Generative Machine Learning for Extreme Climate Scenarios

Dario Oliveira, Daniela Szwarcman, Jorge Luis Guevara Diaz, Campbell Watson, Maysa Macedo, **Bianca Zadrozny**

IBM Research

Introduction

- The frequency, duration and intensity of extreme weather events have increased as the climate system warms.

- In this context, the stochastic generation of extreme weather scenarios is increasingly vital for **risk and resilience**.

  - One example is flood risk – different scenarios of extreme precipitation are needed to drive flood models.

- Extreme heatwaves
- Longer droughts
- Flooding
- More frequent tropical cyclones
- Melting ice caps
- Wildfires

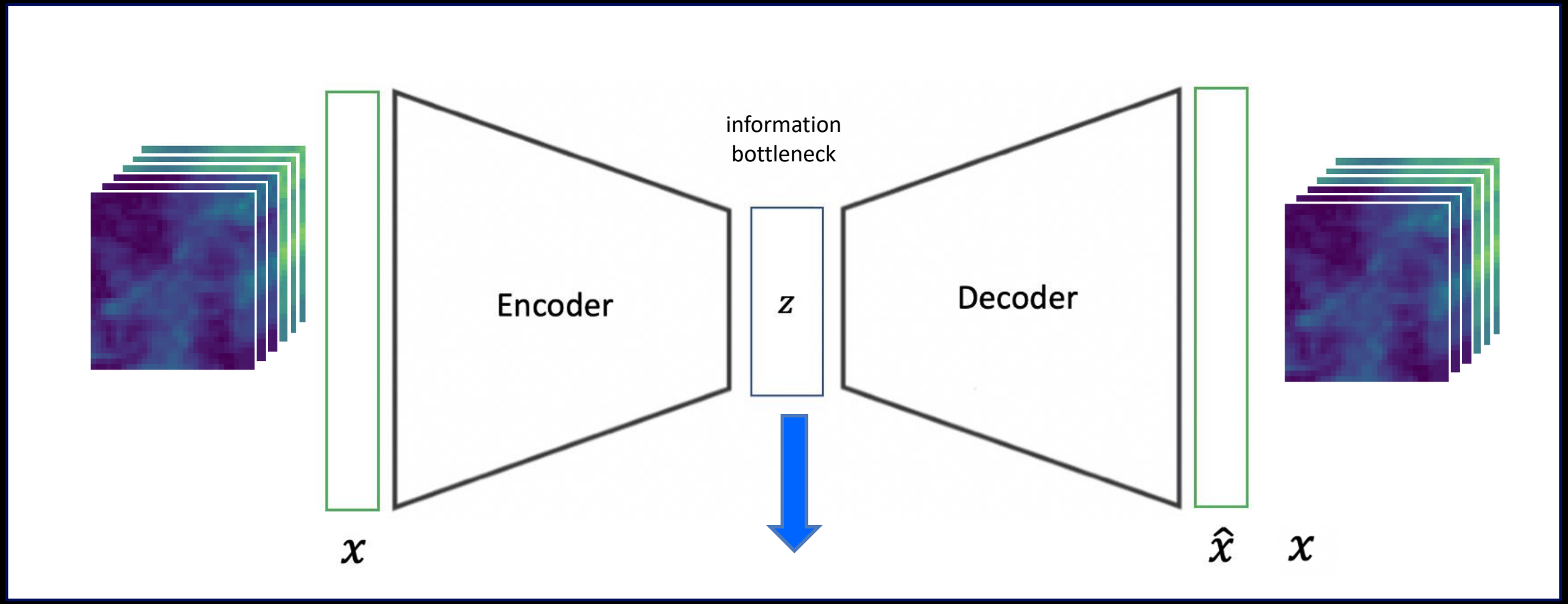
Introduction

- Creating high-quality variations of under-represented extreme events remains a challenge for current weather generators.
  - Here we present a new approach based on variational auto-encoders (VAEs) with a quantized reconstruction loss
  - We present initial results for a case study of precipitation in Palghar, India.

- Extreme heatwaves
- Longer droughts
- Flooding
- More frequent tropical cyclones
- Melting ice caps
- Wildfires

Auto-Encoders

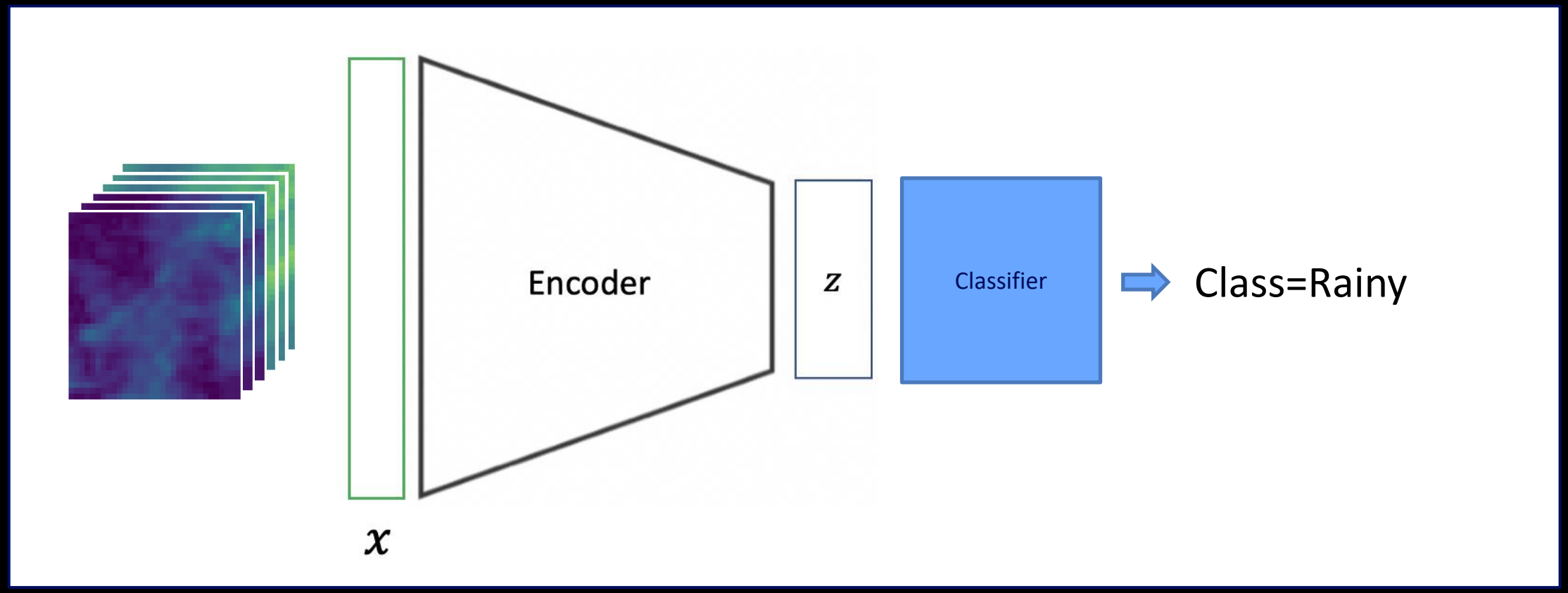
Auto-encoders: models that learn to reconstruct the input data



Z is a very compact, efficient and
discriminant representation of the input known as latent space

Auto-Encoders

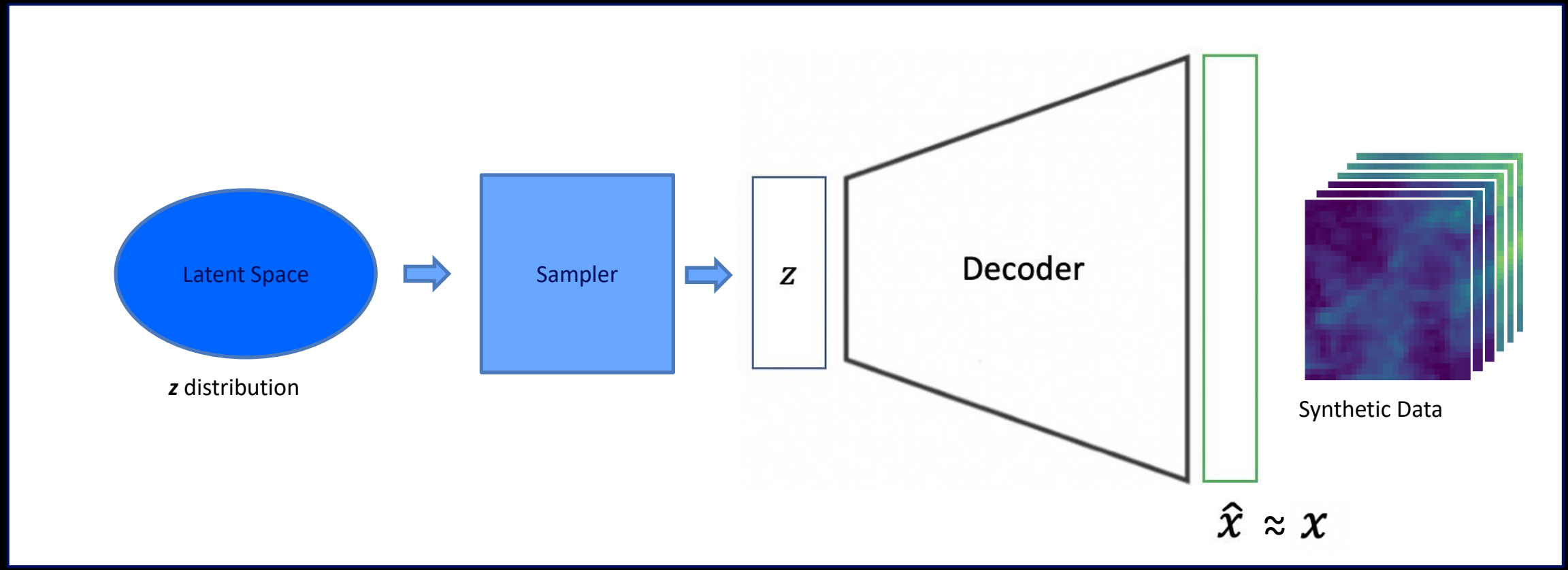
Auto-encoders: models that learn to reconstruct the input data



z can be used for efficient feature extraction

Auto-Encoders

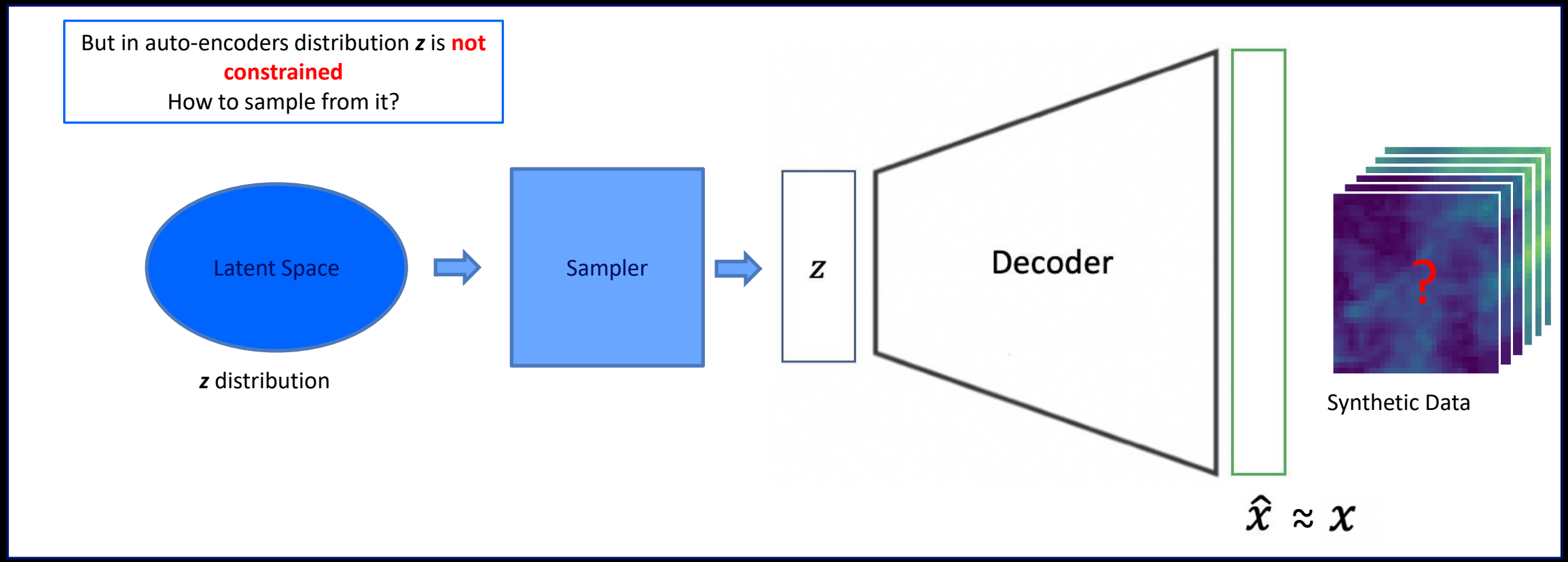
Auto-encoders: models that learn to reconstruct the input data



z can be used for <u>stochastic synthesis</u>

Auto-Encoders

Auto-encoders: models that learn to reconstruct the input data



z can be used for <u>stochastic synthesis</u>

Variational Auto-Encoders - VAE

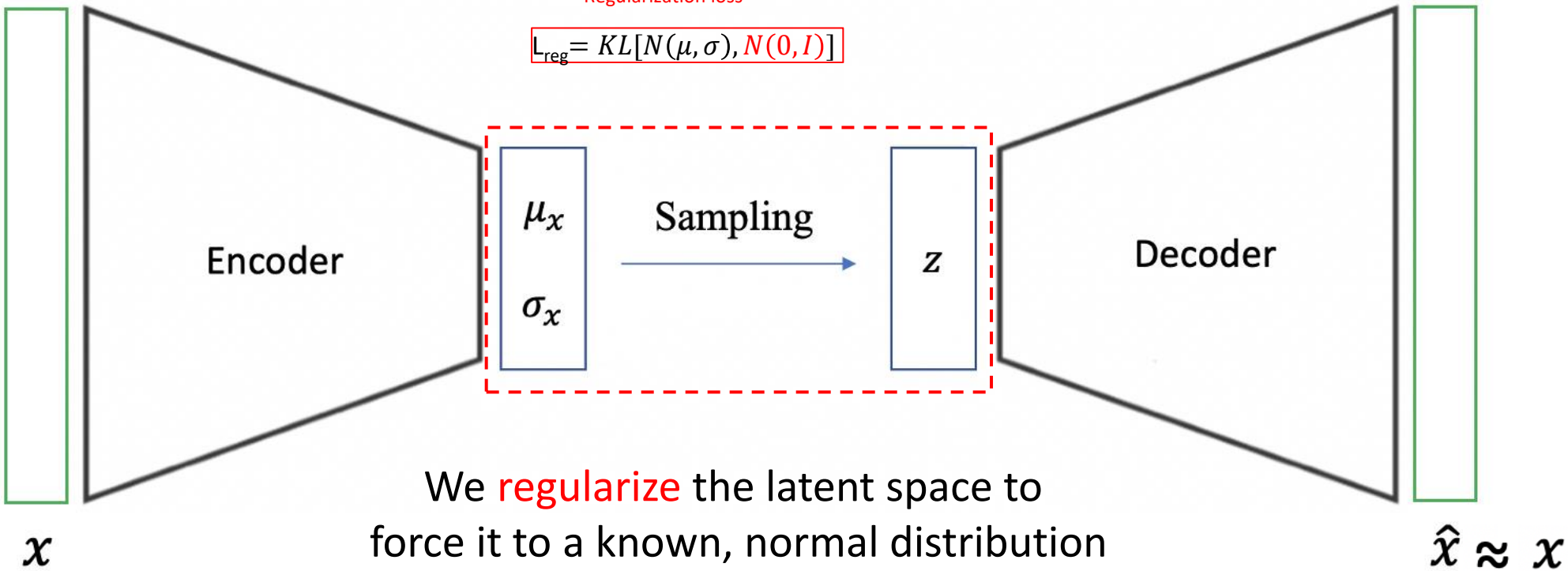
Variational Auto-Encoders: constraining latent space distribution

Reconstruction loss

$L_{rec} = \|x - \hat{x}\|$

Regularization loss

$L_{reg} = KL[N(\mu, \sigma), N(0, I)]$



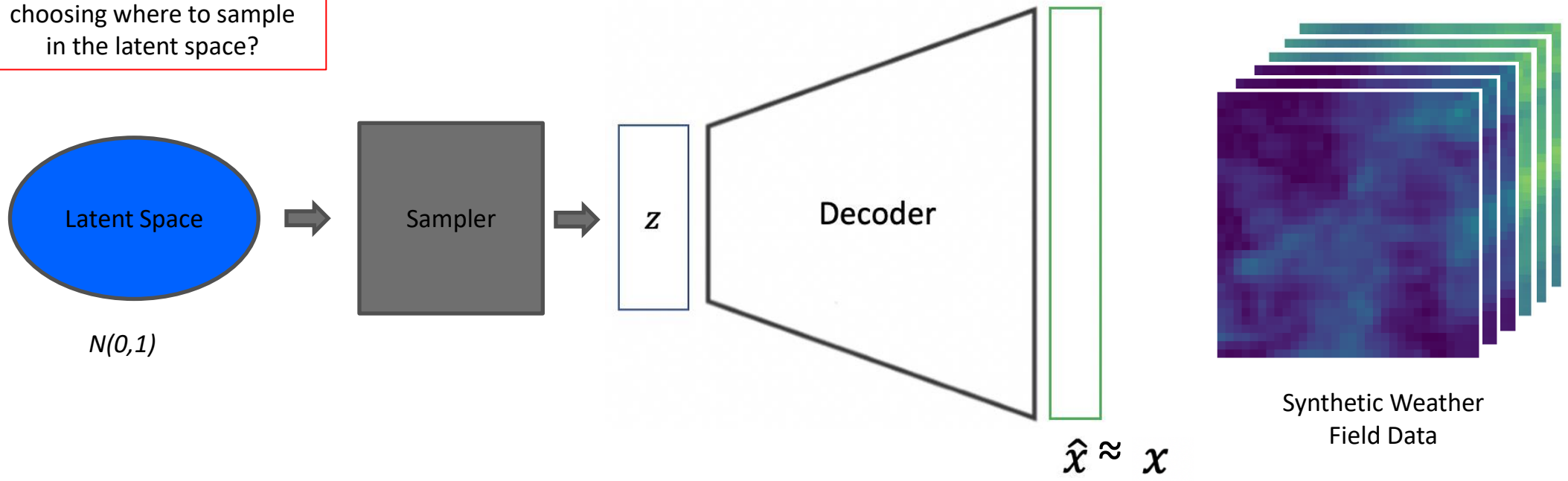
We regularize the latent space to
force it to a known, normal distribution

Now, we can sample from *N(0,I)* for realistic synthesis!

Controling Weather Field Data Synthesis

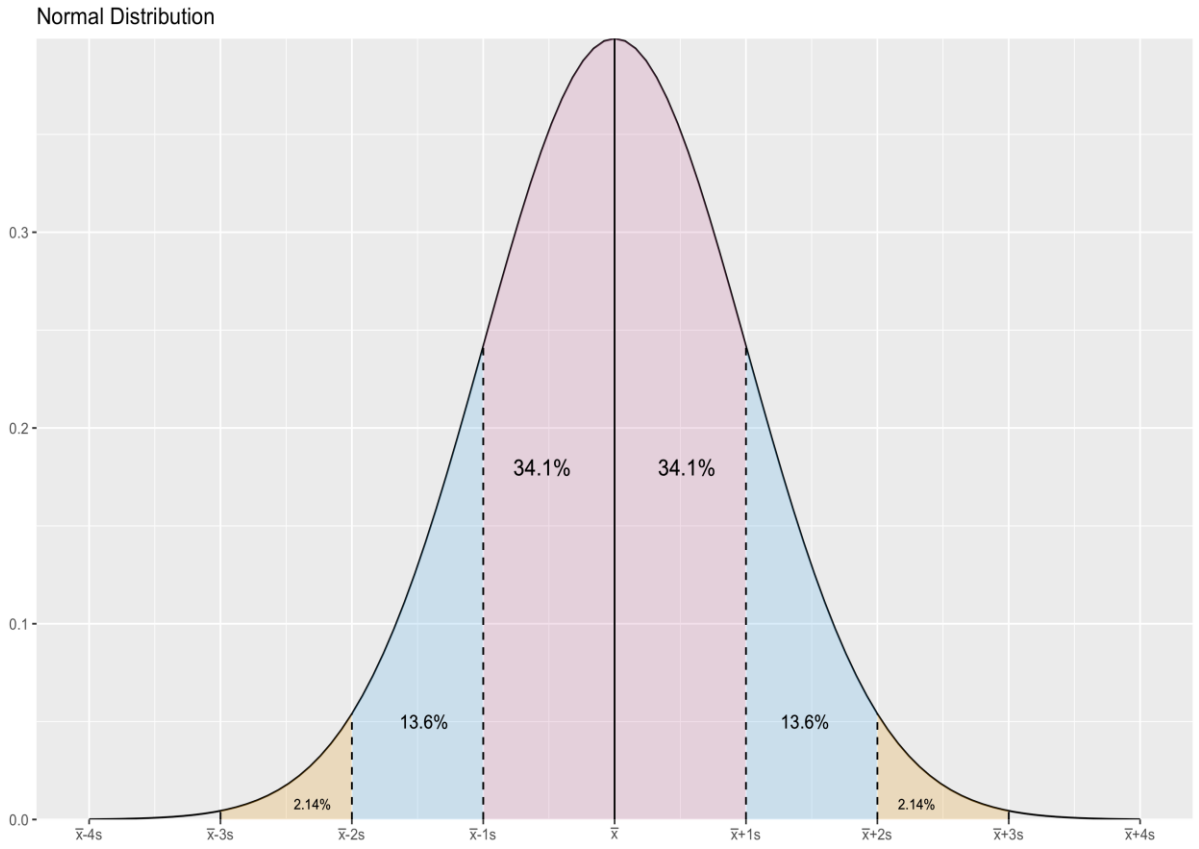
Using trained VAE weather field data synthesis

Can we control synthesis choosing where to sample in the latent space?

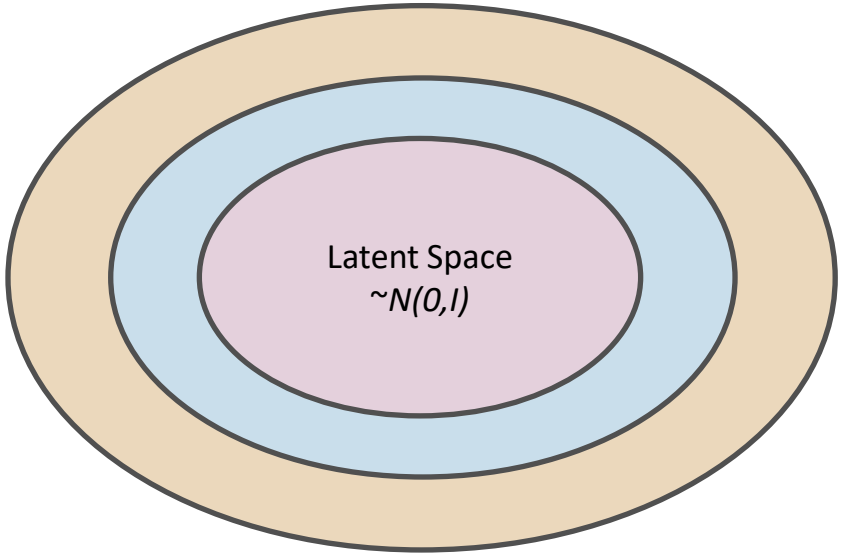


Controling Weather Field Data Synthesis

How is the VAE latent space distribution? ~ *N(0,I)*

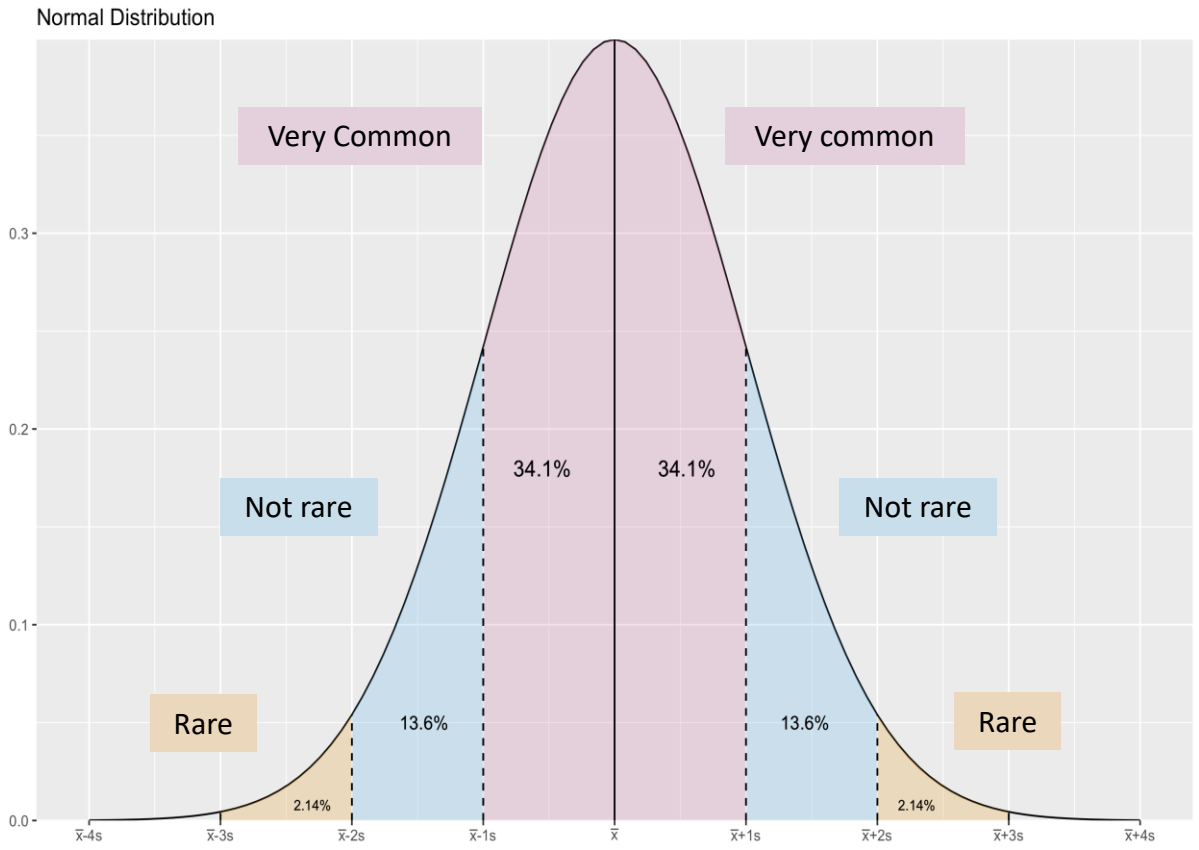


But it is N-dimensional

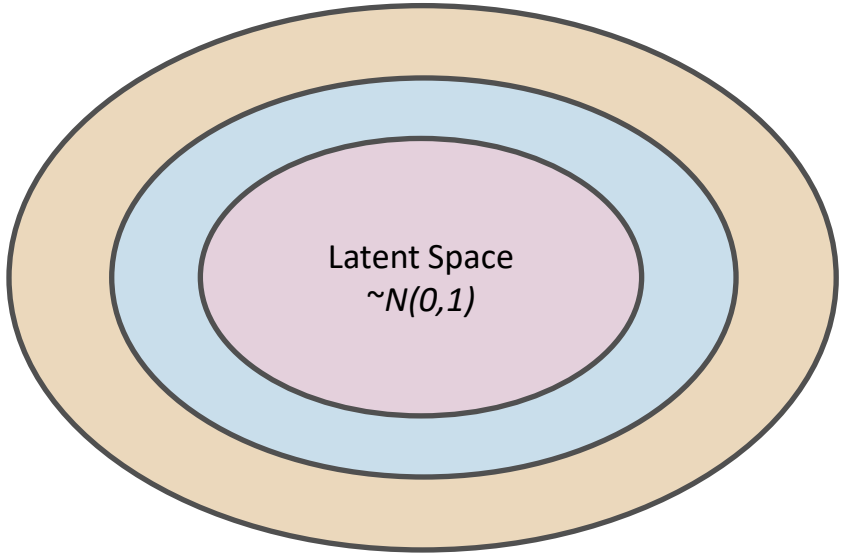


Controling Weather Field Data Synthesis

How are weather events distributed in the latent space ~ *N(0,1)*?

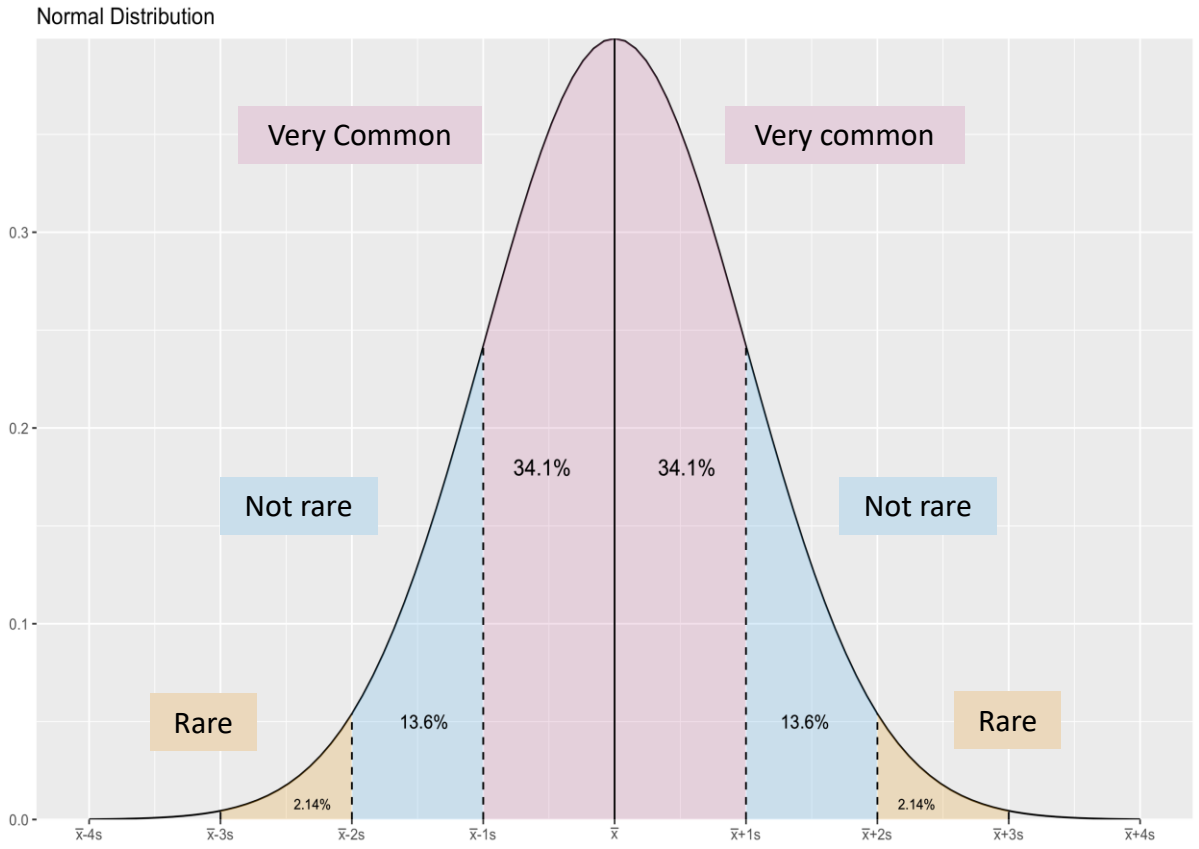


Very common events will necessarily be located near the distribution mean and rare events will be located far from it

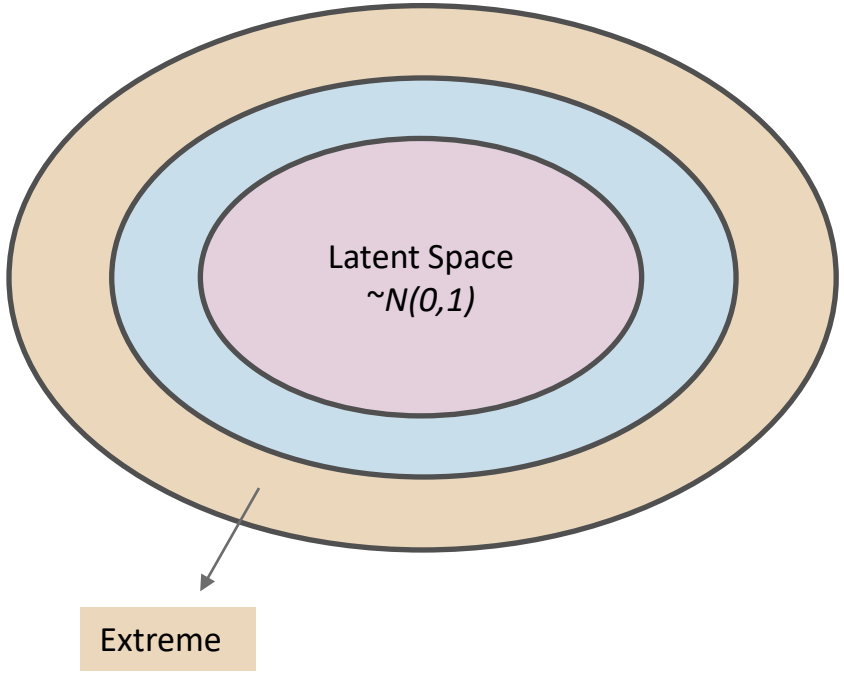


Controling Weather Field Data Synthesis

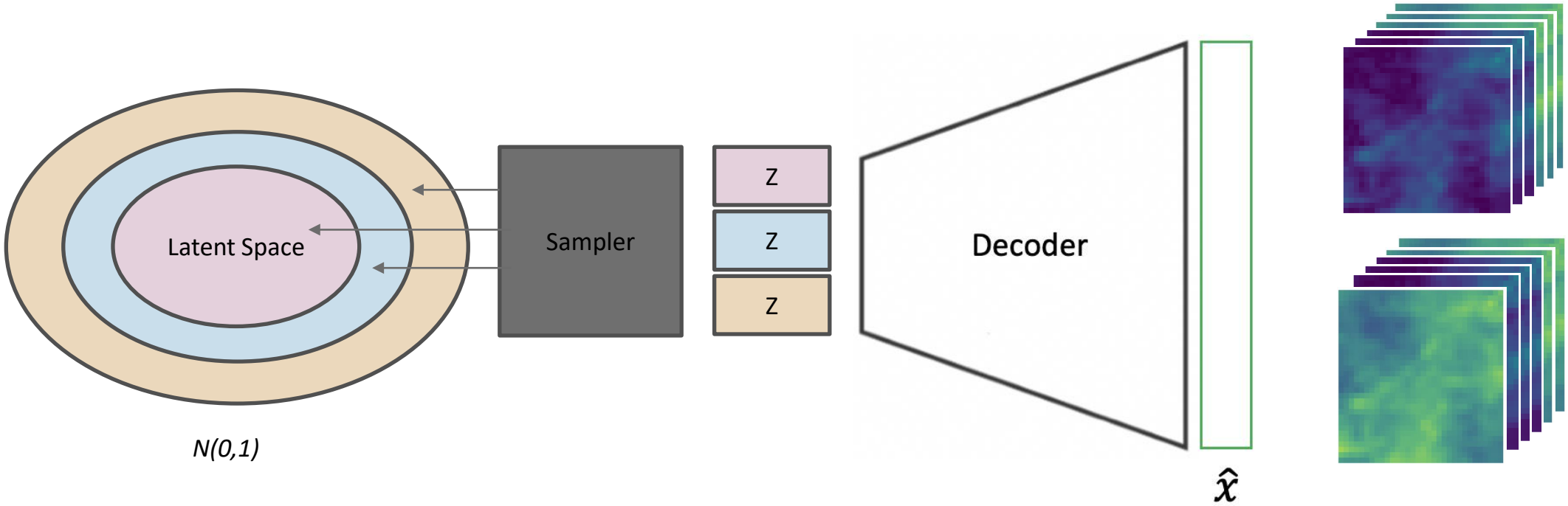
How are climate events distributed in the latent space ~ *N(0,1)*?

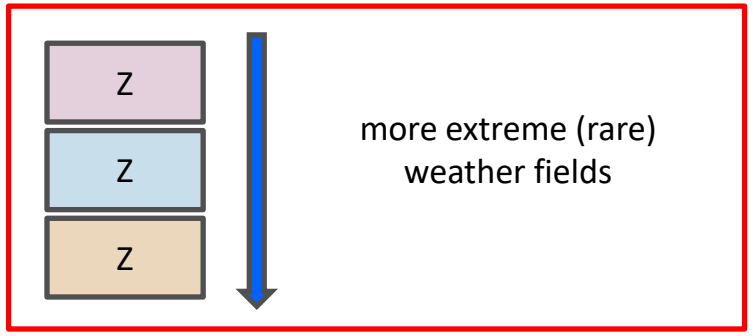


But extreme weather events are also usually rare!



Controling Weather Field Data Synthesis

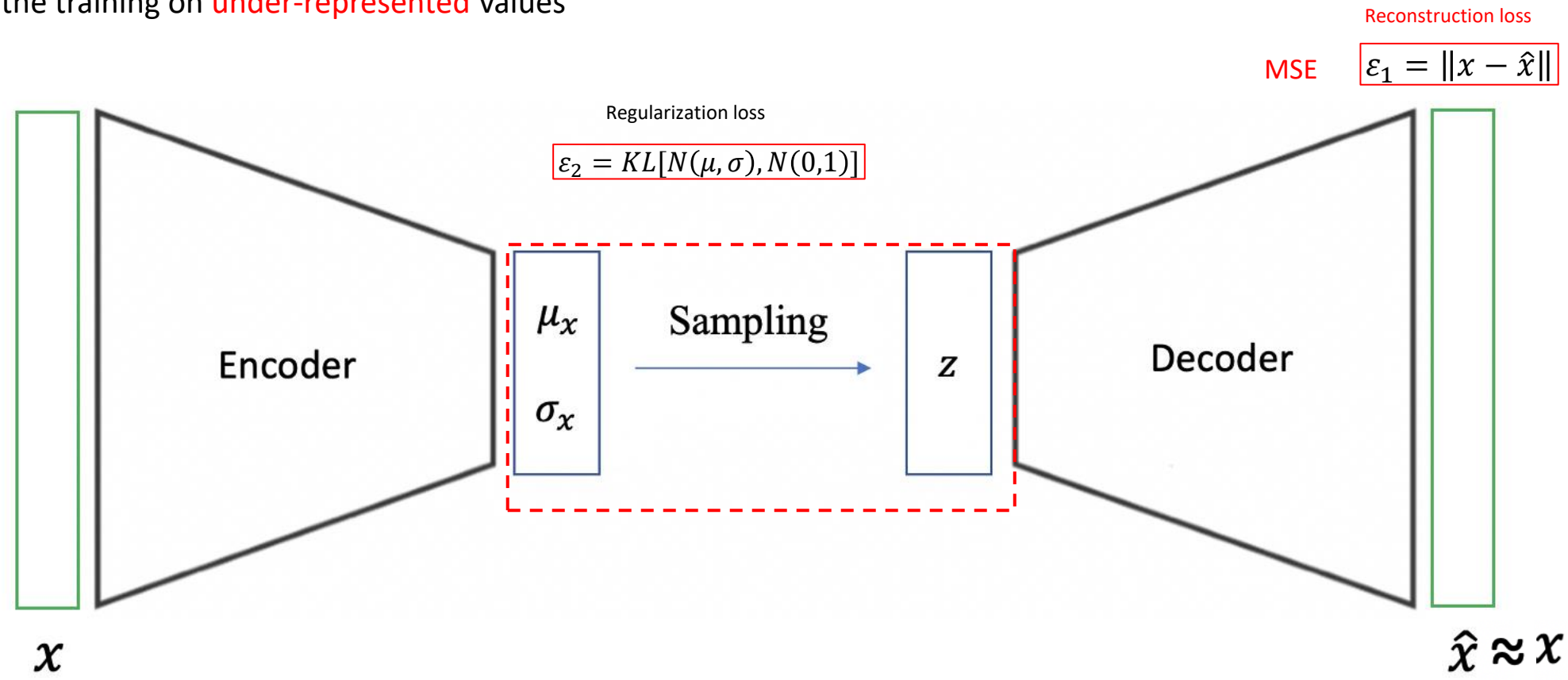




We do not need to define exactly what extreme means!
extreme=rare

Quantized reconstruction losses

- The standard VAE uses MSE as the reconstruction loss
- We propose two different quantized reconstruction losses to focus the training on under-represented values



Quantized reconstruction losses

- The first approach penalizes the reconstruction loss according to the observed frequency
- It bins the data according to frequency and averages the losses in each bin.

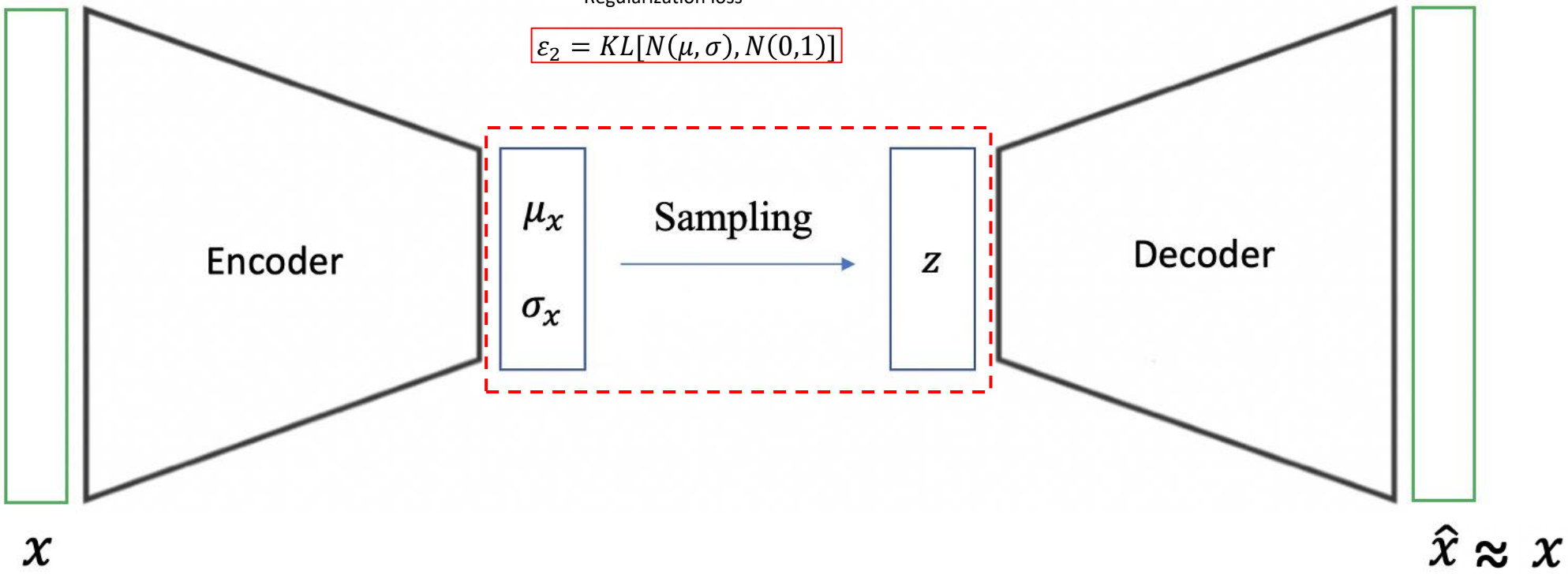
1) Quantized MSE

Reconstruction loss

$$\varepsilon_{1qt} = \sum_{x \in \chi} \sum_{j}^{B} \frac{1}{|\Omega_j|} \sum_{i \in \Omega_j} \|x_i - \hat{x}_i\|$$

Regularization loss

$$\varepsilon_2 = KL[N(\mu, \sigma), N(0,1)]$$



Quantized reconstruction losses

The second approach weights the quantized reconstruction loss by the inverse likelihood of a given value based on the histogram frequency distribution

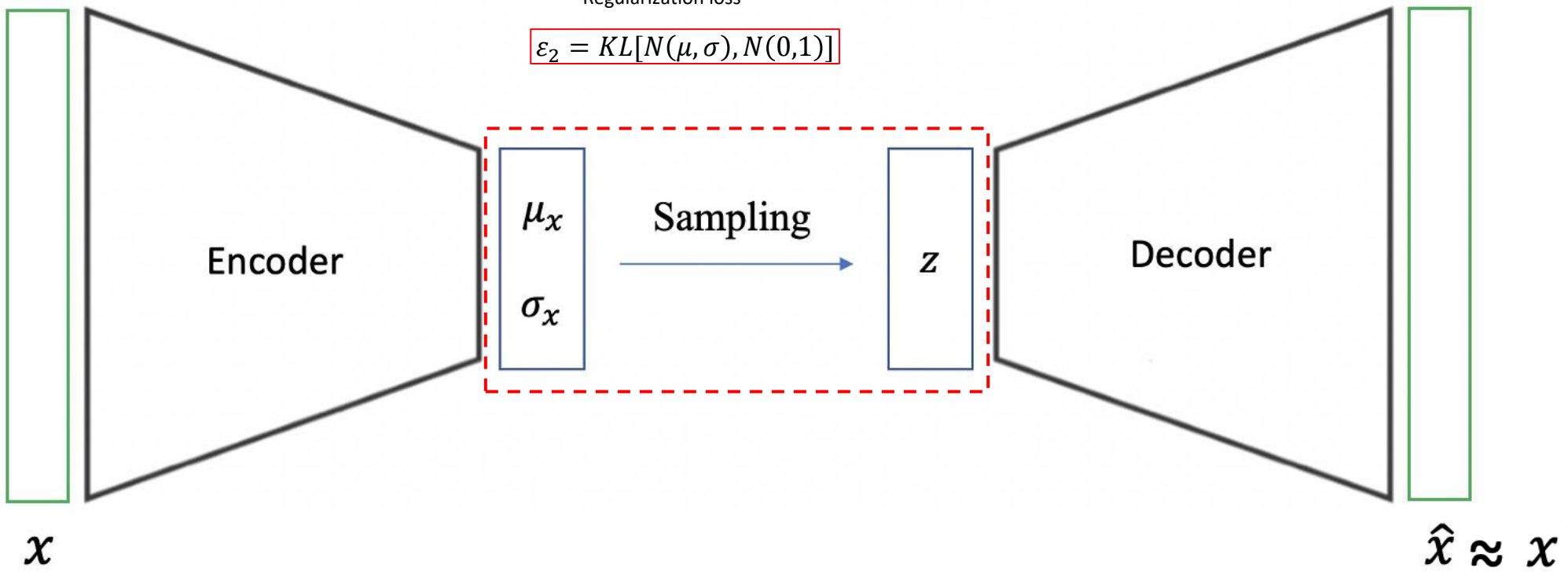
2) Weighted Quantized MSE

Reconstruction loss

$$\varepsilon_{1wqt} = \sum_{x \in \chi} \sum_{j}^{B} \frac{1}{|\Omega_j|} \sum_{i \in \Omega_j} \omega_j(x_i) \cdot \|x_i - \hat{x}_i\|$$

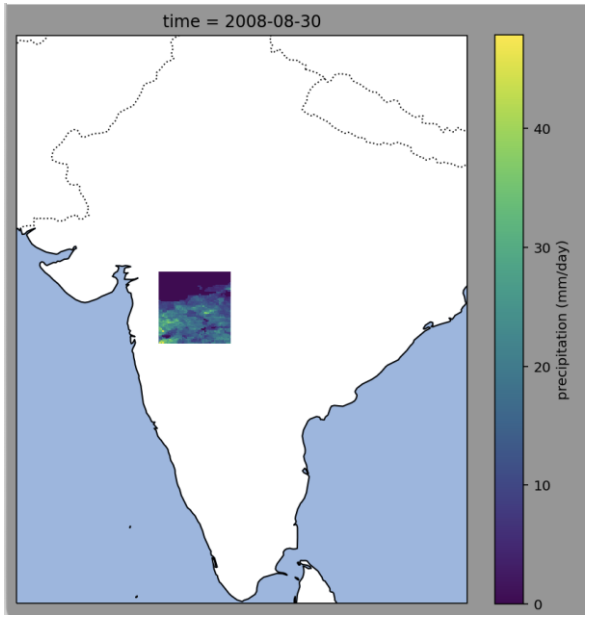
Regularization loss

$$\varepsilon_2 = KL[N(\mu, \sigma), N(0,1)]$$

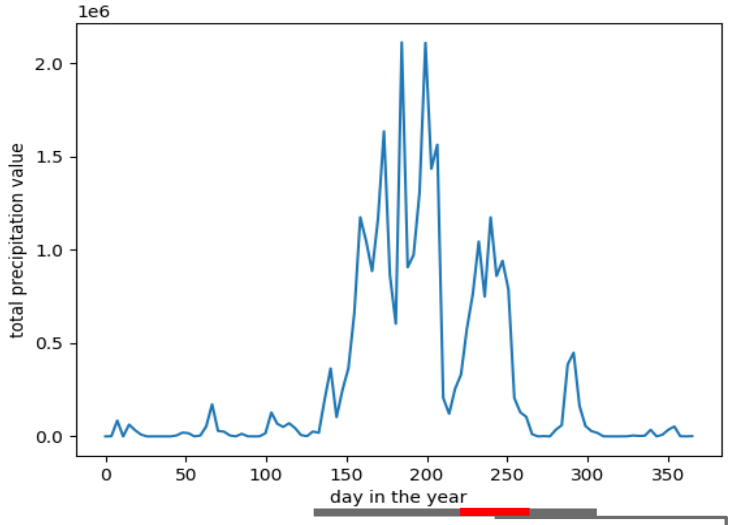


Case study: Palghar Monsoons

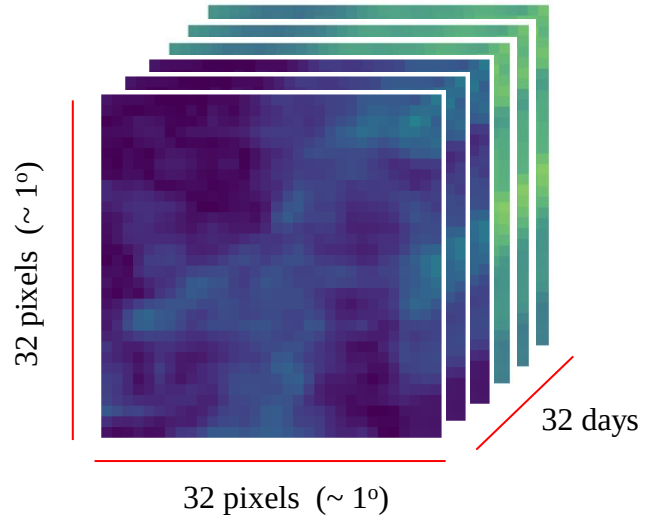
Monsoons in Palghar, Southwest India



Precipitation in a random year



Weather Field Sample



CHIRPS dataset[1] from Palghar, India

39 years of data: 1981-01-01 to 2020-01-01

Training: 1981-2010

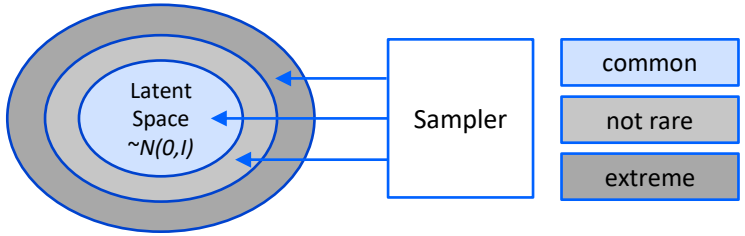
Testing: 2010-2020

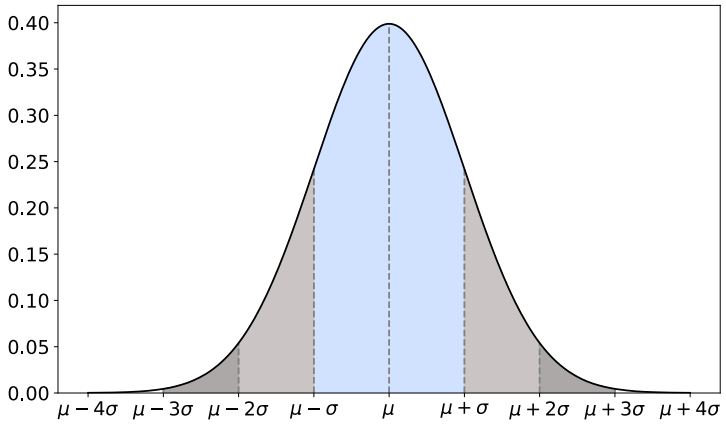
sampling sequences of 32 days

[1] Chris Funk, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell, et al.
The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data*, 2(1):1–21, 2015.

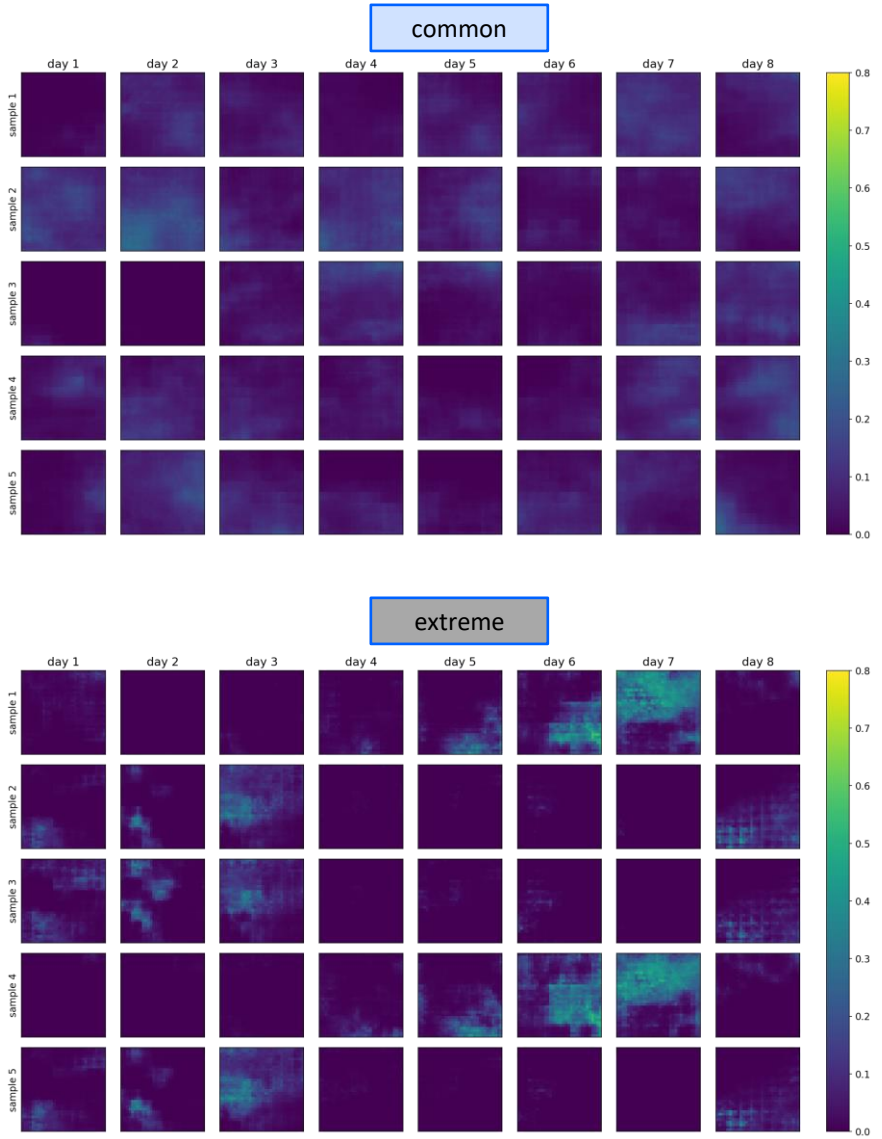
Case Study: Palghar Monsoons - Results

Experiment: with a trained variational autoencoder, control the sampling of *Z* based on *N(0,1)* quantiles

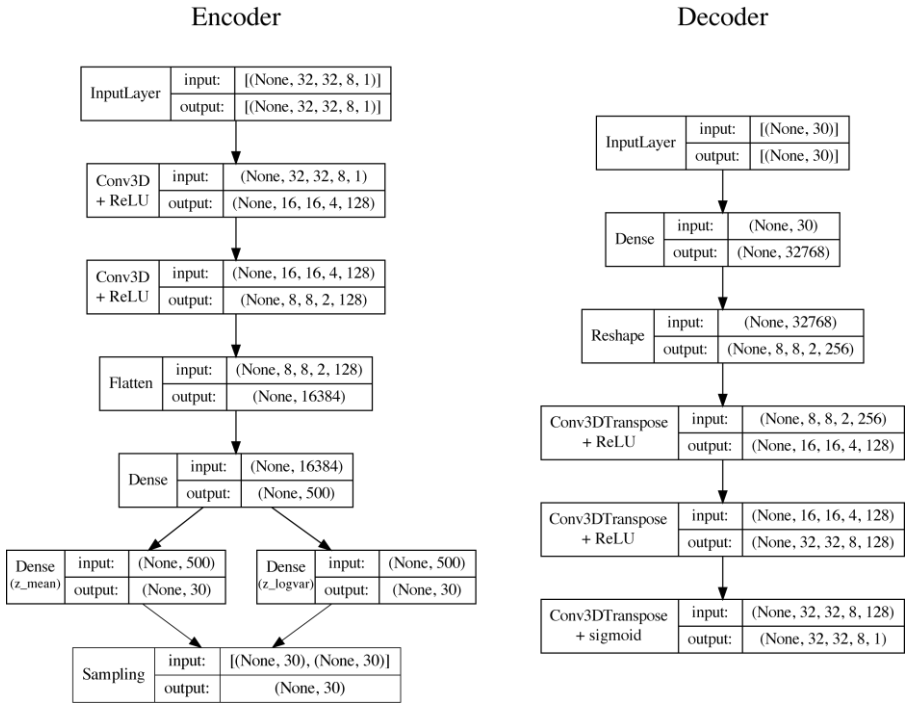




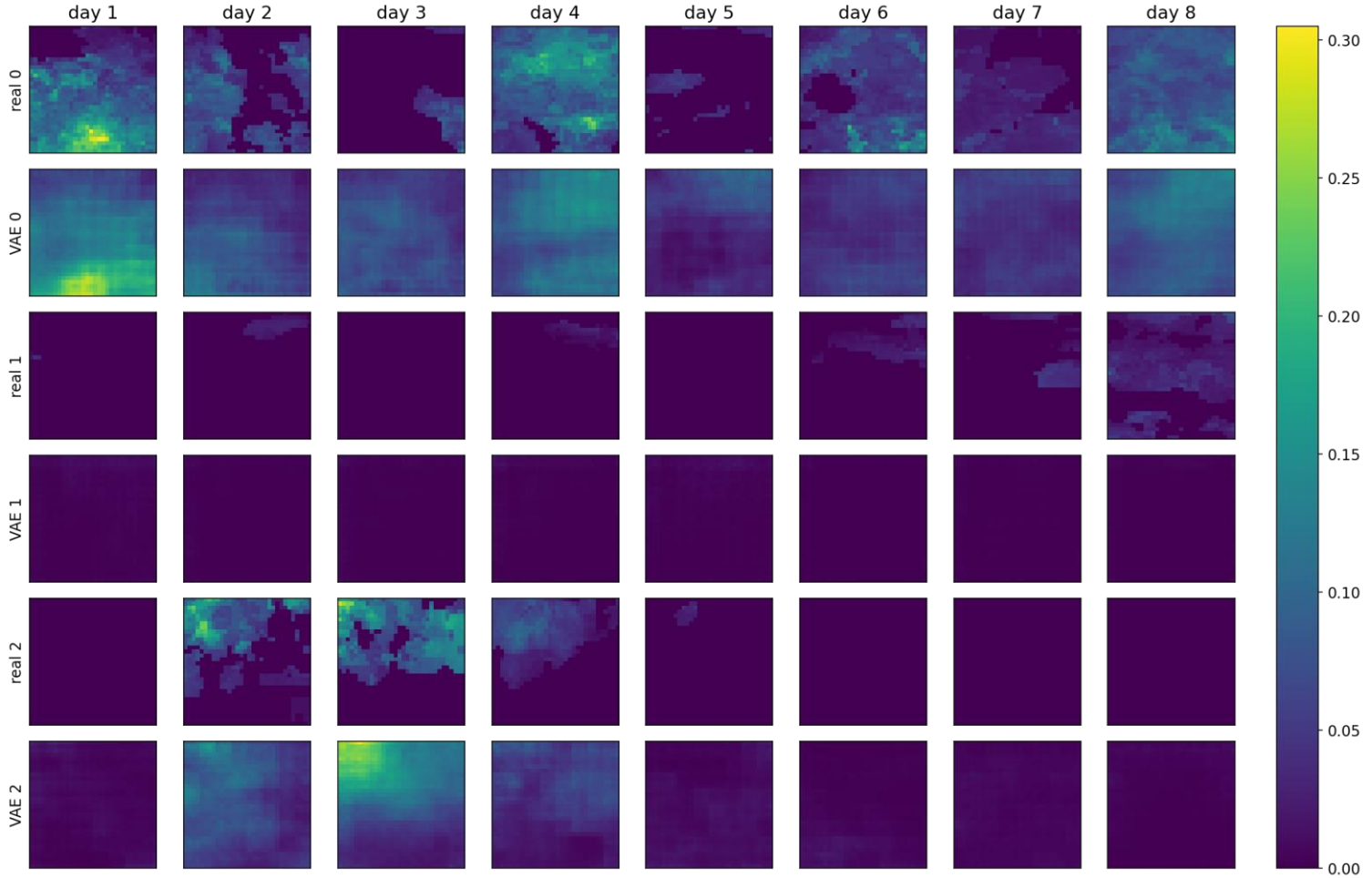
generate weather variables with values corresponding to different level of "extremeness"



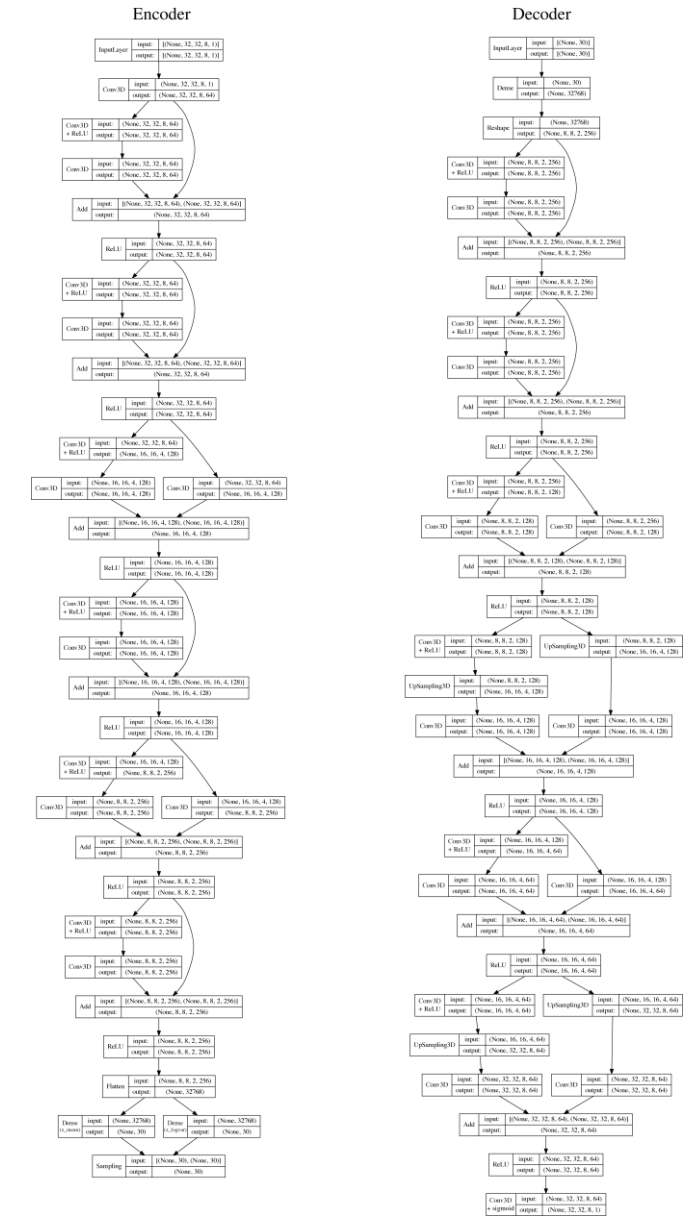
Case Study: Palghar Monsoons - Reconstructions

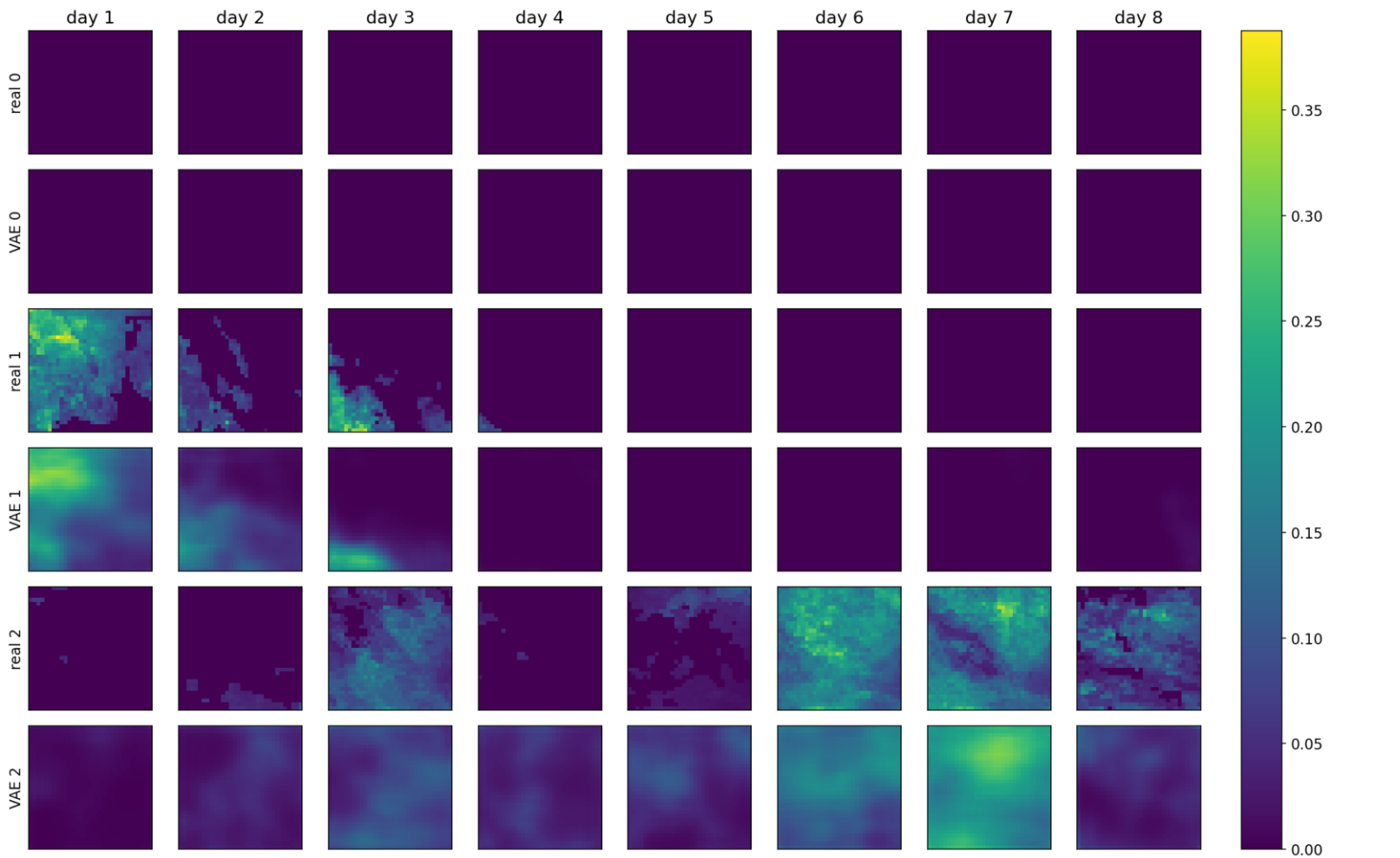


Architecture 1: “Vanilla” VAE



Case Study: Palghar Monsoons - Reconstructions

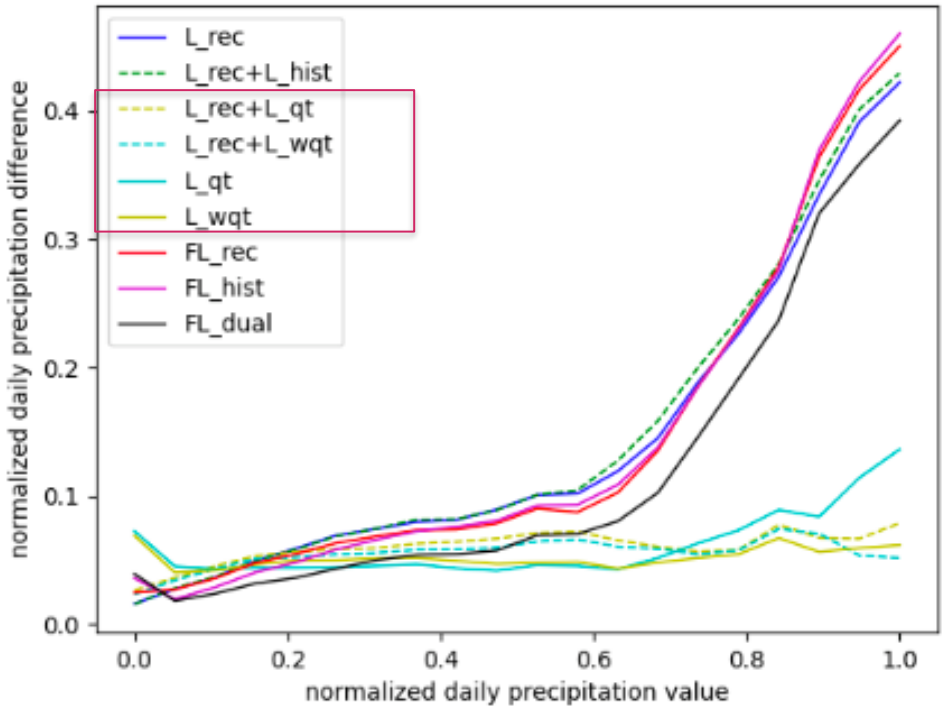




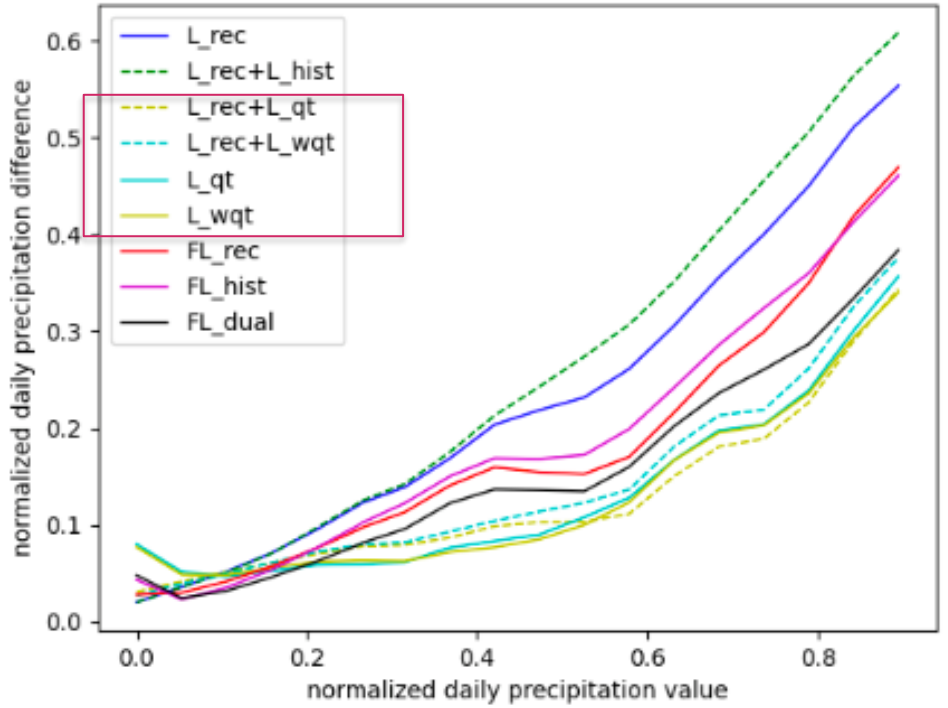
Architecture 2: Residual VAE

Quantized MSE Plots – Training/Test

Training

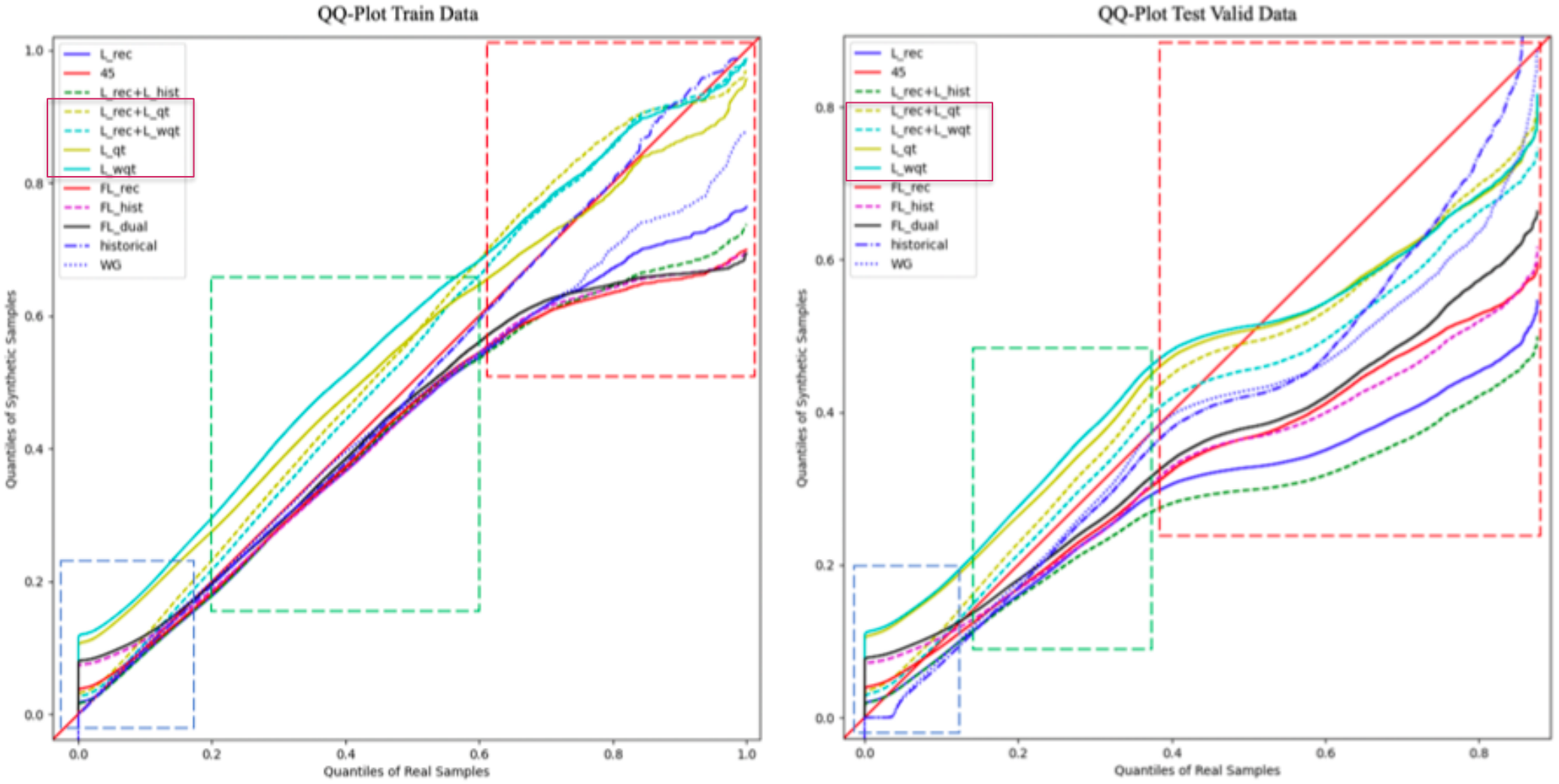


Test



One can notice improved results with VAEs using quantized losses.

Real vs Synthetic Distributions – Training/Test



One can notice improved results with VAEs using quantized losses.

Conclusions

- VAEs are powerful tools for encoding the distribution of daily precipitation data in a region and allowing sampling of extreme cases.

- The use of our proposed quantized reconstruction losses improves performance consistently over standard losses for the generation of realistic extreme weather data.

- We are currently experimenting with different architectures and parameters to further understand their effect on reconstruction and ability to generate event at the tail of the distribution.

Thanks!

biancaz@br.ibm.com