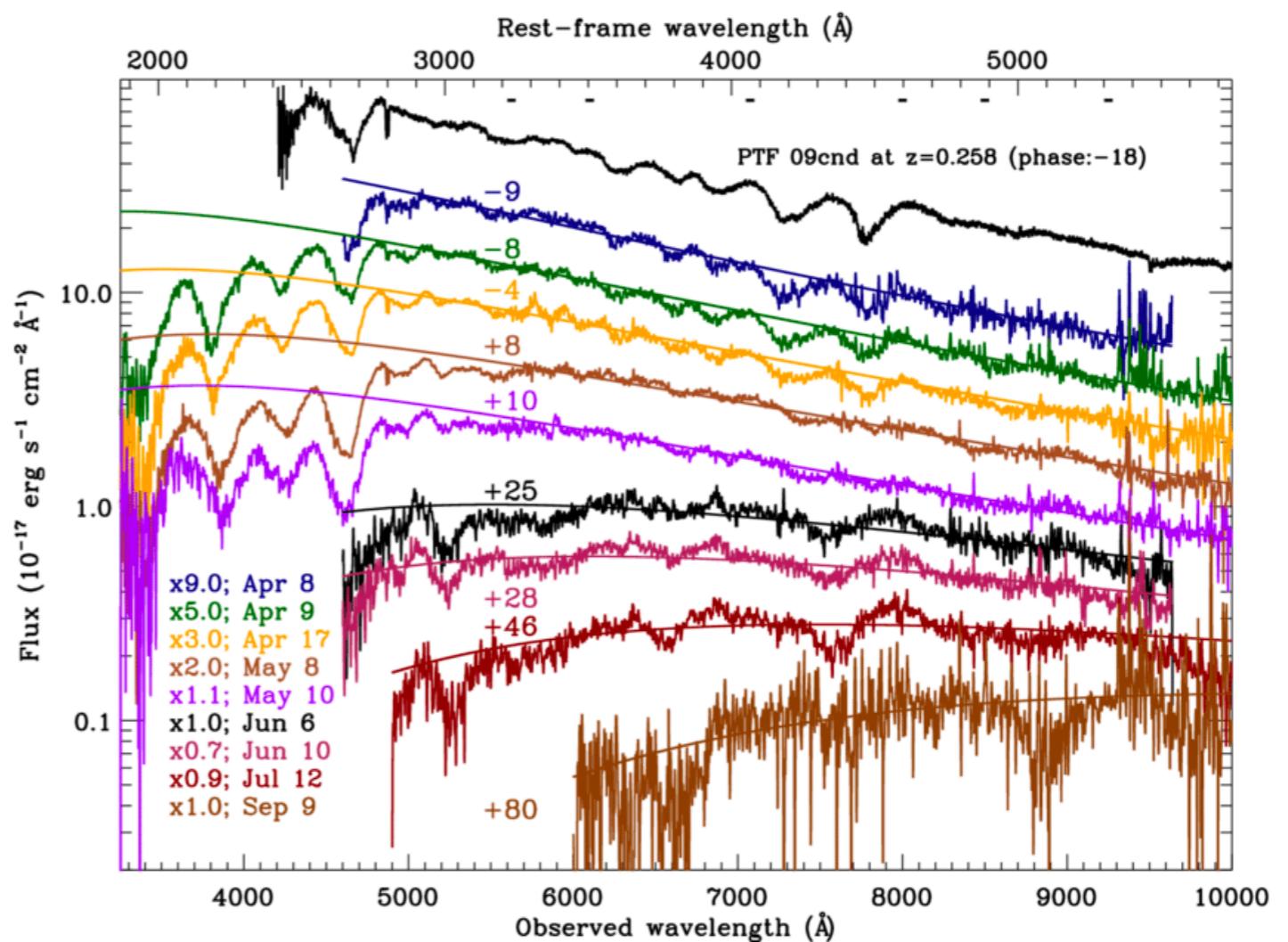


- Galaxies: spiral and elliptical.
- Supernovae: type Ia, type II, etc.
- AGN: type I and type II.

# Simple structures: sequences

A sequence: a set of properties that change continuously as a function of some parameter.

Sequences as a function of time:

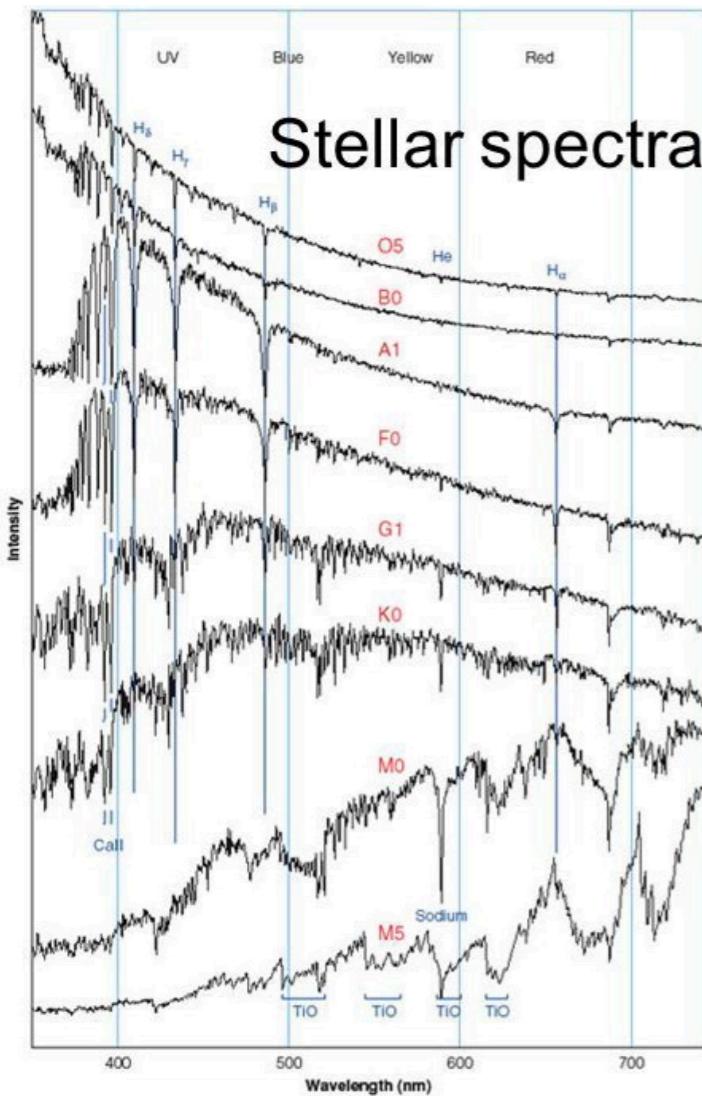


# Simple structures: sequences

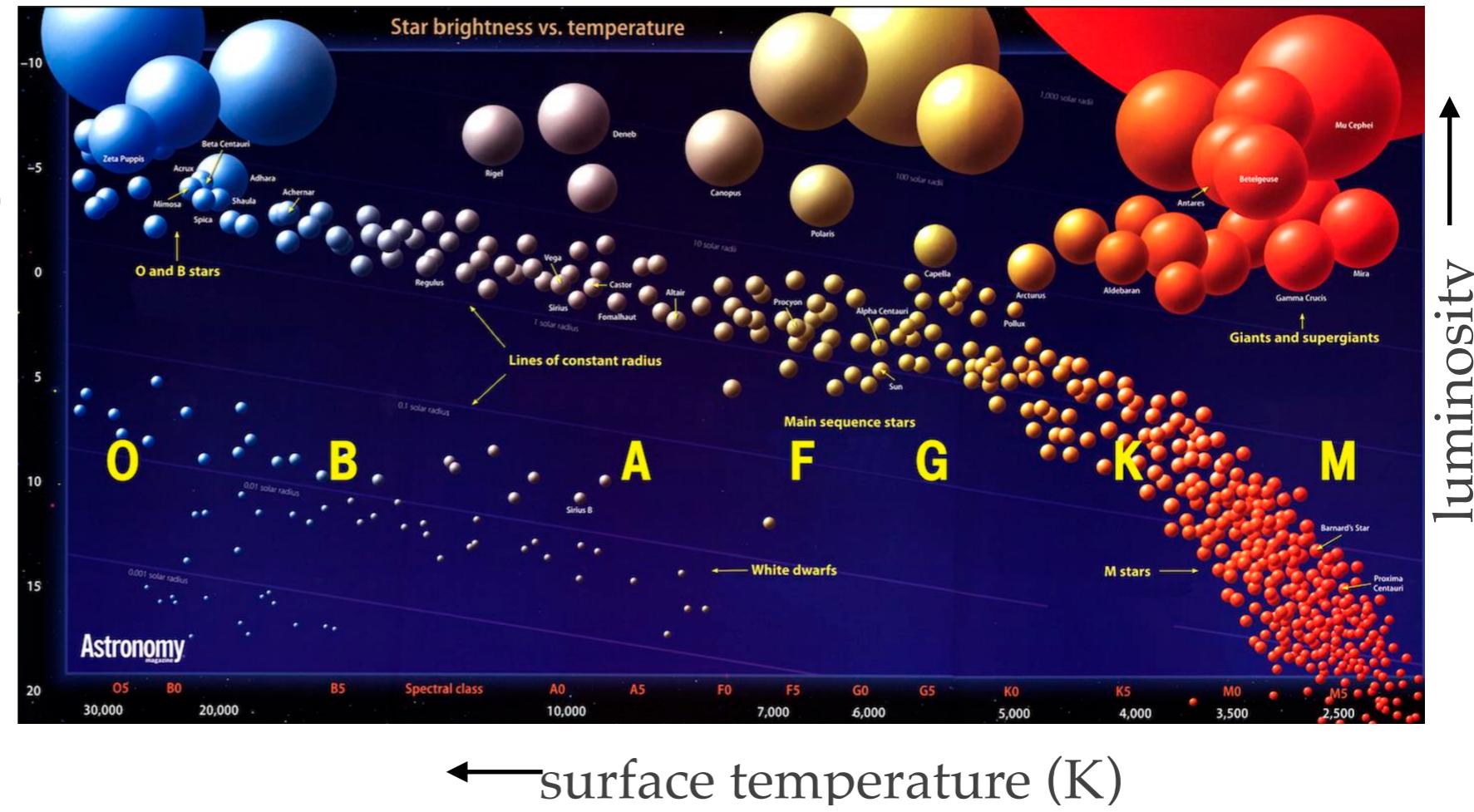
Sequences as a function of time:



# Simple structures: sequences



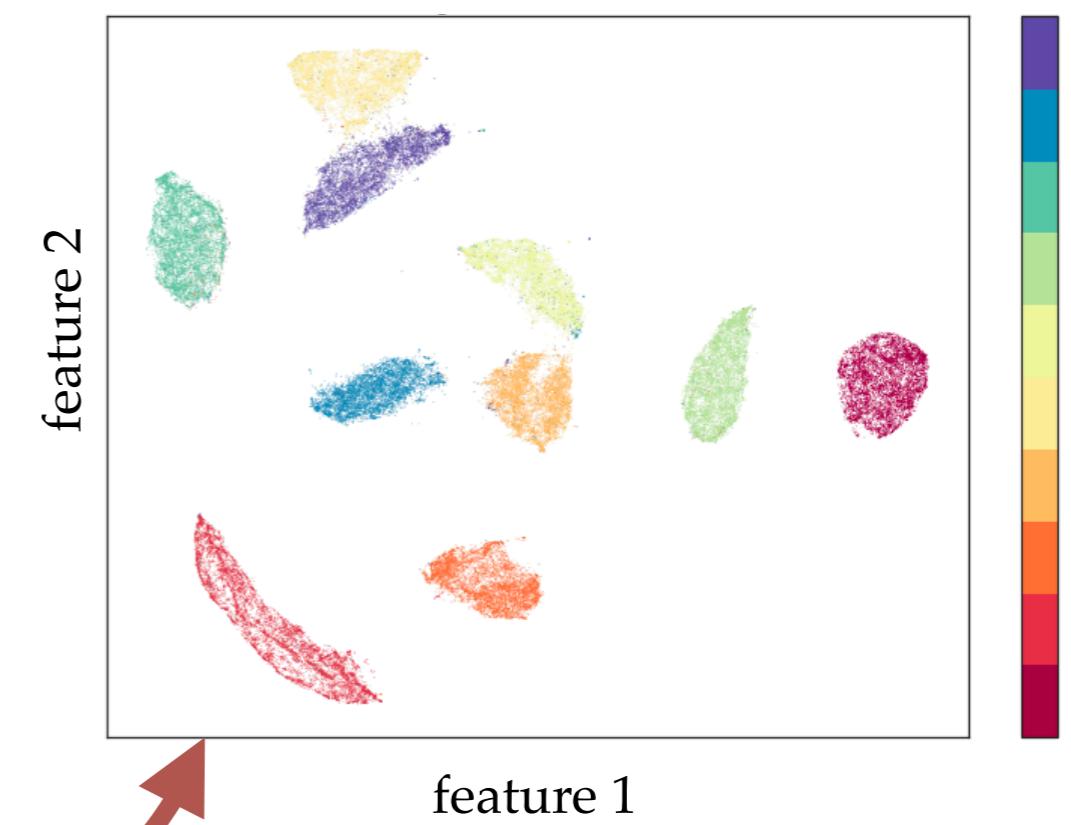
absolute magnitude



# Dimensionality Reduction Algorithms

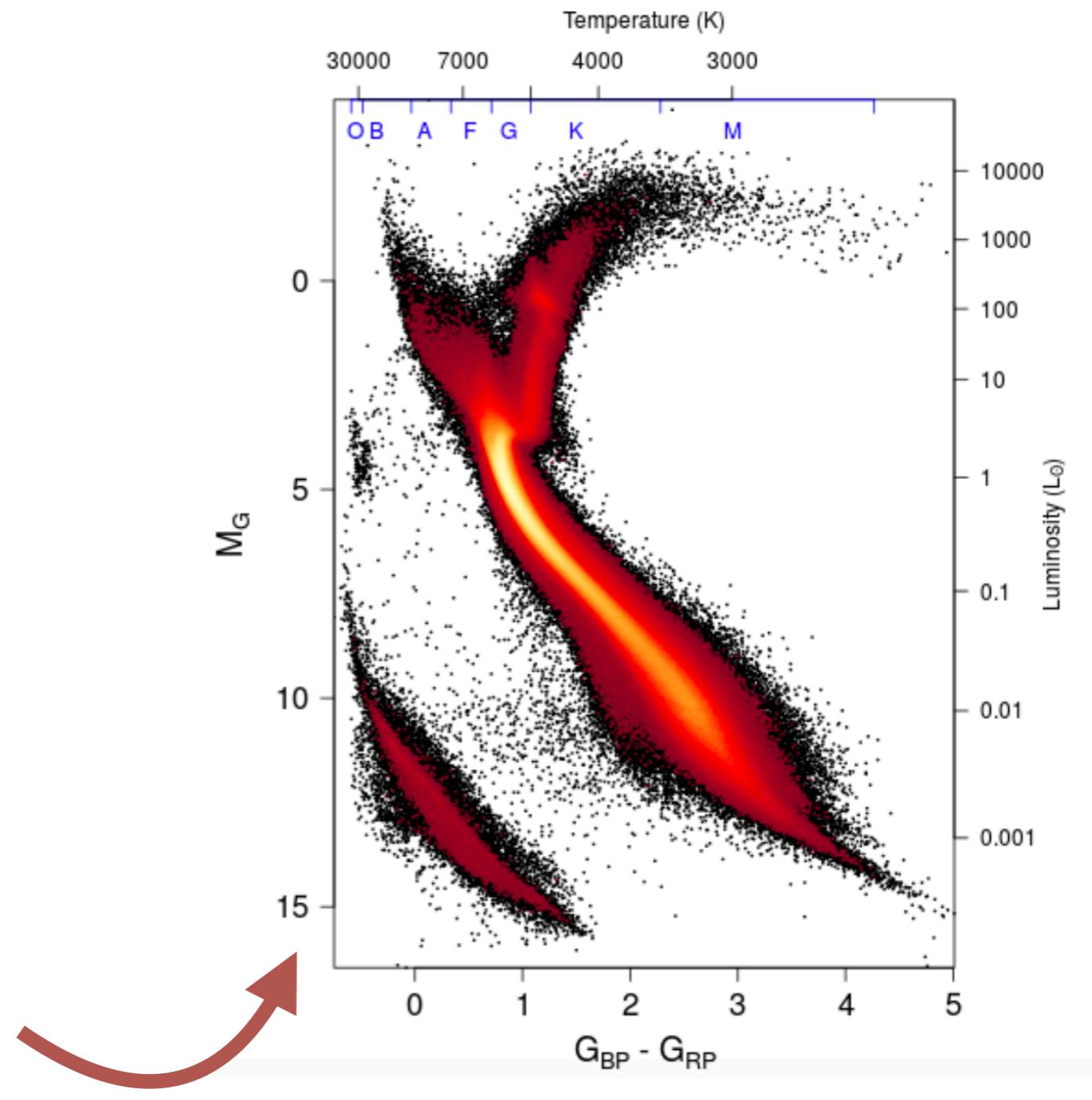
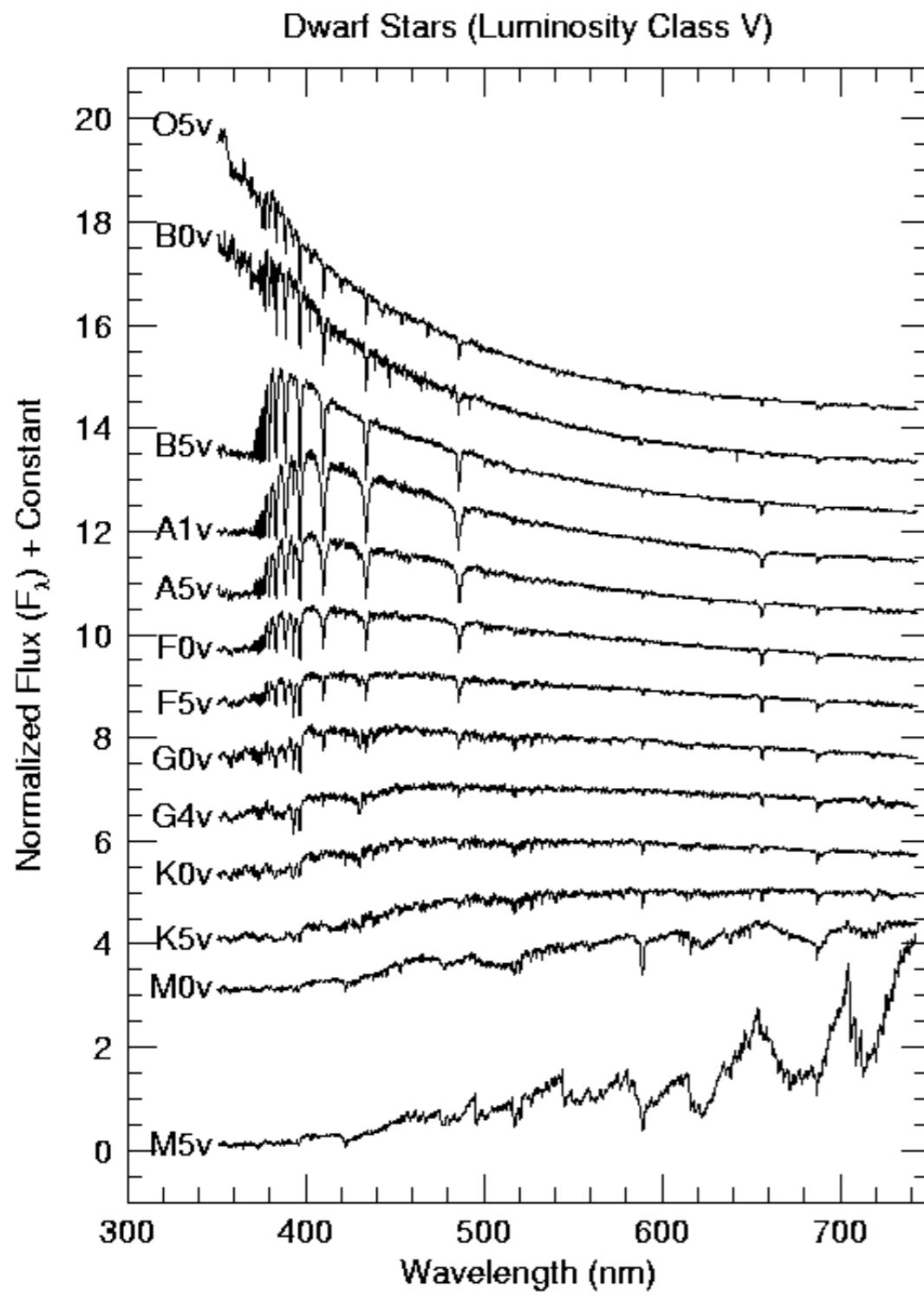
- ❖ PCA, tSNE & UMAP are used to embed high-dimensional data into a low dimensional space.
- ❖ The resulting embedding depends on several hyper-parameters and external choices, such as the distance metric.

MNIST dataset: 28x28 features per image



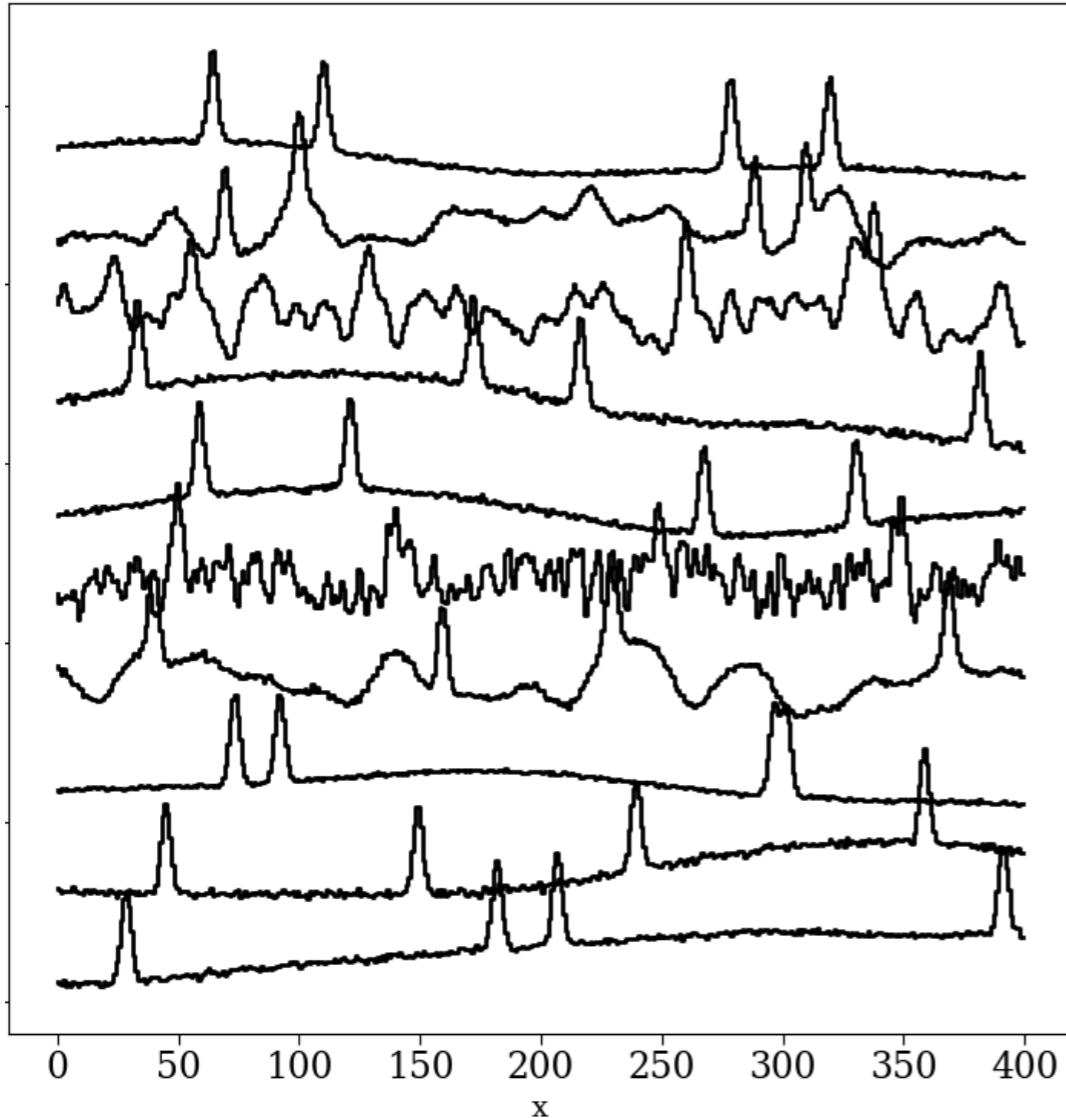
From: <https://umap-learn.readthedocs.io/>

# Our Ultimate Goal:



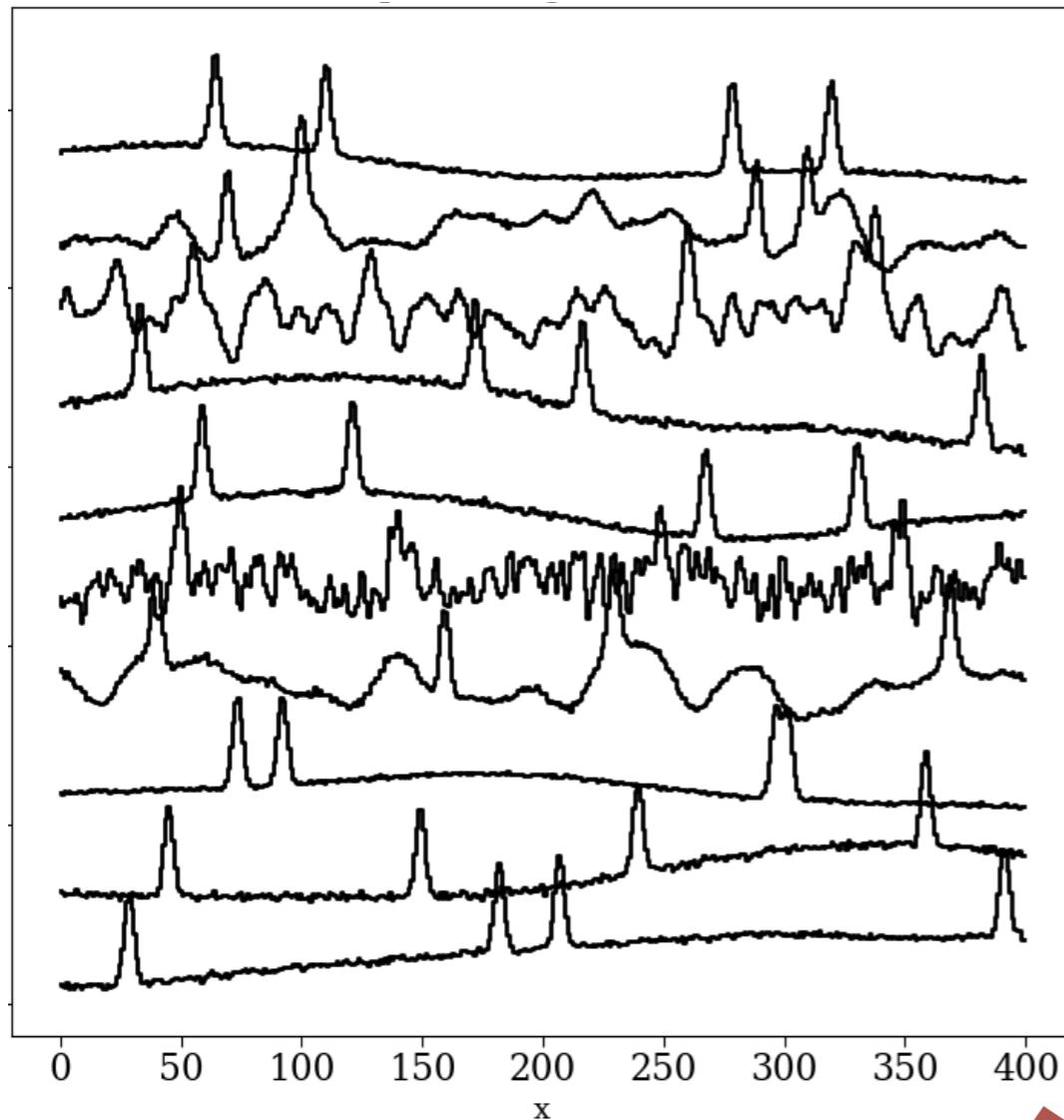
# The Sequencer Algorithm

Input dataset,  
random order

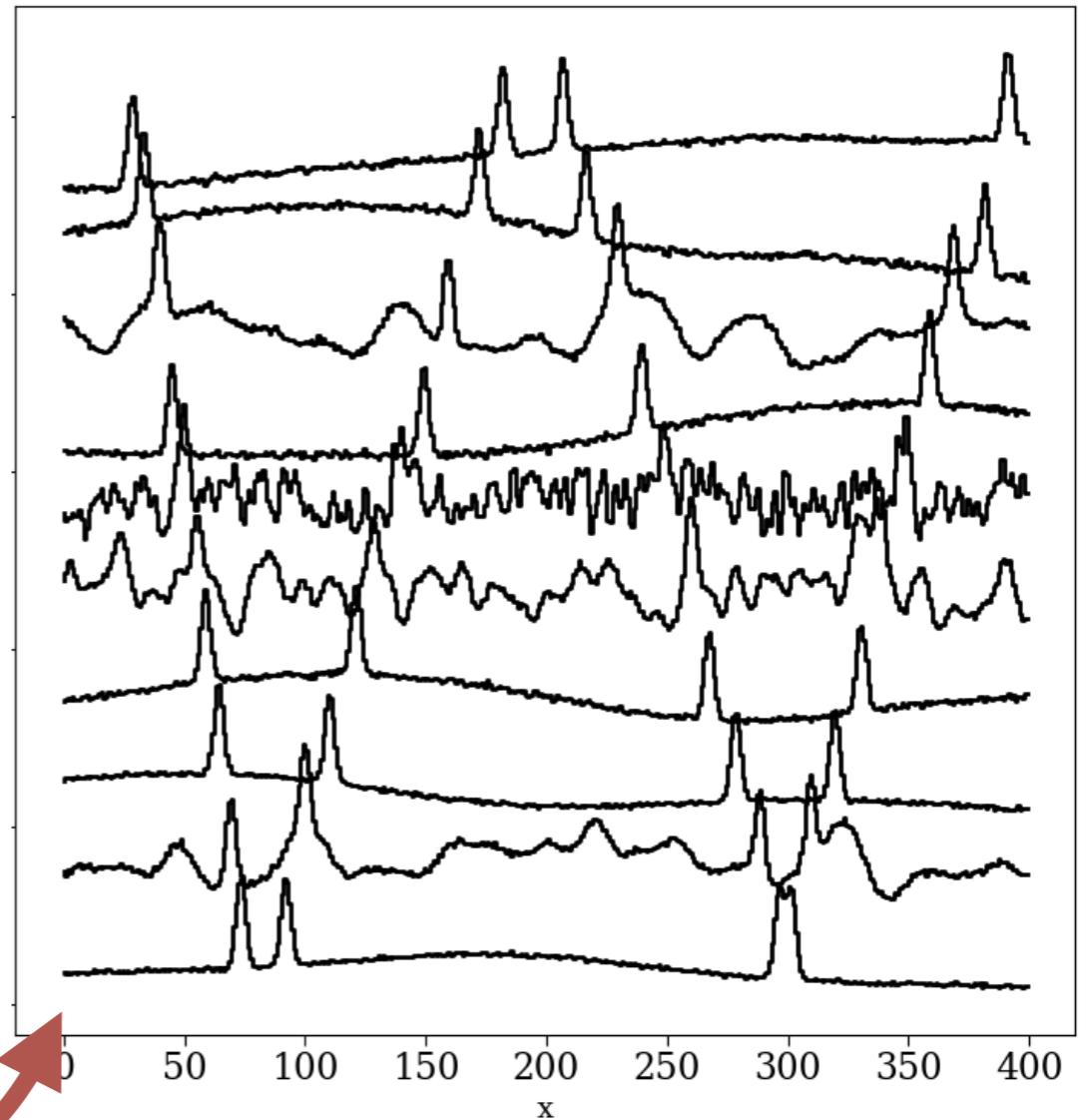


# The Sequencer Algorithm

Input dataset,  
random order

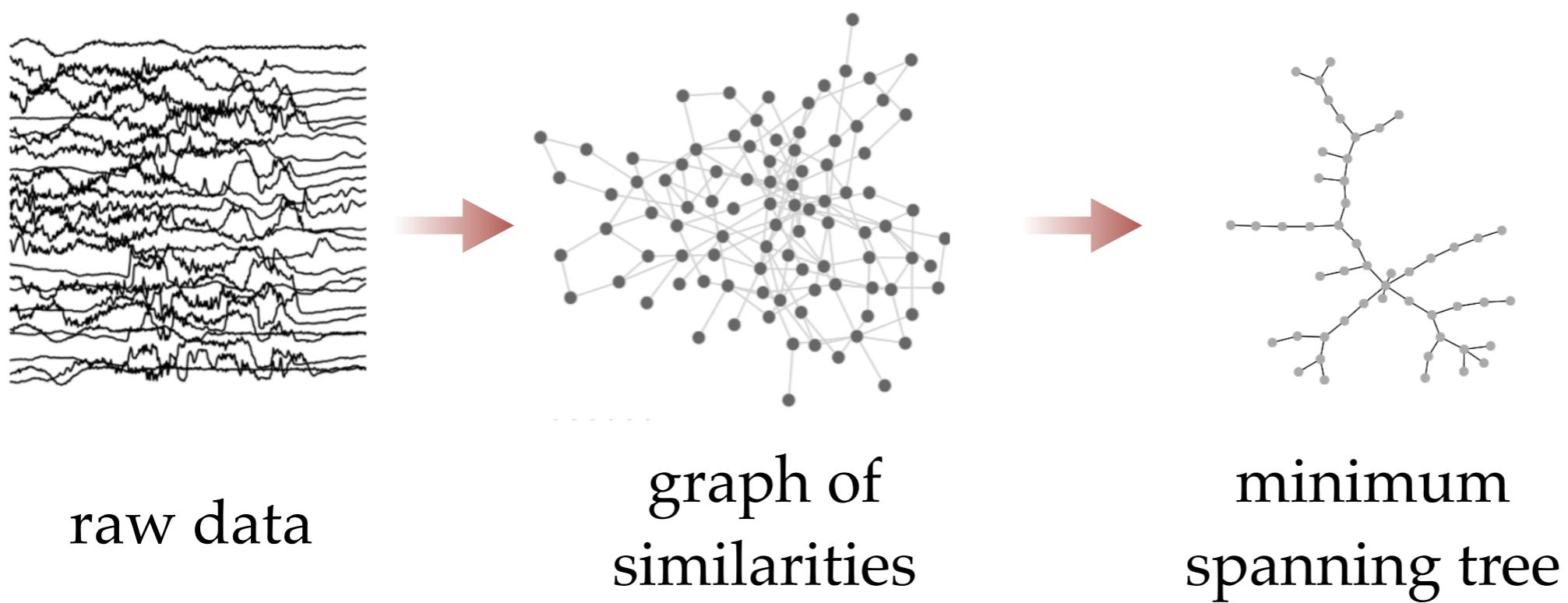


Ordered dataset according  
to detected sequence

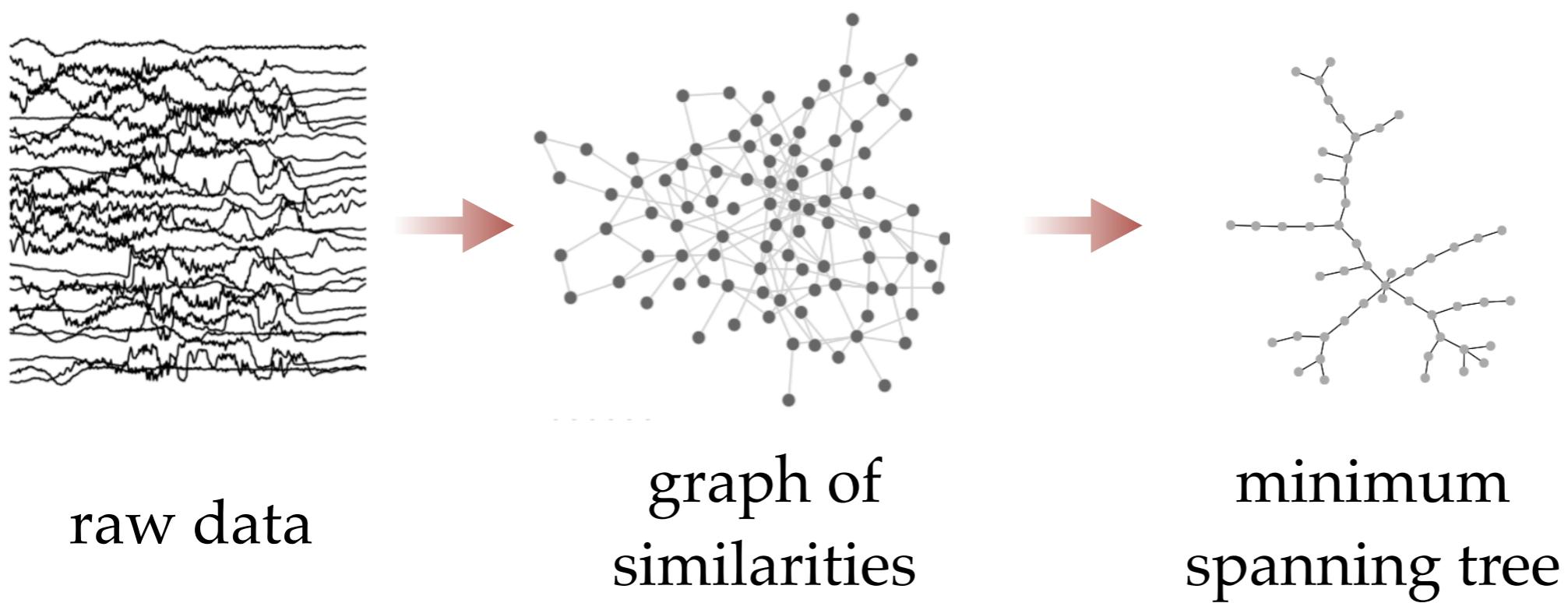


Sequencer

# The Sequencer Algorithm

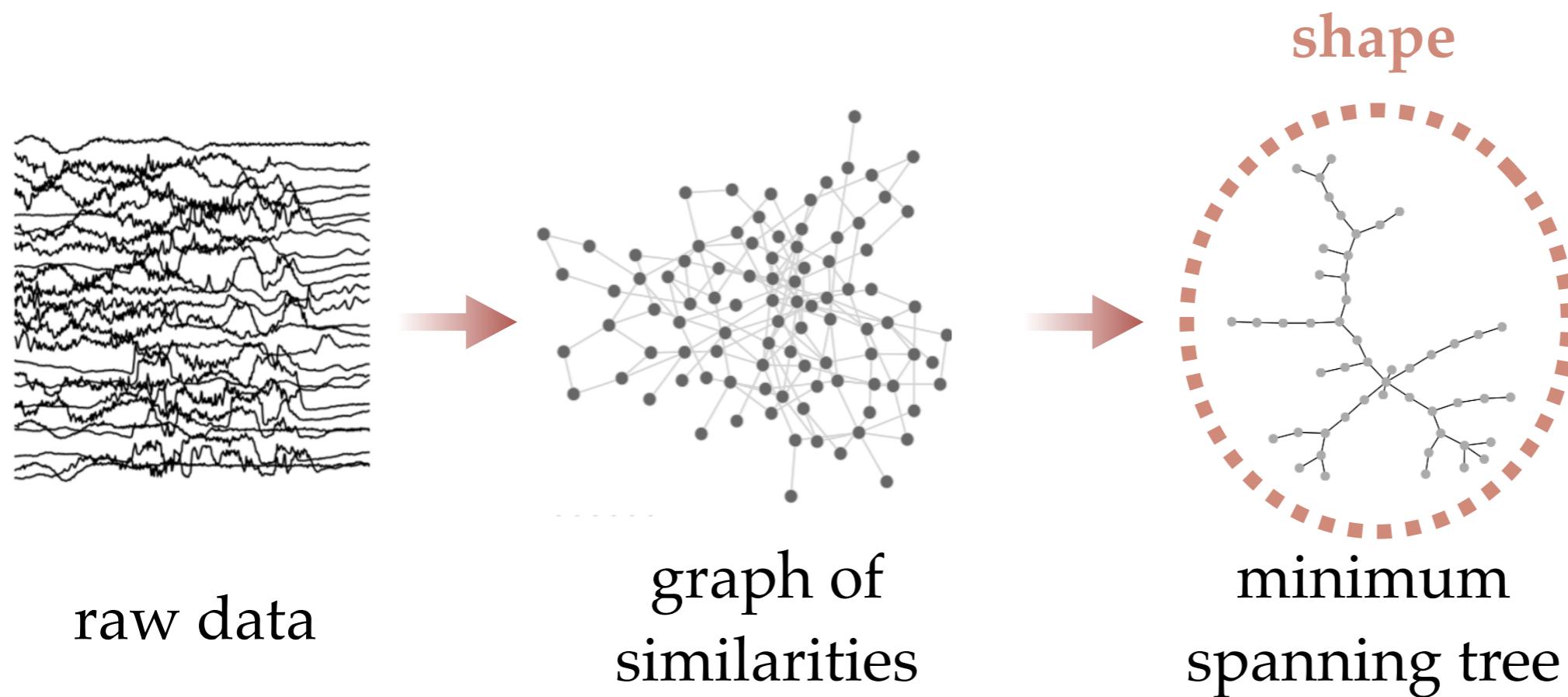


# The Sequencer Algorithm

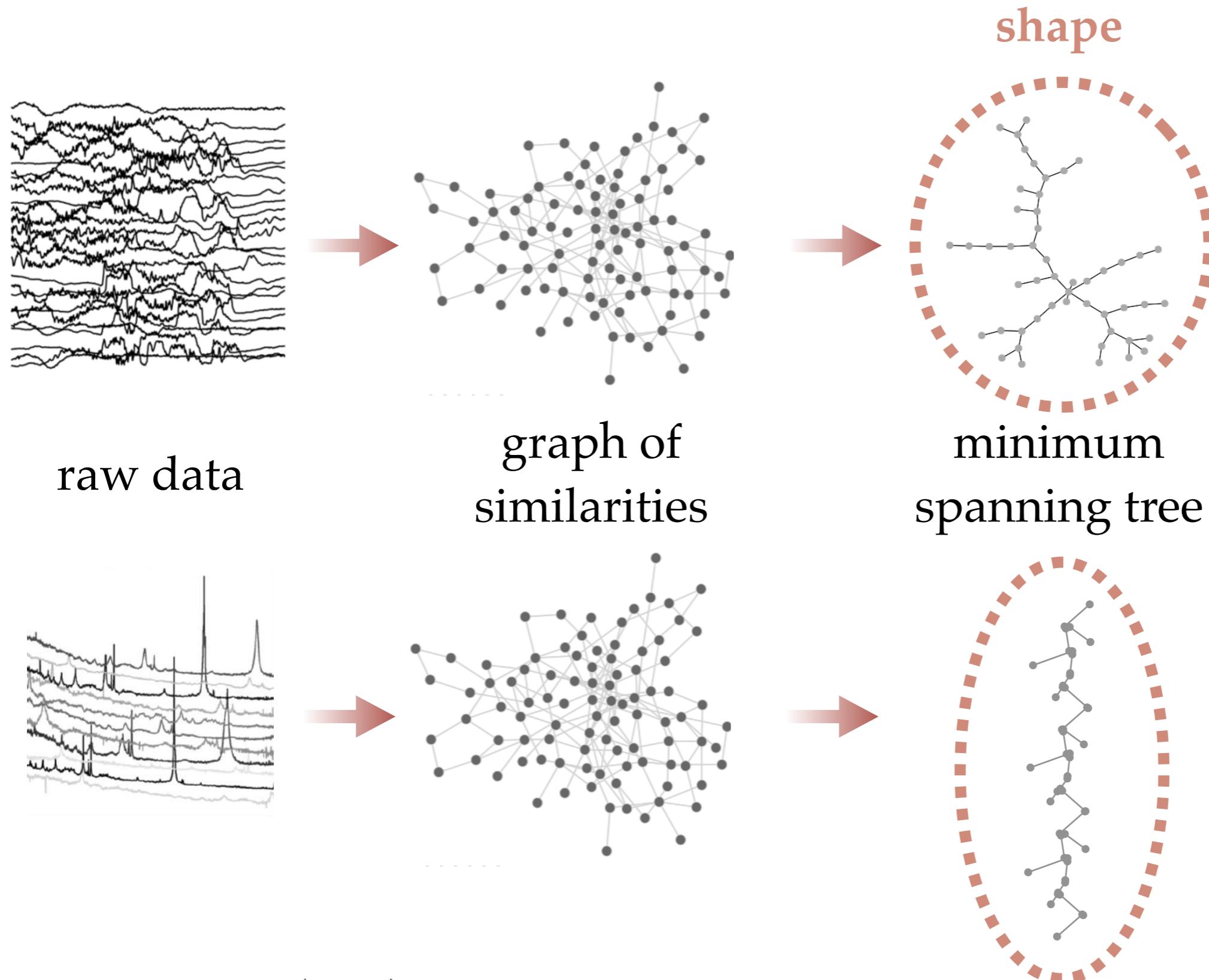


MST: the subset of the edges in a fully-connected graph that connects all the nodes together, without any cycles, and with the minimum possible total edge weight.

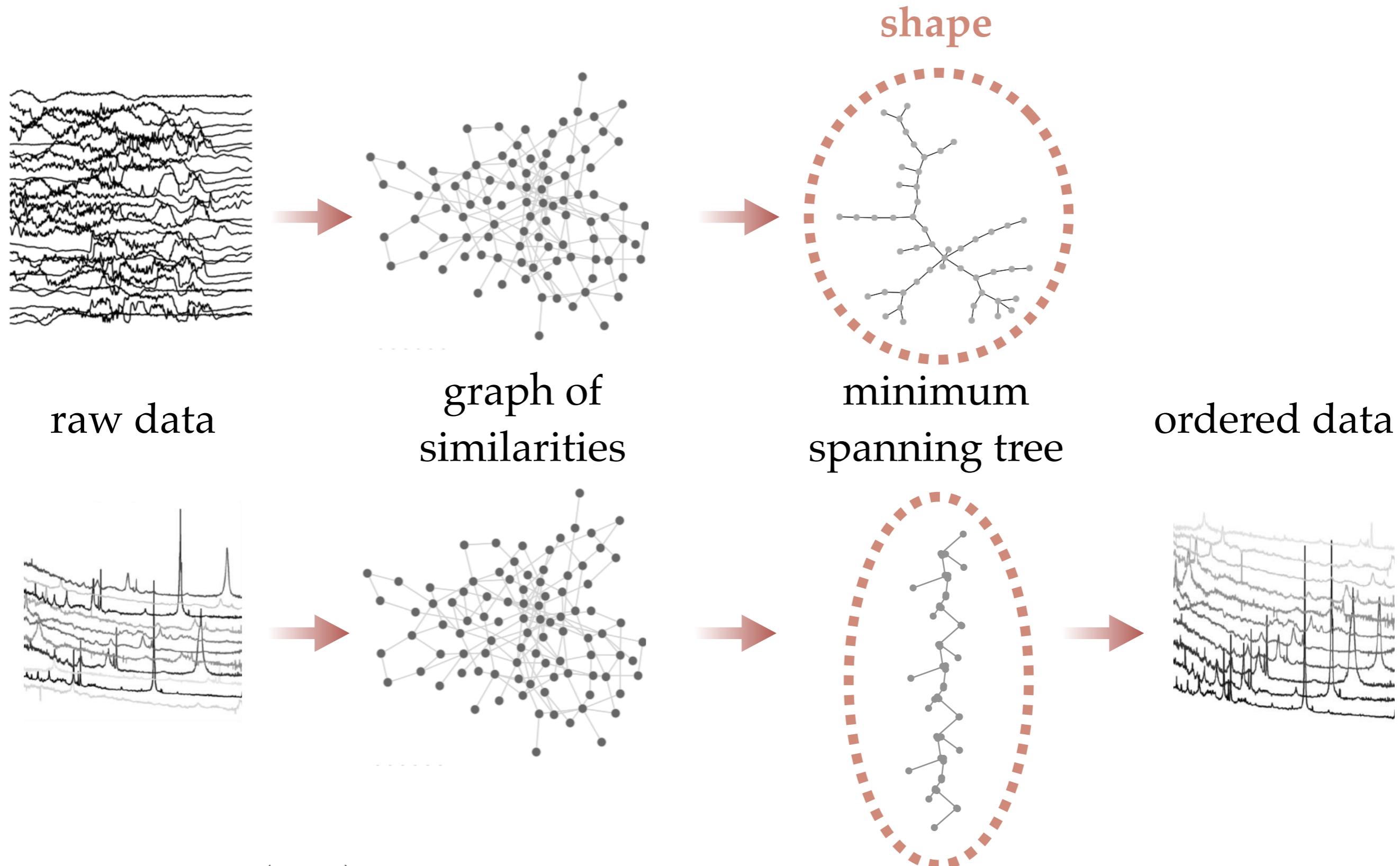
# The Sequencer Algorithm



# The Sequencer Algorithm

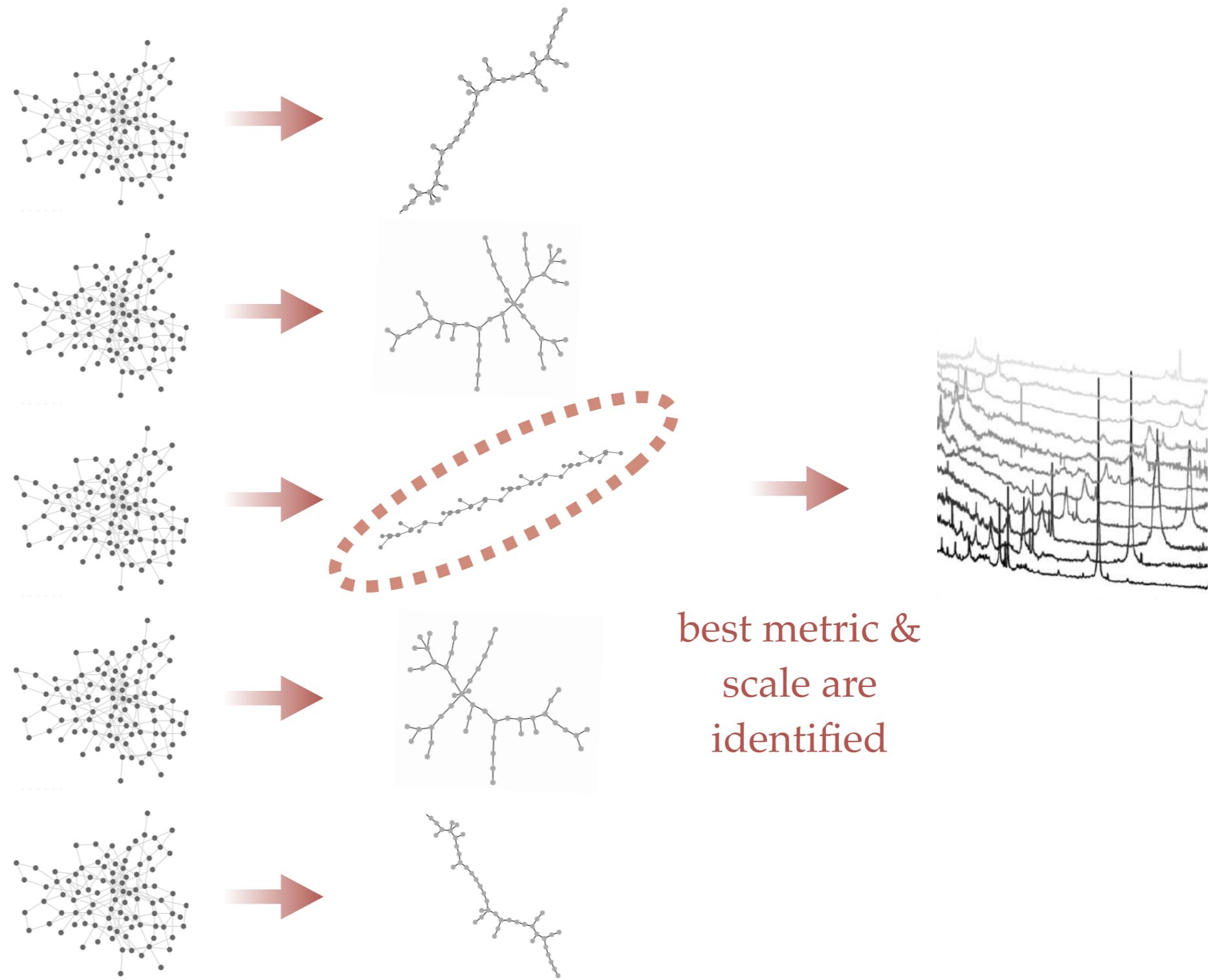
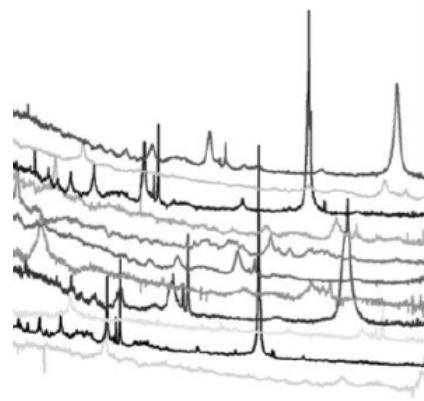


# The Sequencer Algorithm

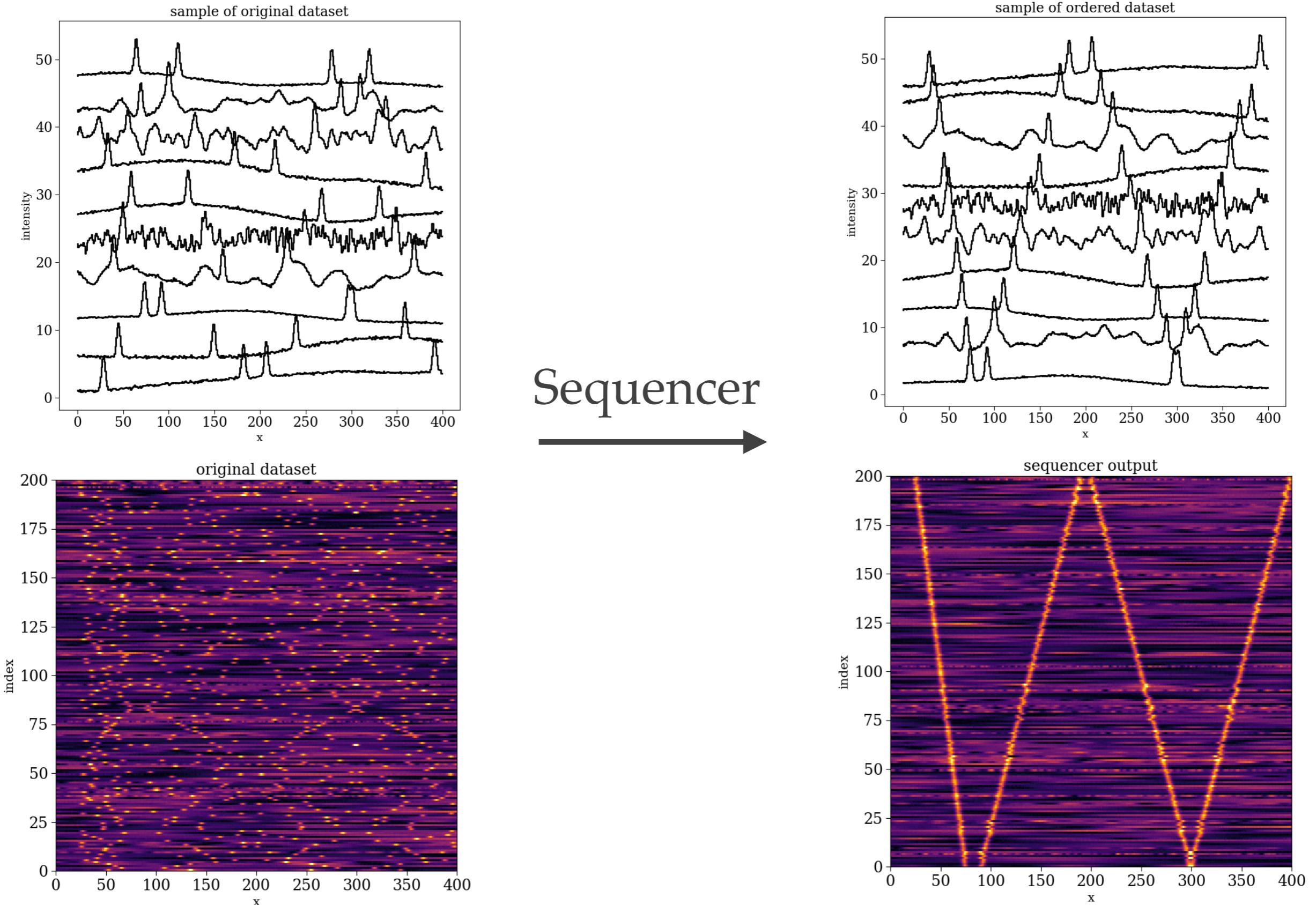


# The Sequencer Algorithm

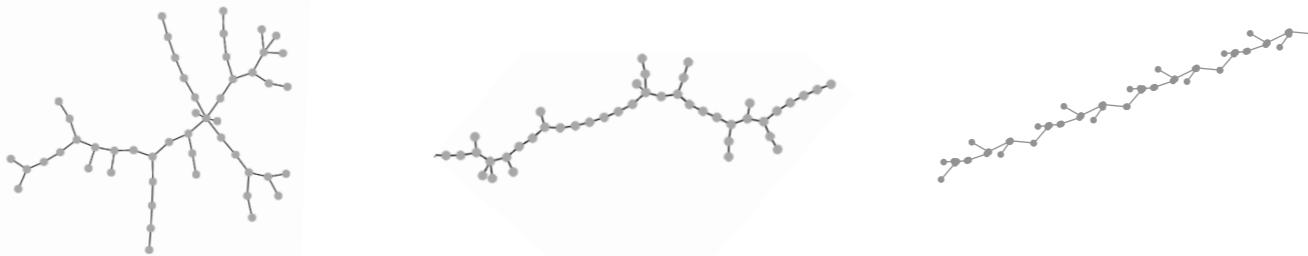
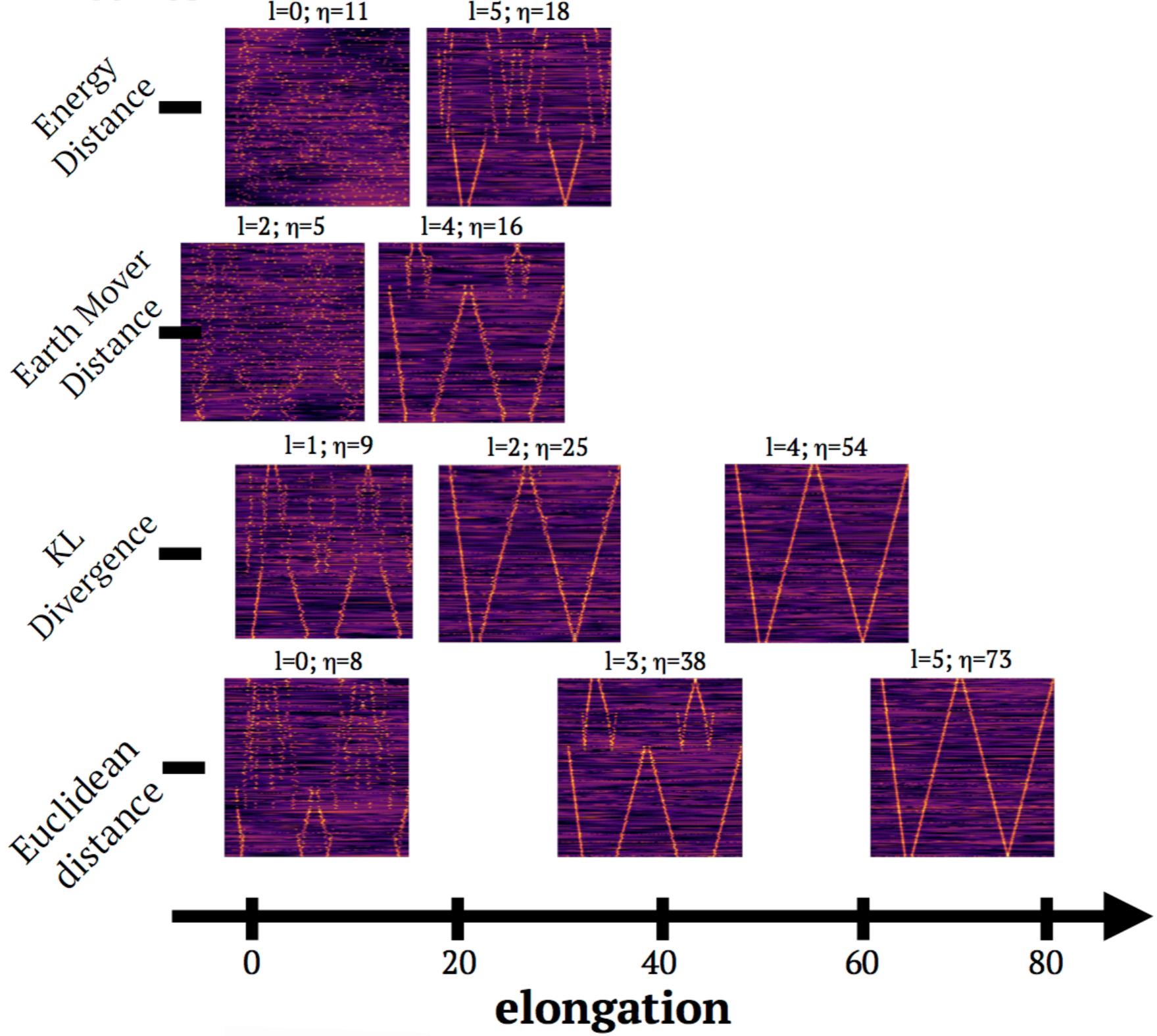
for different  
metrics & scales



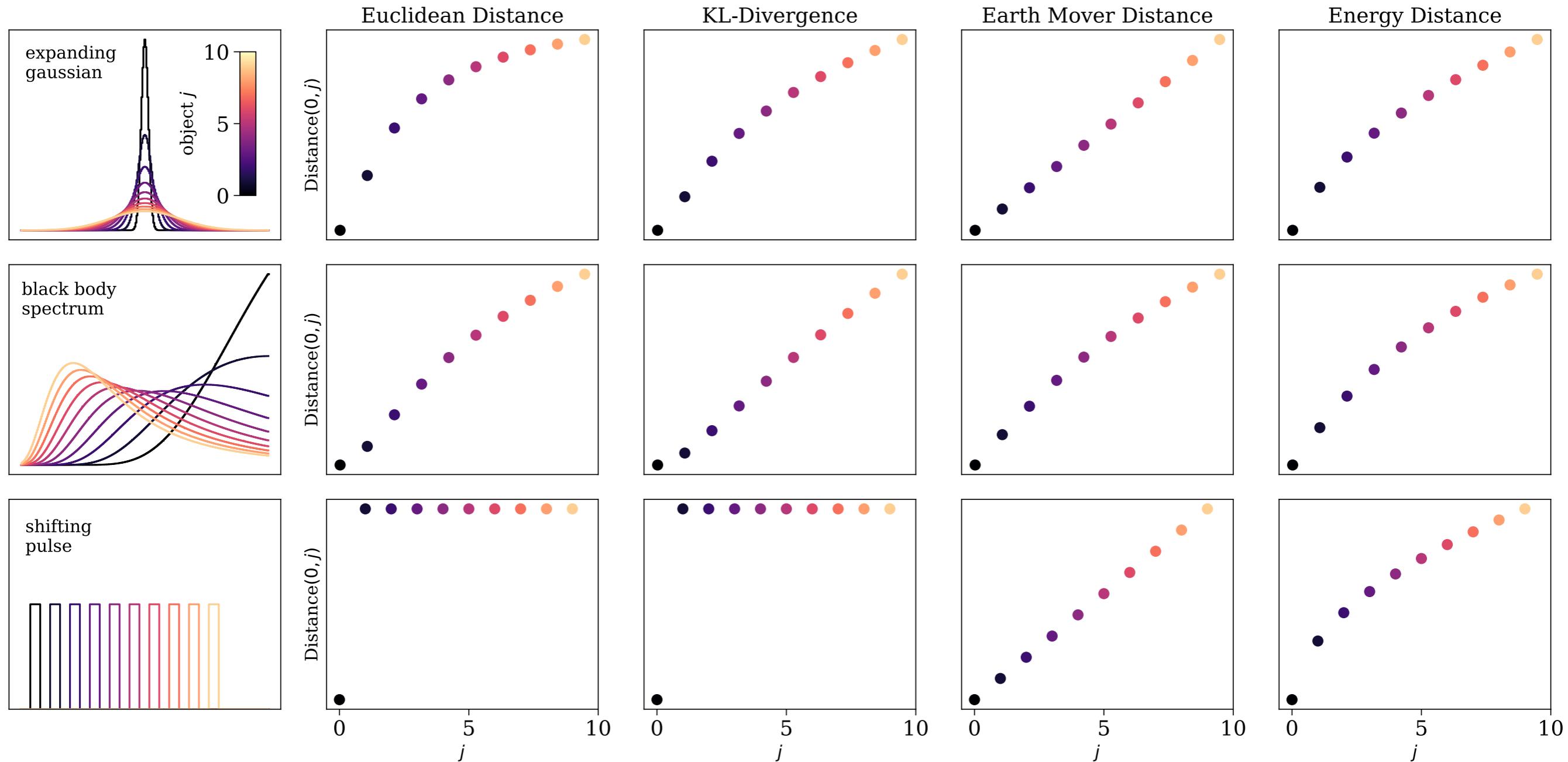
# Example with Simulated Data



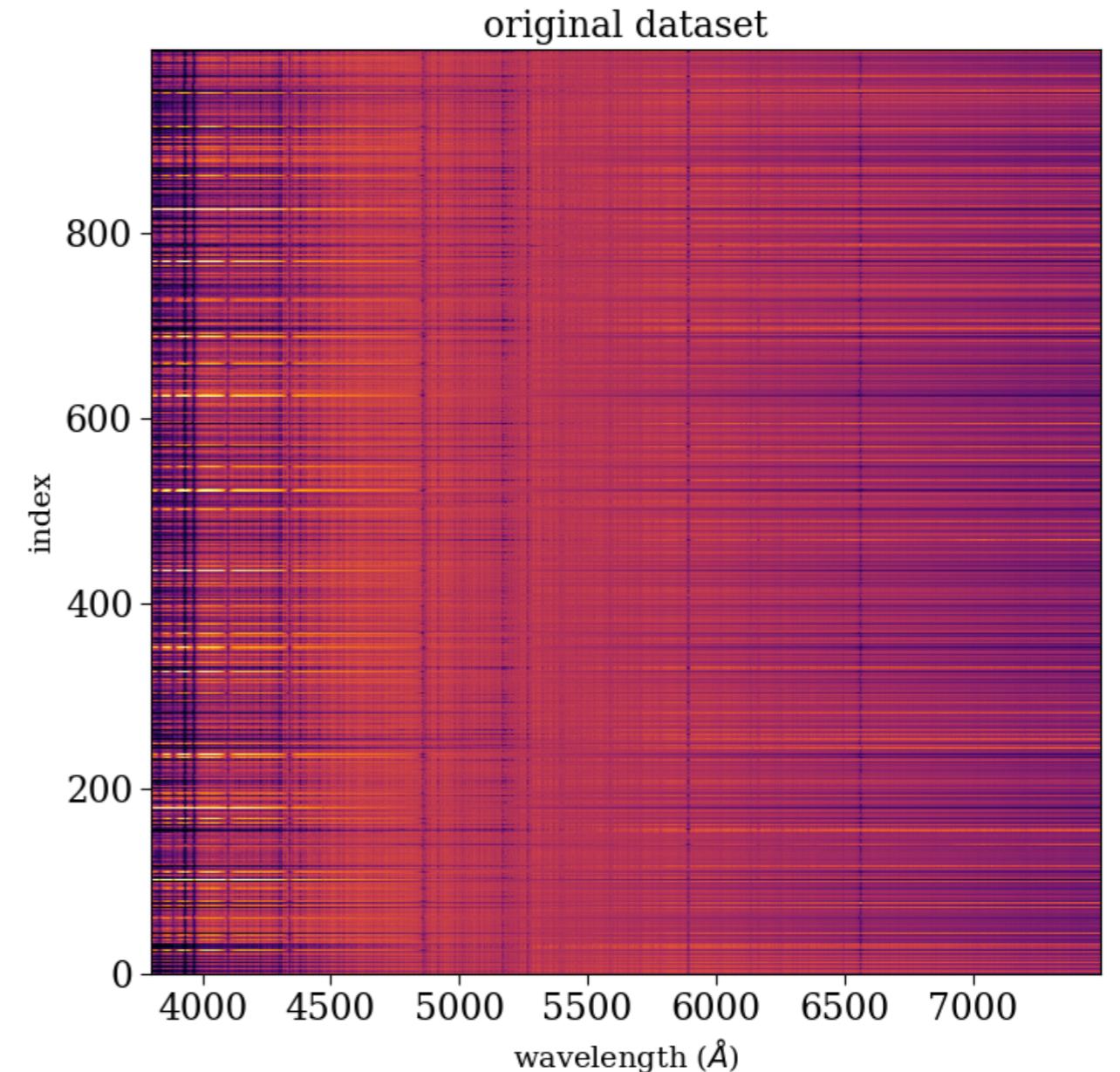
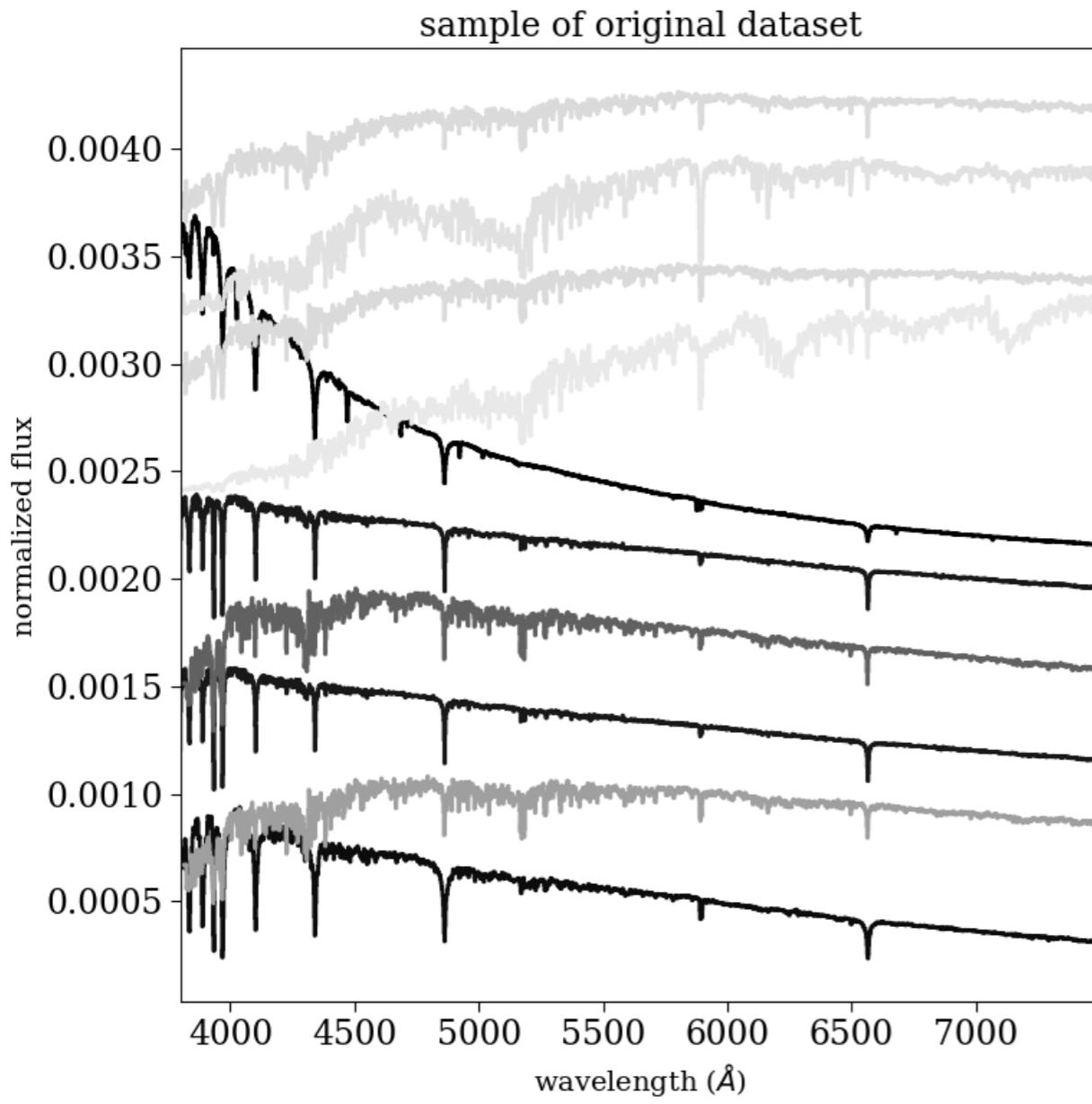
# Metrics



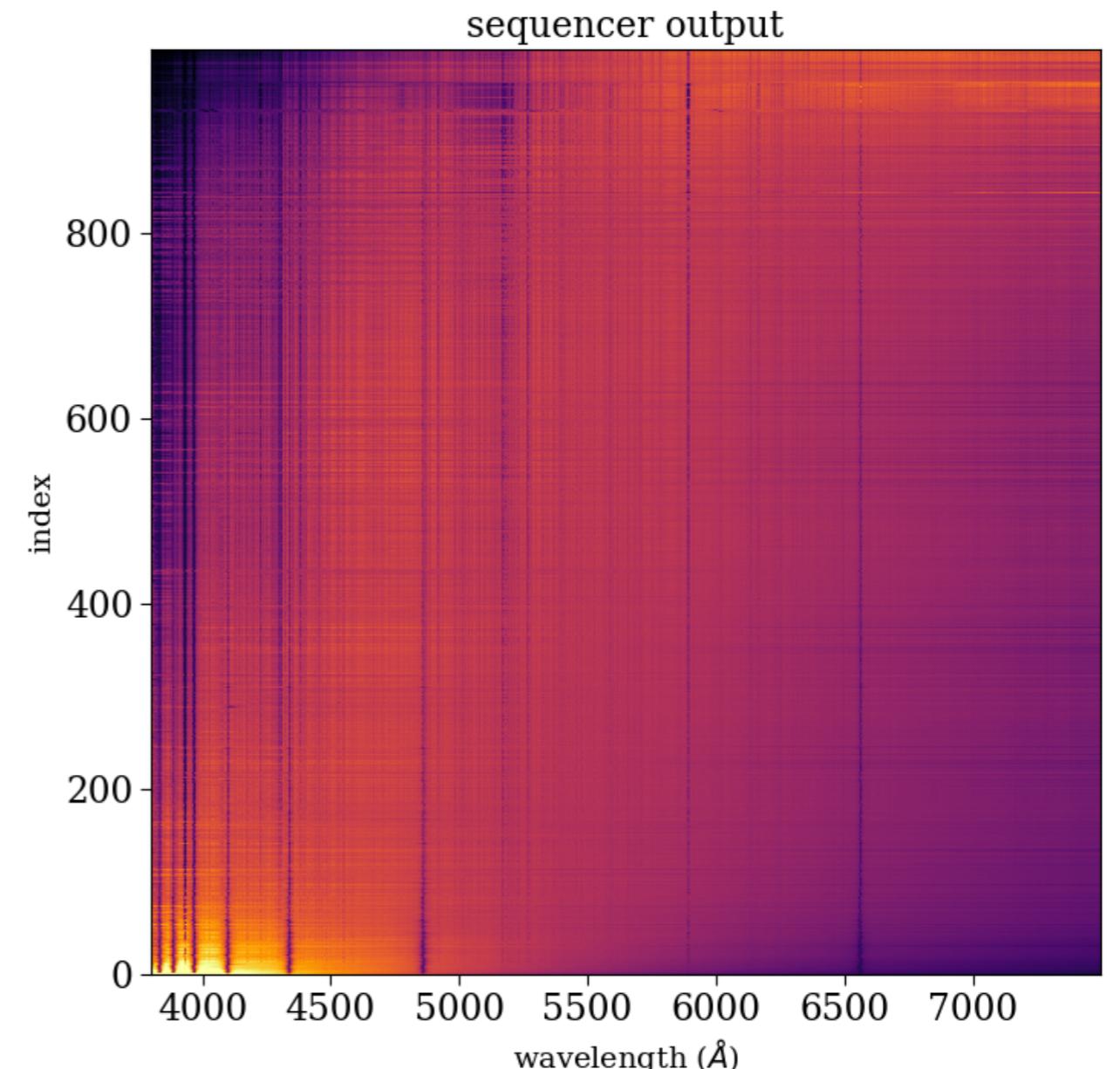
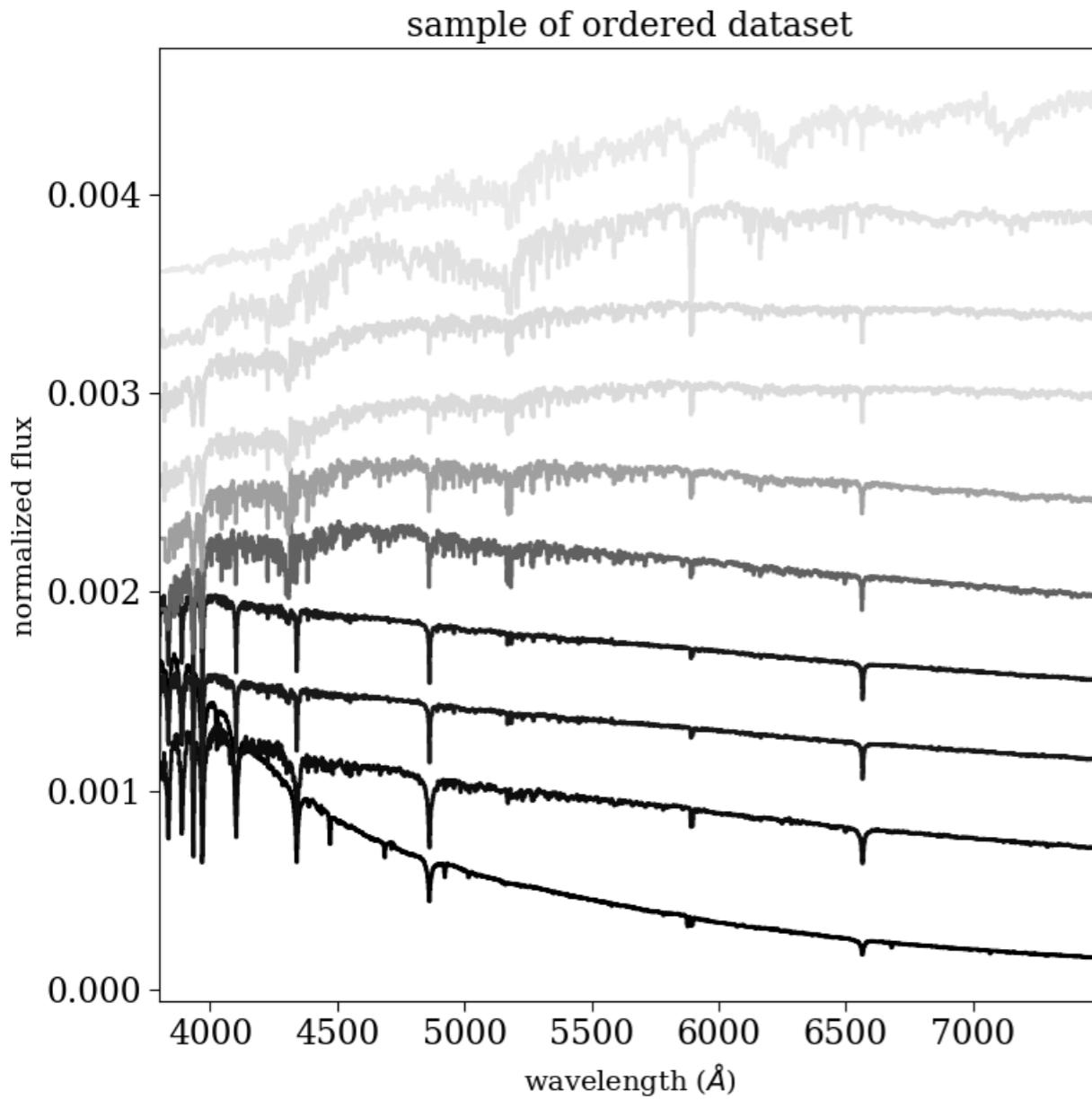
# Different distance measures



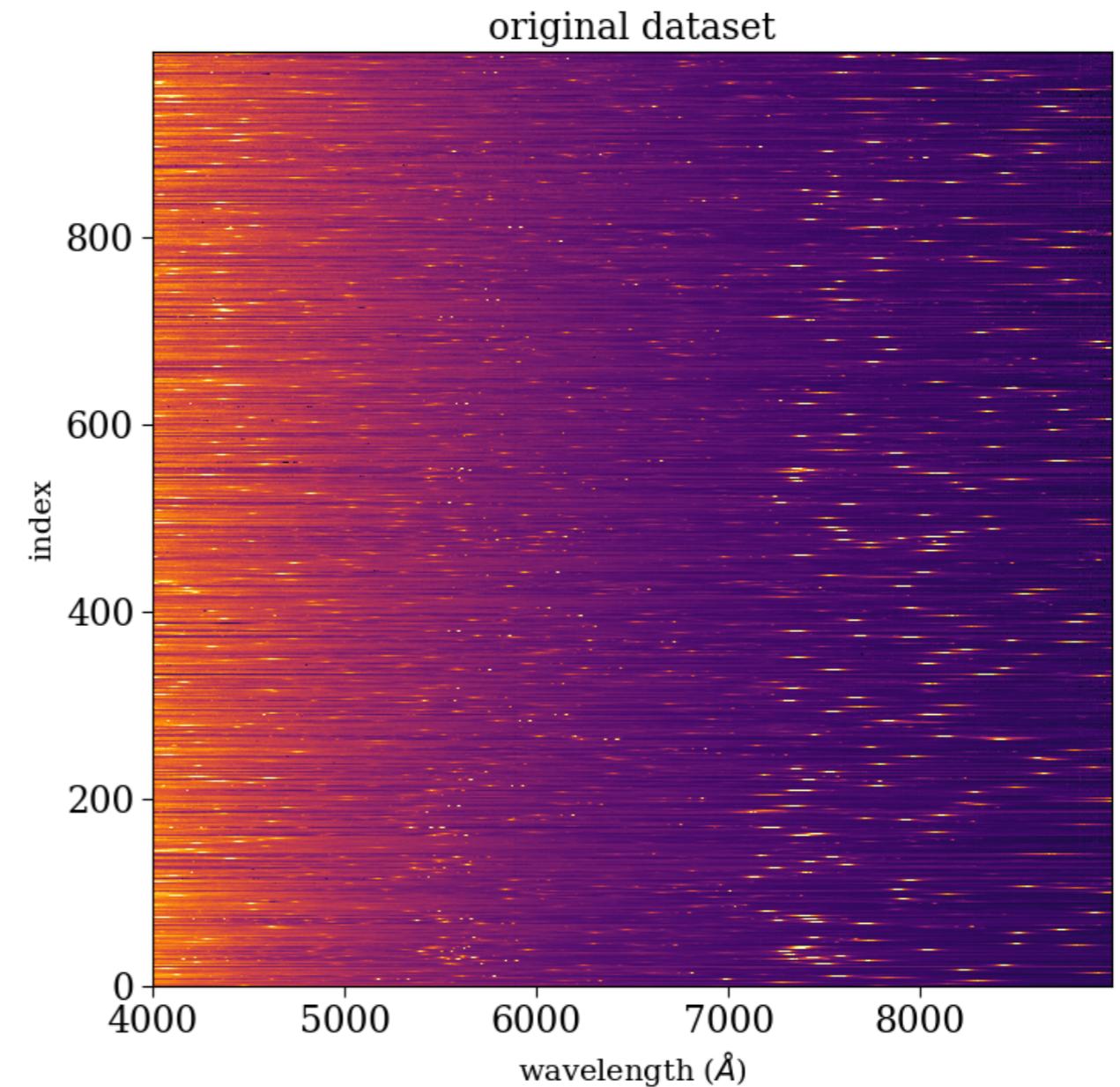
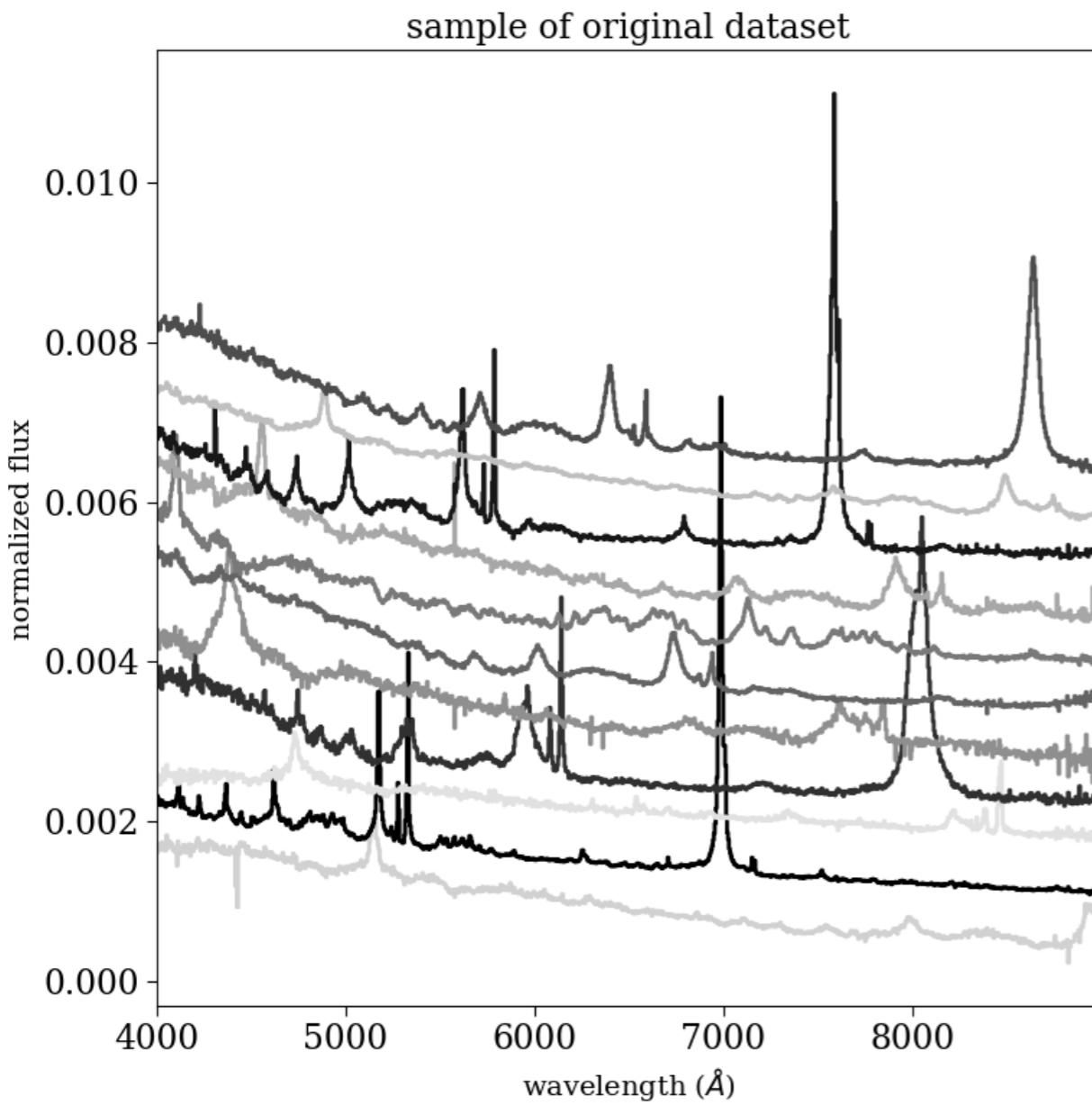
# Examples from Astronomy: stellar sequence



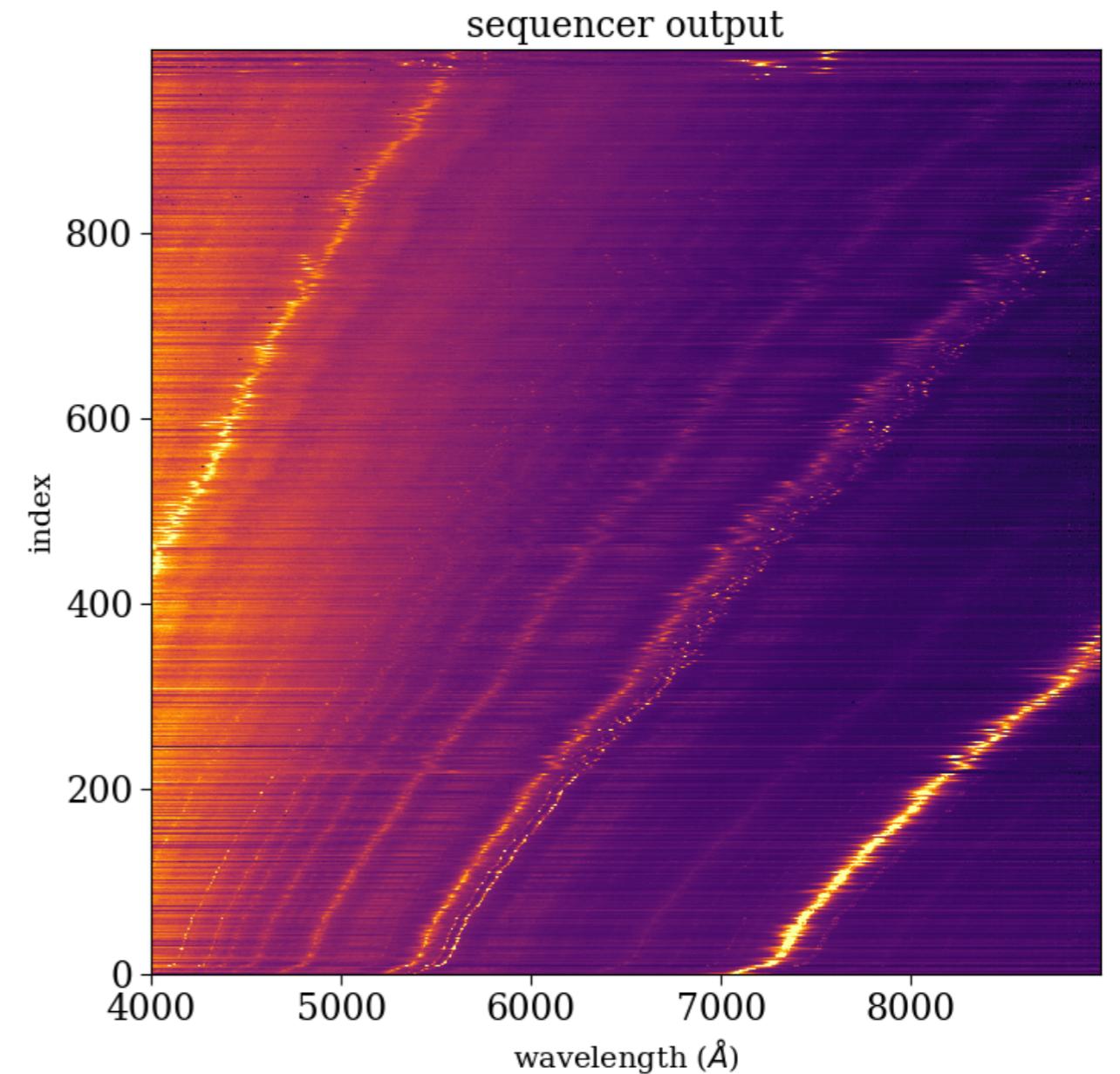
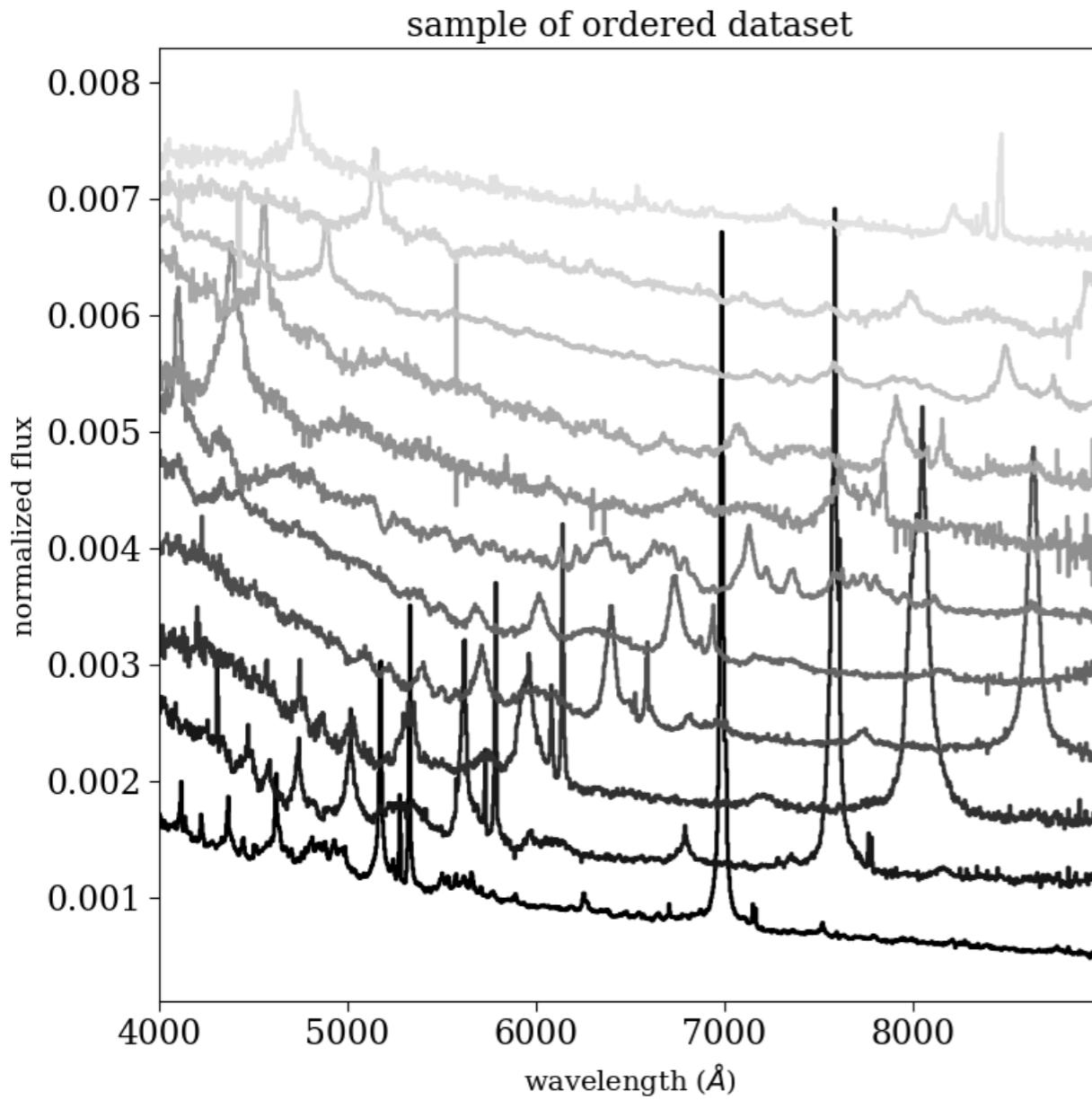
# Examples from Astronomy: stellar sequence



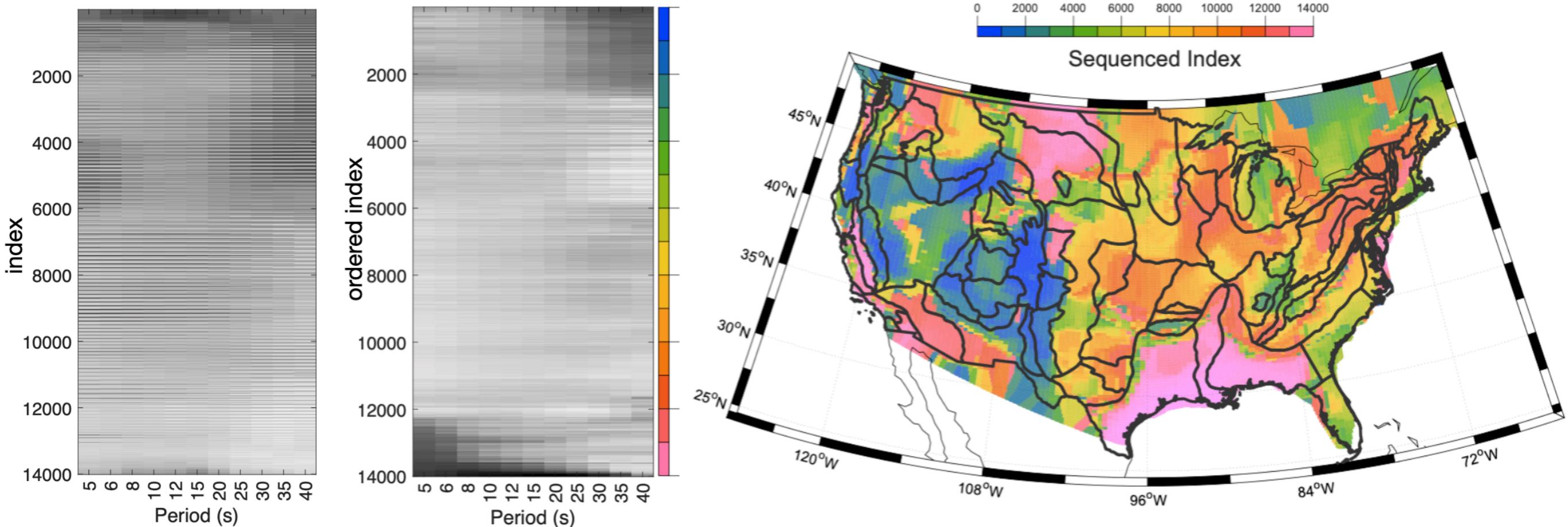
# Examples from Astronomy: redshift sequence



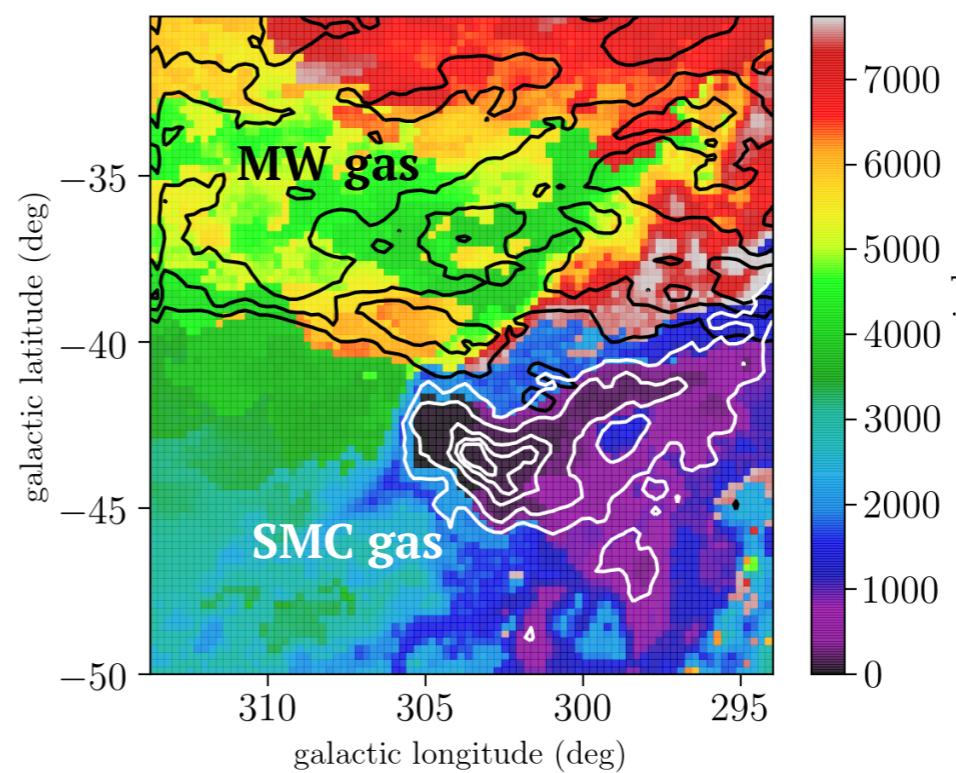
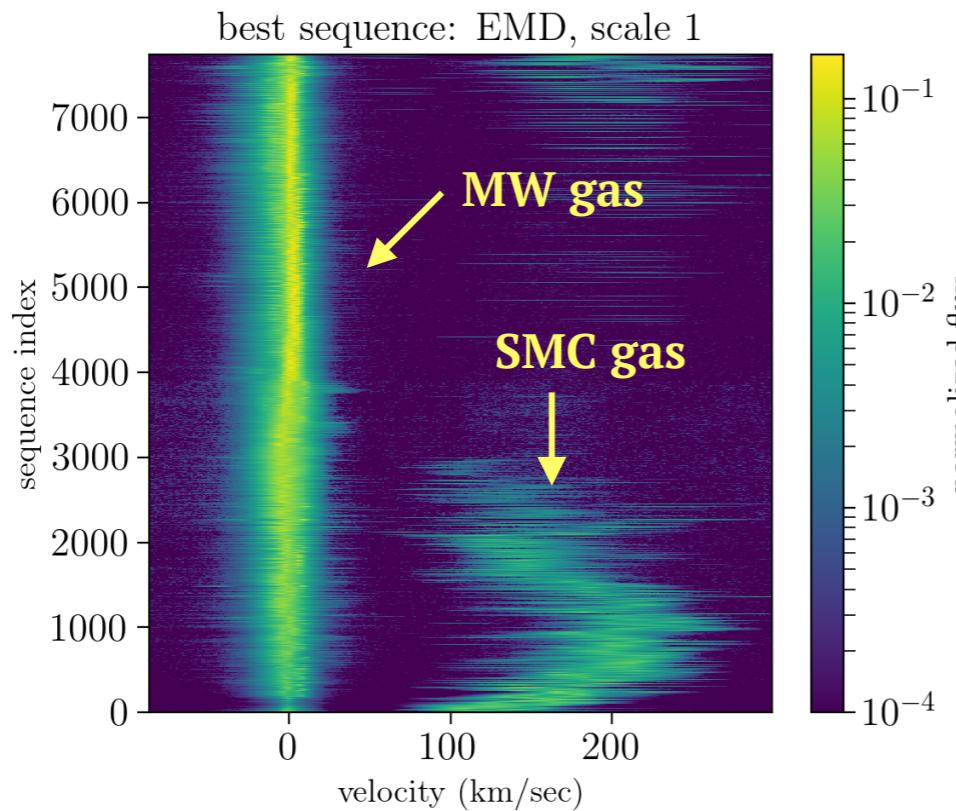
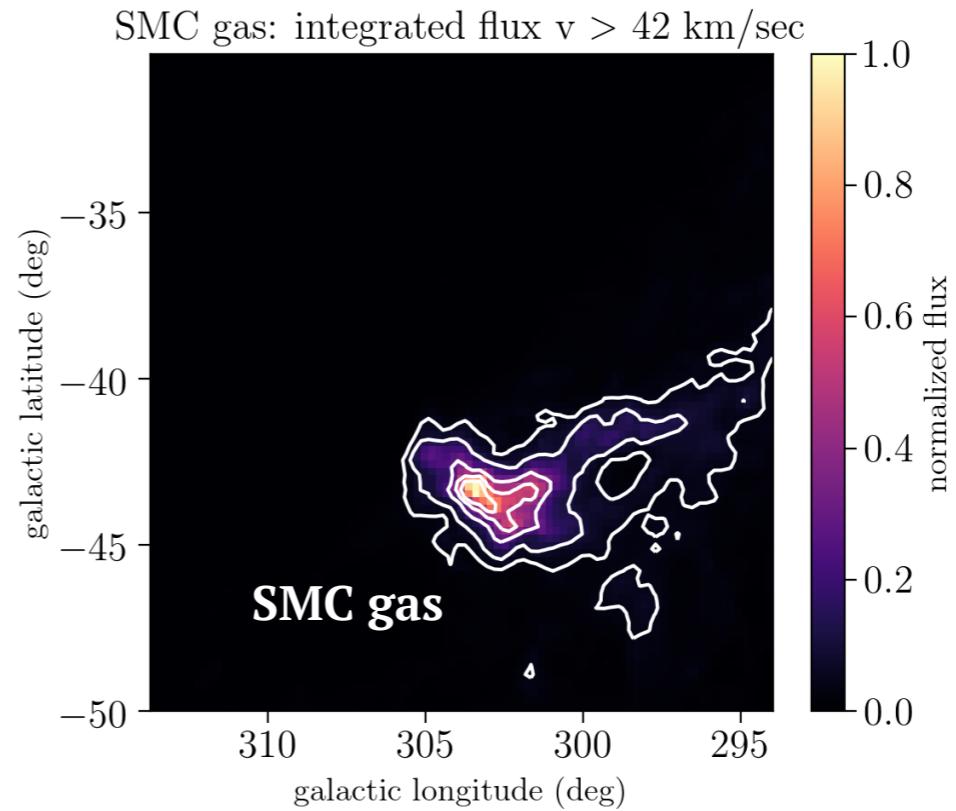
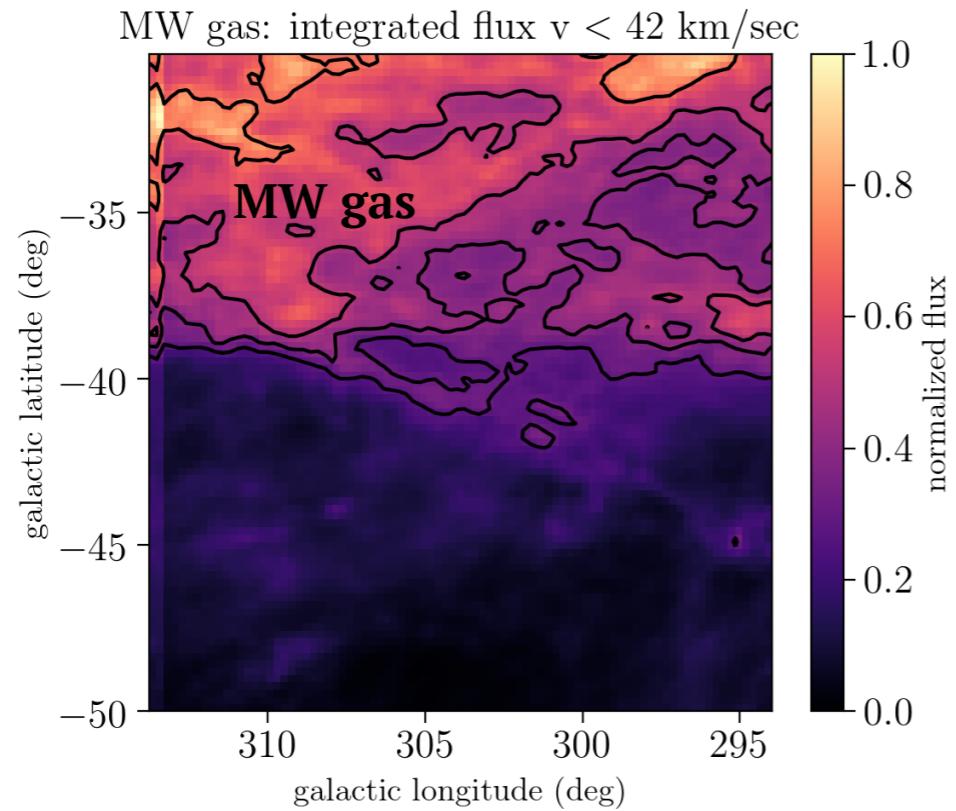
# Examples from Astronomy: redshift sequence



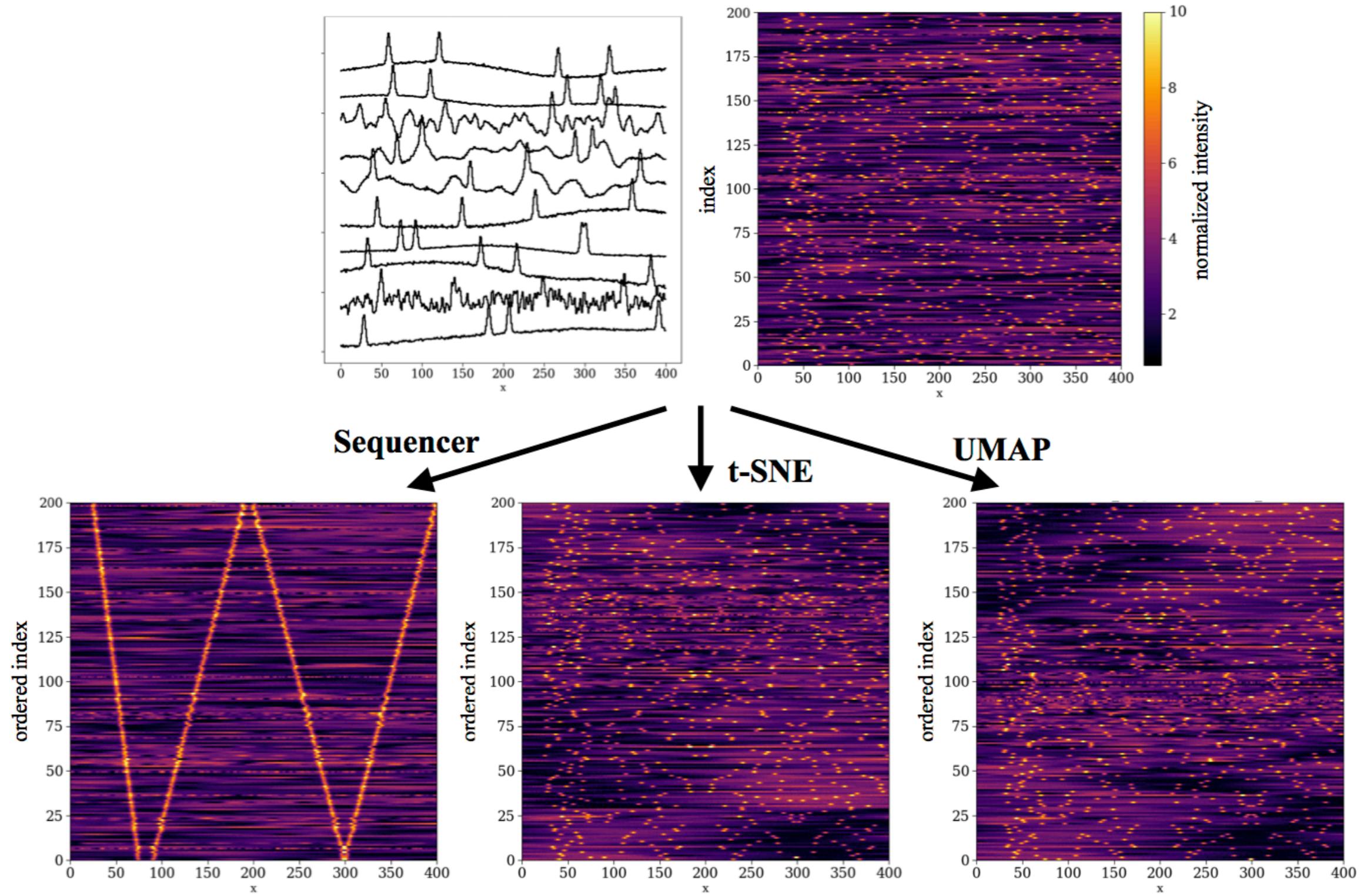
# Example from structural seismology



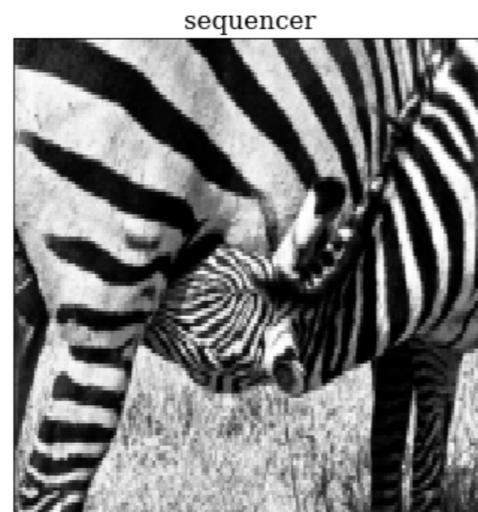
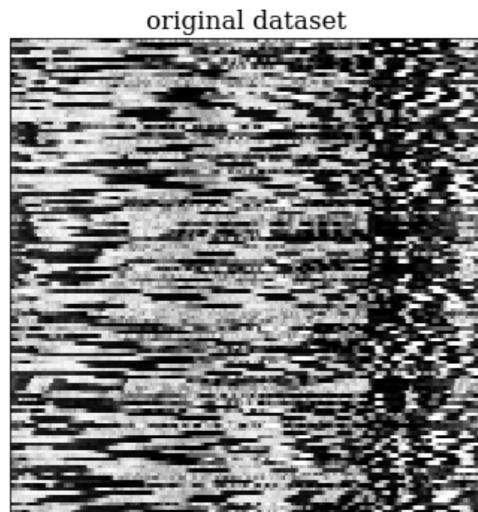
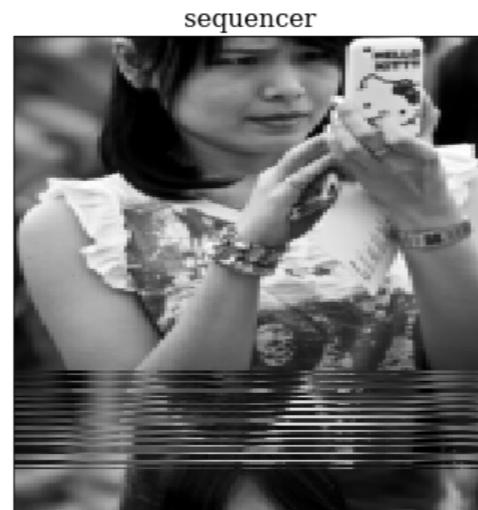
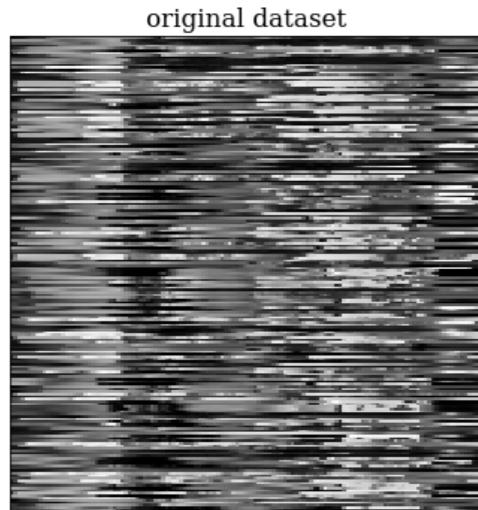
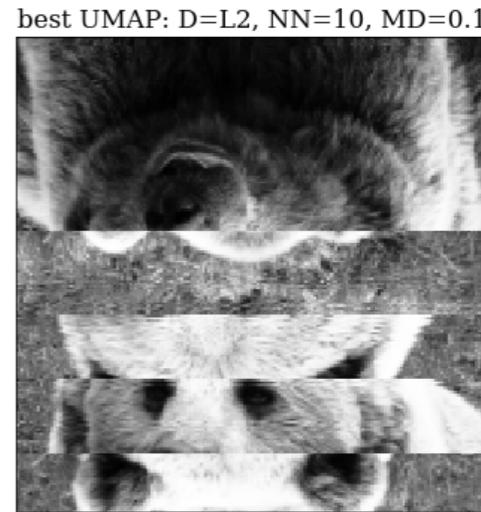
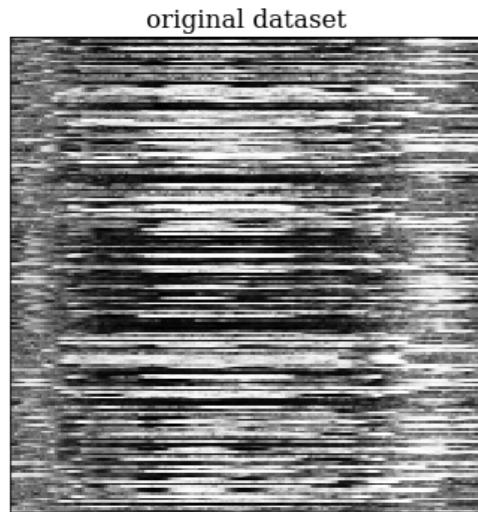
# Color-coding IFU data



# Comparison to tSNE/UMAP



# Comparison to tSNE/UMAP

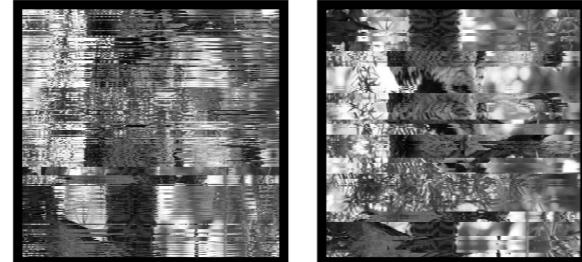


# Comparison to tSNE/UMAP

## Metrics

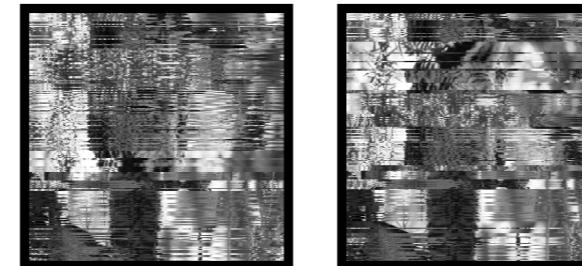
Energy  
Distance

LR=50, P=100,  $\eta=0.17$    LR=100, P=10,  $\eta=0.26$



Earth Mover  
Distance

LR=100, P=100,  $\eta=0.19$    LR=10, P=50,  $\eta=0.34$



KL  
Divergence

LR=150, P=150,  $\eta=0.2$



LR=100, P=50,  $\eta=0.55$



LR=150, P=10,  $\eta=0.95$

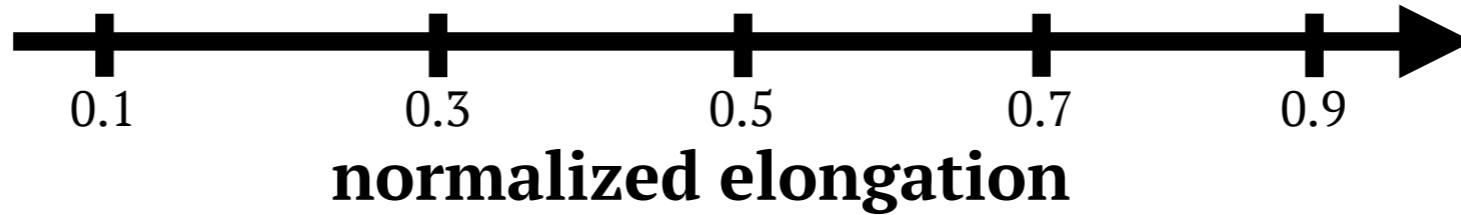


Euclidean  
distance

LR=100, P=150,  $\eta=0.42$    LR=150, P=50,  $\eta=0.62$



LR=10, P=10,  $\eta=1$



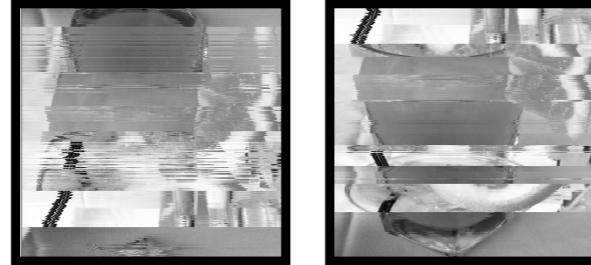
## t-SNE

# Comparison to tSNE/UMAP

## Metrics

Energy Distance ■  
Earth Mover Distance ■  
KL Divergence ■  
Euclidean distance ■

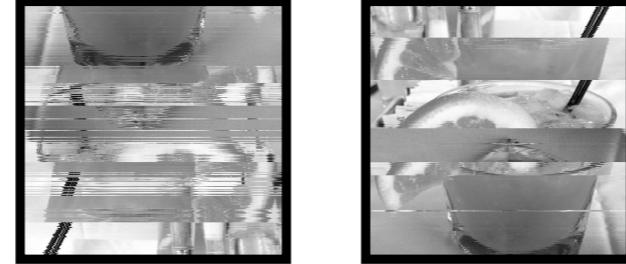
NN=100, MD=0.5,  $\eta=0.15$  NN=10, MD=0.2,  $\eta=0.33$



NN=30, MD=0.5,  $\eta=0.23$  NN=5, MD=0.2,  $\eta=0.46$



NN=100, MD=0.8,  $\eta=0.23$  NN=50, MD=0.05,  $\eta=0.45$

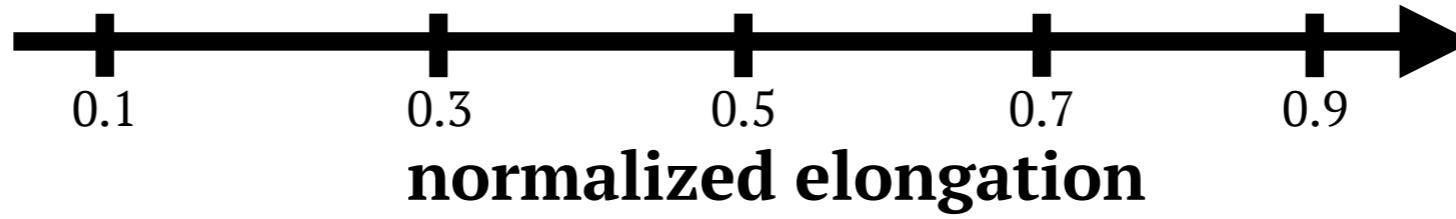


## UMAP

NN=30, MD=0.1,  $\eta=0.81$



NN=30, MD=0.05,  $\eta=0.9$



# Final Remarks

---

- ❖ The input data can be 1D, 2D, or N-D.
- ❖ Comparison with tSNE & UMAP:
  - ❖ The Sequencer output is a 1D embedding, while tSNE and UMAP can provide with a 2D embedding.
  - ❖ The hyper-parameters of tSNE and UMAP can be optimized using the same trick as for the Sequencer.
- ❖ The code is available at: <https://github.com/dalya/Sequencer/>.
- ❖ Online interface: <http://sequencer.org>.