

What is this about?

Task

❖ Spectral Energy Distributions (SEDs) :

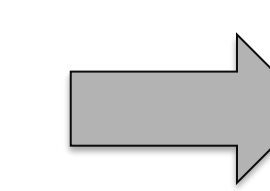
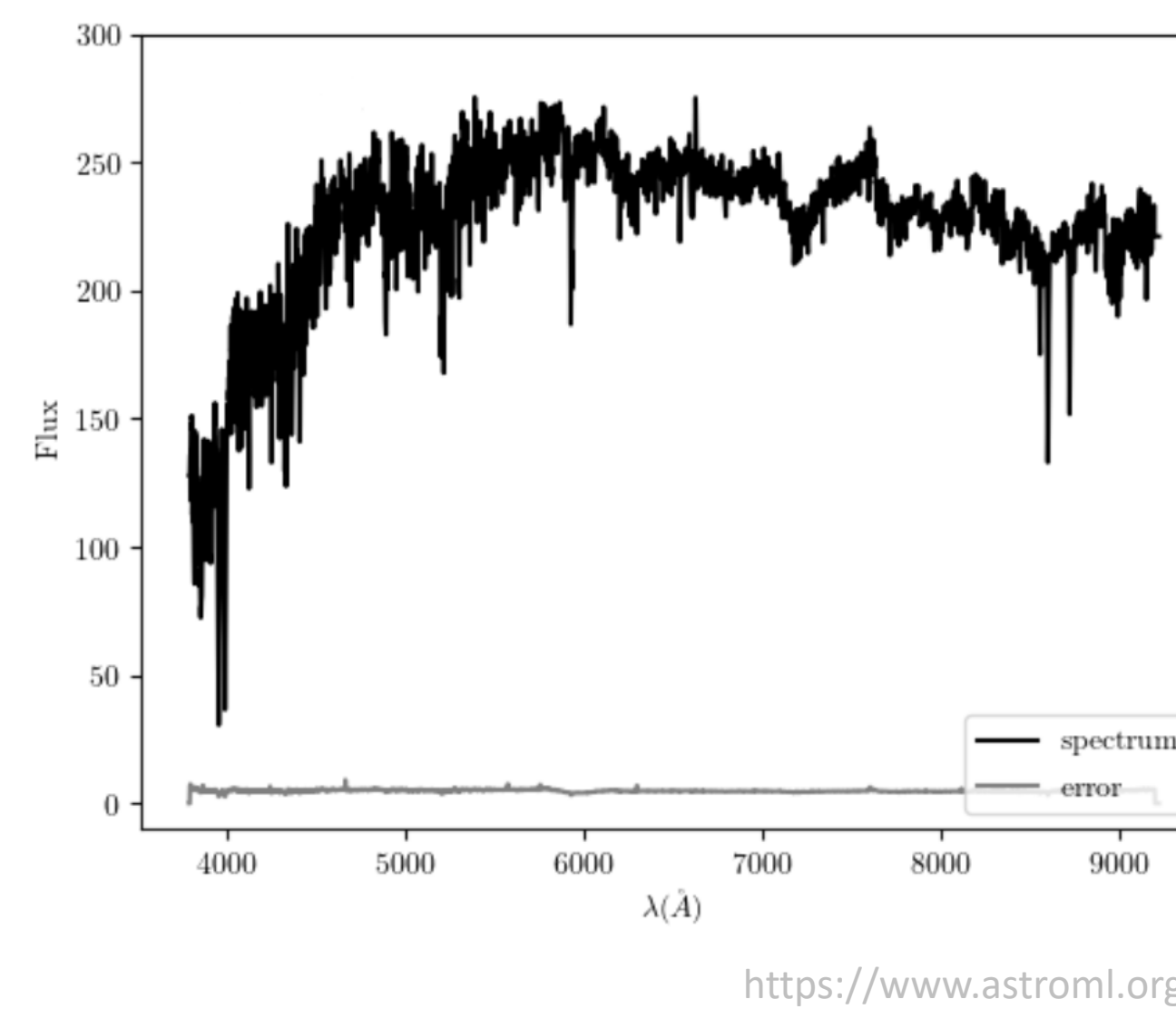
- Energy measurements of astronomical objects, for different wavelengths
- Useful to discover properties and evolution of astronomical objects
- Noise/Interference/Expensive Measurement Procedure => erroneous/incomplete samples

❖ Goal:

- Correct/Enhance SEDs via Deep Learning => generalization + robustness
- Idea: Deep Generative Model as Structural Prior
- Properties such as high-frequency and irregularity of SEDs => potential challenge for deep learning

3-Step Approach

- 1) Preprocessing step -> create a complete dataset, *iterative PCA*
- 2) Train a Deep Generative Model
- 3) Test-time Reconstruction from incoming *noisy/incomplete* samples



The Method

Formulation

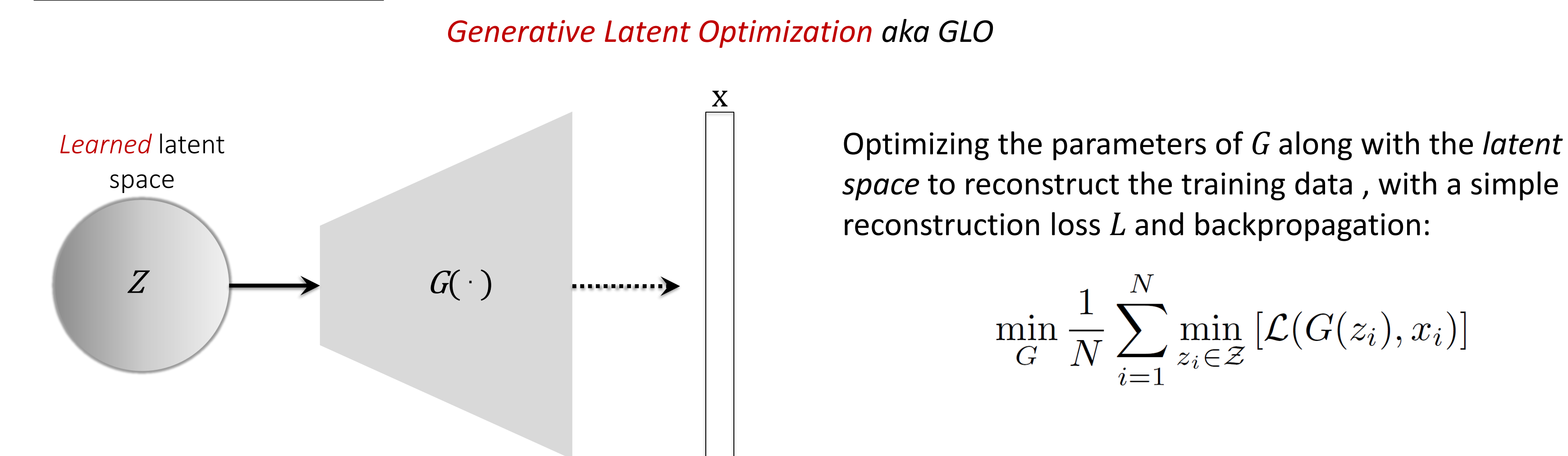
$$y = Ax^* + \eta$$

Given (noisy) linear measurements y (m -dimensional), via a known procedure A Goal to reconstruct x^* (d -dimensional)

=> **Compressed Sensing** ($d \gg m$)

Also, the formulation models realistic **inverse problems** e.g., denoising, inpainting, super-resolution

The Generative Network



To impose **structure in latent space** during training, project the *latent codes* onto the *unit sphere*.

Reconstruction Methods

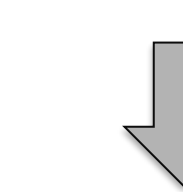
Estimate a latent code by **minimizing the measurement error** (backpropagation) => corresponds to a signal in the domain of G that *approximately agrees with the measurements*

$$\hat{z} = \arg \min_{z \in \mathcal{Z}} \frac{1}{m} \|AG(z) - y\|_2^2$$

(a) Explicit **projection** onto the unit sphere

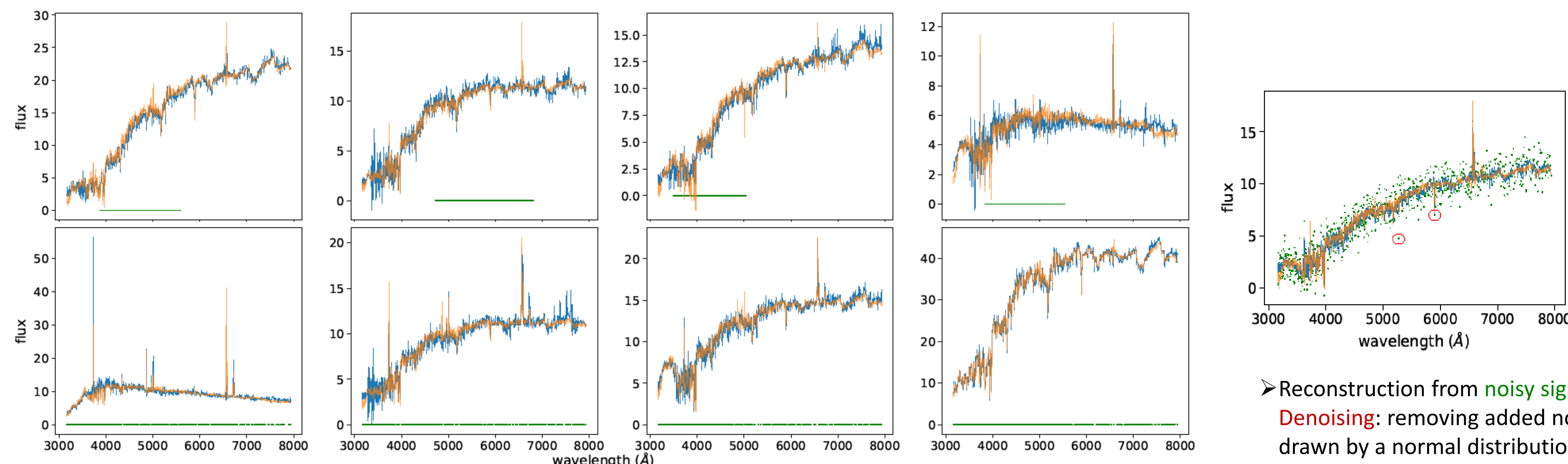
$$\hat{z} = \arg \min_{z \in \mathcal{Z}} \frac{1}{m} \|AG(z) - y\|_2^2 + \lambda \|z\|_2^2$$

(b) Implicit regularization with extra **regularization** term

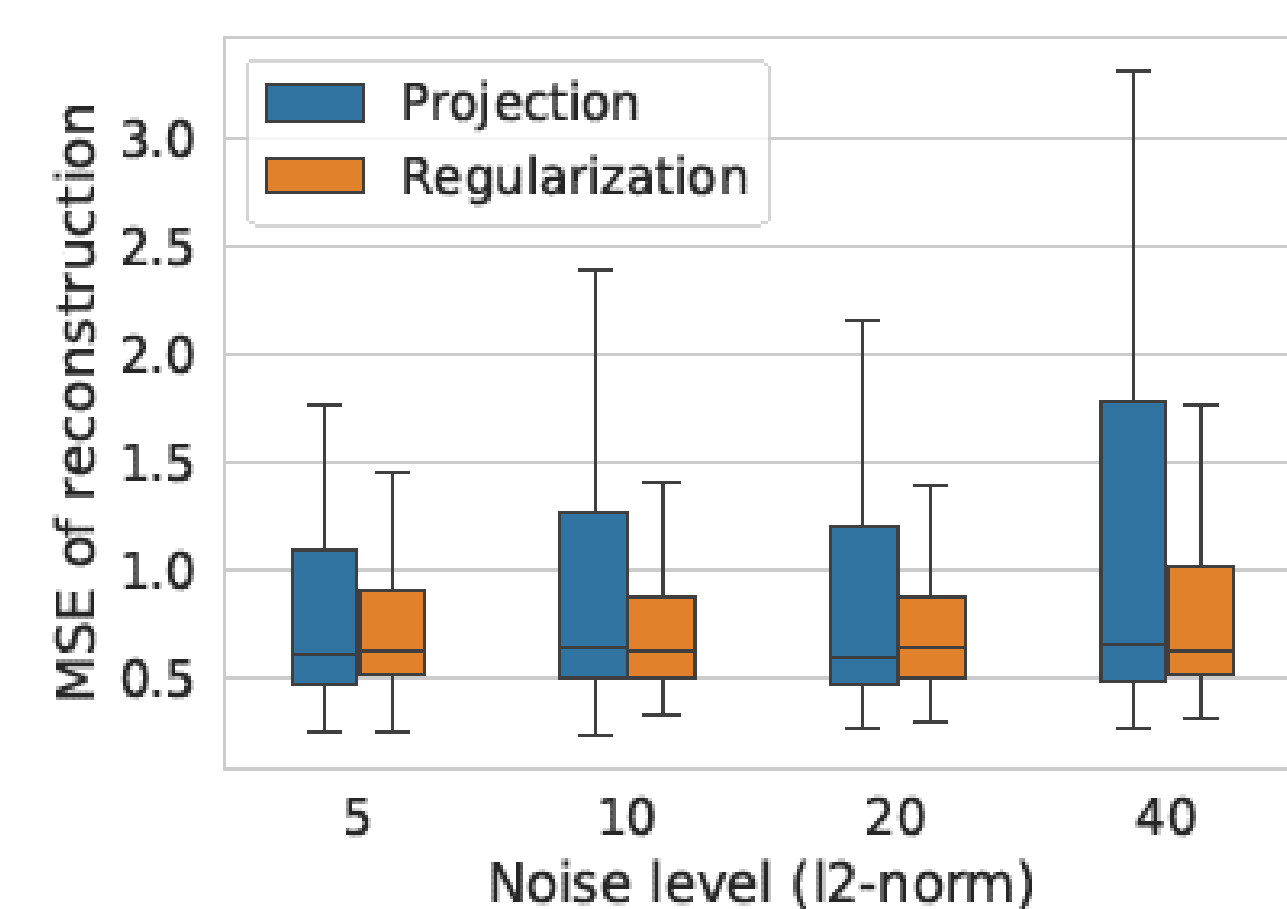
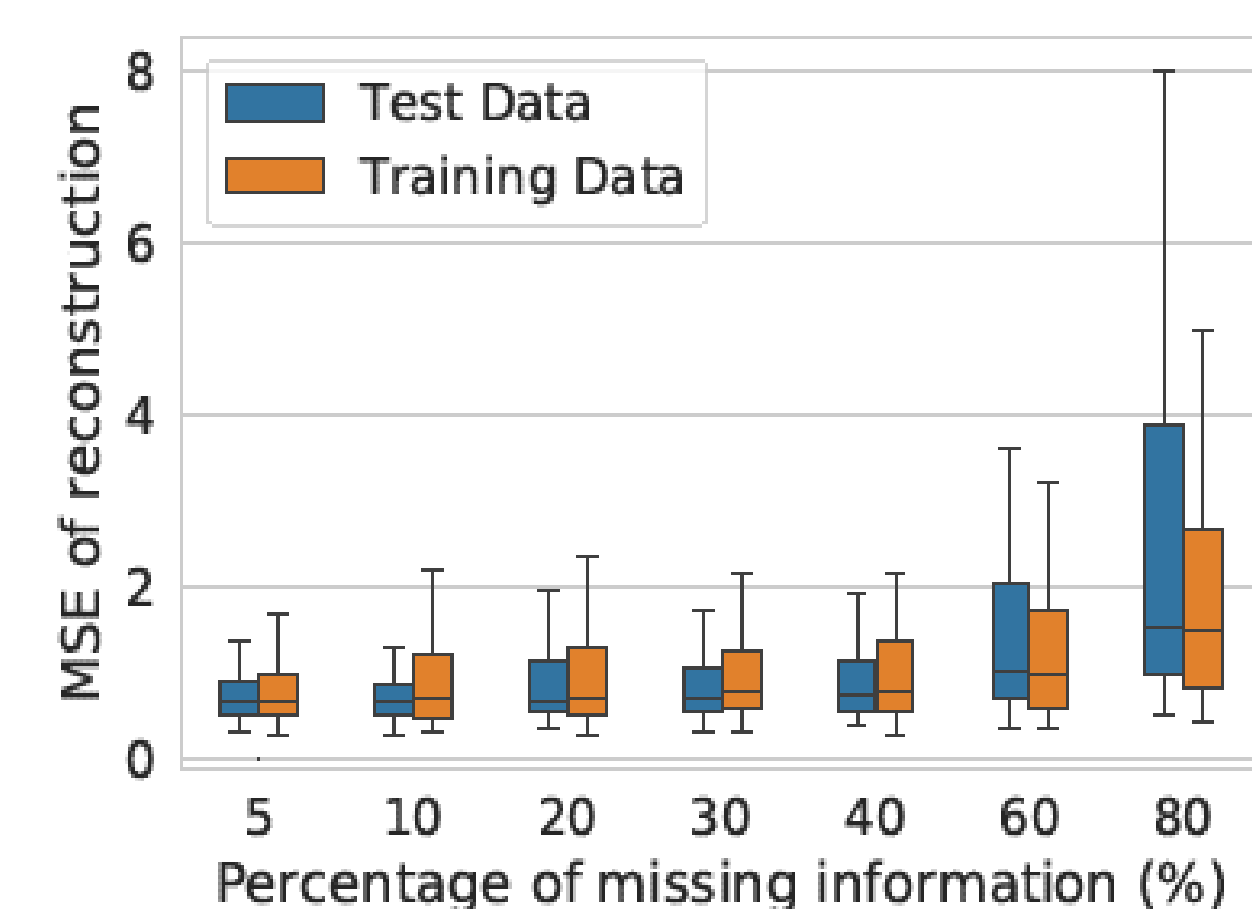


Some Preliminary Results

— Original Signal — Reconstruction • Missing Information



➤ Reconstruction for **40%** missing information
(TOP) **Inpainting**: measurements are missing in a continuous window
(BOTTOM) **Super-resolution**: randomly selected measurements are missing



➤ MSE of reconstruction for 100 randomly selected SEDs on:

(LEFT) **Inpainting**, for different levels of missing information (%) => displays **generalization** capability

(RIGHT) **Denoising**, for different levels of added noise (l2-norm) => **regularization** appears more **flexible**

The Setup

Data

- ❖ Sloan Digital Sky Survey or **SDSS Corrected Spectra** dataset by the astroML library
 - SEDs of **4000 galaxies**, moved to restframe
 - Preprocessed with *iterative PCA*
 - Resampled to 1000 wavelengths (3000 - 8000Å)

- ❖ The original dataset consists of innately incomplete and/or corrupted SEDs => **lack of ground truth**

Training

- ❖ Training setup:
 - Architecture: 7-layer feed-forward network with leakyReLU activations, **50-dim** latent space
 - Optimization: Adam, 10000 epochs, batches of 64 SEDs
 - Learning rate: 0.1 for the network - 0.01 for the latent codes
 - Reconstruction loss: Mean Squared Error (MSE)

- ❖ Reconstruction setup similar to training, *limited to 1000 epochs*.