# Gaussian Assimilation of non-Gaussian Image Data via Pre-Processing by Variational Auto-Encoder (VAE)

**Daisuke Hotta**

***MRI/JMA***

# Ultimate goal: Assimilation of 2D images

- Examples:
  - Satellite images (brightness temperature)
  - Radar images (reflectivity)

- Challenges:
  - Non-Gaussianity
  - Dimensional redundancy
  - Interpixel correlations

Existing strategies
→ Gaussian Anamorphosis (GA)
→ Thinning, Linear dimension reduction by EOF (aka PCA)

# Background: Gaussian anamorphosis (GA)

Difficulty in assimilating non-Gaussian data:

Prior distribution p(x) and/or likelihood p(y|x) are not Gaussian (or cannot be well approximated by Gaussian distributions)

→ Conventional DA methods based on Gaussian assumption (Var or EnKF) do not function well

Idea of Gaussian anamorphosis:

- Transform x (model states) and/or y (observables) so that the prior p(x) and likelihood p(y|x) become closer to Gaussian
- then perform Gaussian-based DA in the transformed space, and transform the resultant analysis (or posterior distribution) back to the physical space

Limitation: variable transforms can only be constructed for univariate case in practice (construction for multidimensional case requires prohibitively large data set or simiply unfeasibly expensive)

# Ultimate goal: Assimilation of 2D images

- Examples:
  - Satellite images (brightness temperature)
  - Radar images (reflectivity)

- Challenges:
  - Non-Gaussianity
  - Dimensional redundancy
  - Interpixel correlations

Wish to simultaneously achieve the following:
- Variable transforms that brings distributions closer to Gaussian
- Dimension reduction
- Removal of error correlation

A solution has been proposed in ML community: Variational Auto-Encoder (VAE)

# Variational Auto-Encoder (VAE)

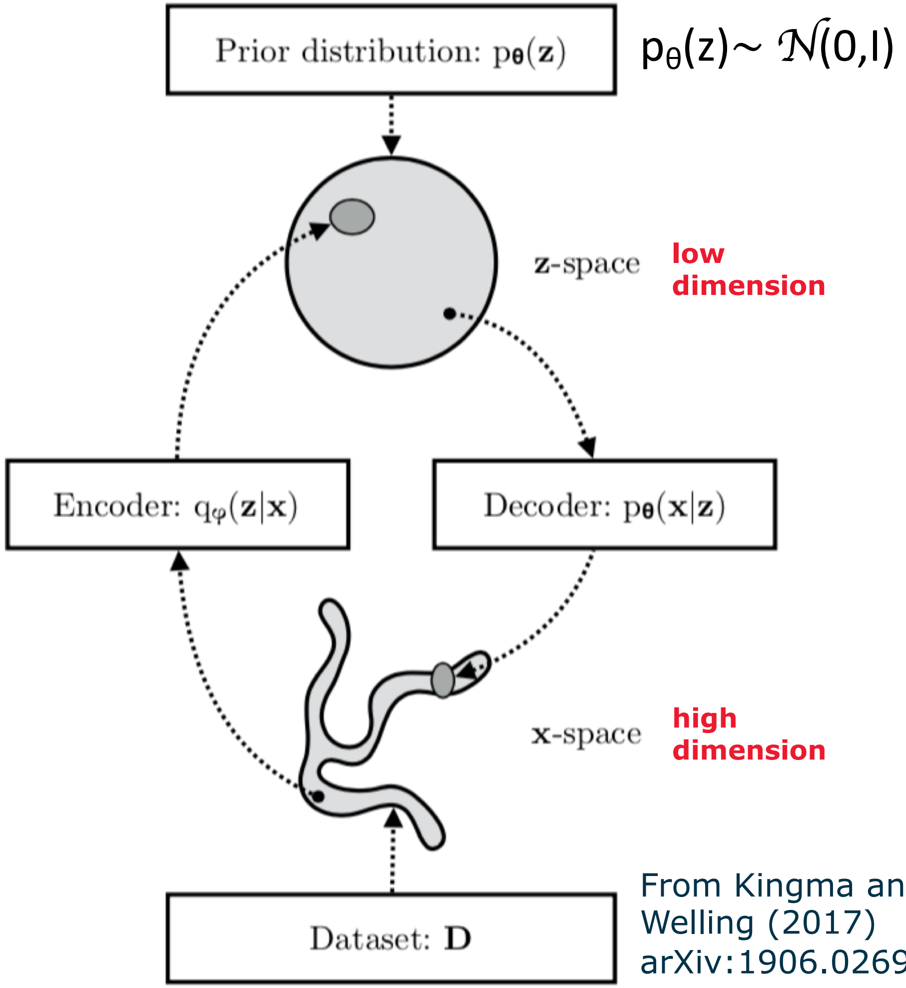
Generative Model which assumes:

- We have low-dimensional latent variable **z**, each element of which is uncorrelated and follows independent standard Gaussian distribution $\mathbf{z} \sim N(0,\mathbf{I})$
- High-dimensional complex data **x** are generated by nonlinearly transforming such **z**

Under this assumption, VAE learns a good Gaussian approximation of $p(\mathbf{x}|\mathbf{z})$ and $p(\mathbf{z}|\mathbf{x})$ solely from samples of **x**.

- No knowledge about **z** is required (unsupervised learning).
- Still, the size of **z** needs to be externally determined (or tuned).

# VAE: Schematic illustration




From Kingma and Welling (2017) arXiv:1906.02691

Assumption:

- High-dimensional complex data **x** are generated from low-dimensional Gaussian variables **z**∼N(0,**I**)
- Correspondence between such **x** and **z** are stochastically determined by p(**x**|**z**) and p(**z**|**x**).
- We can only observe **x**. Instances of **z** are not available.
- Under such situation, we wish to somehow obtain good approximations of p(**x**|**z**) and p(**z**|**x**).

- Now, by learning realizations *X* of **x**, we can obtain $p_\theta(\mathbf{x}|\mathbf{z})$ and $q_\varphi(\mathbf{z}|\mathbf{x})$ that well approximate p(**x**|**z**) and p(**z**|**x**).
- $p_\theta(\mathbf{x}|\mathbf{z})$ and $q_\varphi(\mathbf{z}|\mathbf{x})$ are both Gaussian distribution, whose mean and covariance are neural networks, whose weight parameters are to be estimated by machine learning.

- The encoder $q_\varphi(\mathbf{z}|\mathbf{x})$ stochastically transforms non-Gaussian **x** into Gaussian **z**. In this sense, this can be interpreted as a multidimensional extension to Gaussain Anamorphosis (← the only originality of this research)

# Non-Gaussian Data assimilation via VAE: fundamental idea

- Just like DA methods that use Gaussian Anamorphosis, transform non-Gaussian data (observation and/or background) by the mean of the VAE encoder $q_\varphi(\mathbf{z}|\mathbf{x})$
- then perform regular Gaussian-based data assimilation (Var or EnKF) in the transformed **z**-space by the mean of the VAE decoder $p_\theta(\mathbf{x}|\mathbf{z})$).

- Note: Encoder and decoder incur transformation errors on its own.
- But VAE neatly quantifies such errors as the variances of $p_\theta(\mathbf{x}|\mathbf{z})$ and $q_\varphi(\mathbf{z}|\mathbf{x})$
- → Can incorporate transformation errors by adding encoder variances to the diagonal of **R** and **B** in **z**-space
- This acts like an automated QC in the sense that the data that are incompatible with the training data (=outliers) are automatically assigned large error variances and hence are not used in assimilation.

# Non-Gaussian DA with pre- and post-processing by VAE

- For simplicity, consider a special case where all state variables are observed ($H$=id)

**(off-line preparation)**

- Suppose a large amount of observation $\mathbf{y}^o$ from climatological distribution are available.
- We train VAE feeding climatological $\mathbf{y}^o$ data as input to obtain the encoder $q_\varphi(\mathbf{z}|\mathbf{x})$ and decoder $p_\theta(\mathbf{x}|\mathbf{z})$.
- From the assumption that $H$=id, state variables $\mathbf{x}$ and observables $\mathbf{y}$ share common encoder and decoder.

**(procedure at each assimilation step)**

1) Prepare $\mathbf{z}$-space observation $\mathbf{z}^o$ as an encoder mean of $\mathbf{y}^o$
2) Similarly, transform prior $\mathbf{x}^b$ into $\mathbf{z}^b$ in $\mathbf{z}$-space
3) Perform Gaussian-based DA (Var or EnKF) in $\mathbf{z}$-space, with following tweaks (deviation from typical use of Gaussian-Anamorphosis):
   - add $\mathbf{x}\to\mathbf{z}$-encoder variance $\sigma^2(\mathbf{x}^b)$ to the diagonal of $\mathbf{B}$
   - add $\mathbf{y}\to\mathbf{z}$-encoder variance $\sigma^2(\mathbf{y}^o)$ to the diagonal of $\mathbf{R}$
4) Transform $\mathbf{z}$-space posterior back into physical space by VAE decoder

Gaussian world

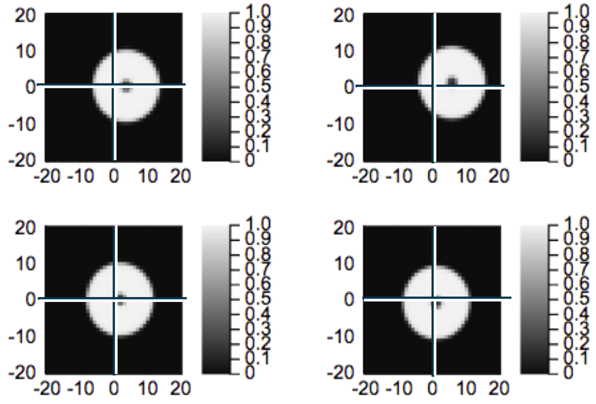
$$\begin{tikzcd} z_x \arrow[r, mapsto, "\mathrm{id}"] \arrow[d, "f(z_x)"'] & z_y \\ x \arrow[r, mapsto, "H=\mathrm{id}"'] & y \arrow[u, "\mu(y)"'] \end{tikzcd}$$

non-Gaussian world

# Idealized Experiment: Correction of positional error by assimilation of image data

# Idealized Experiment: Problem set-up

- Consider a disk with an “eye” (idealization of a tropical cyclone),
- whose center position is climatologically Gaussian distributed with $\mathcal{N}((0,0),\ 2^2 \times \mathbf{I}_2)$.
- Assume that background is twice more accurate than the climatology, $\mathcal{N}((0,0),\ \mathbf{I}_2)$,
- and the observation is even more accurate by a factor of two but with displaced mean, $\mathcal{N}(4,4),\ 1/2^2 \times \mathbf{I}_2)$.
- The true latent variables are the center position.
- Given a center position, an image with 41x41 pixels is generated whose pixel values are in [0, 1]

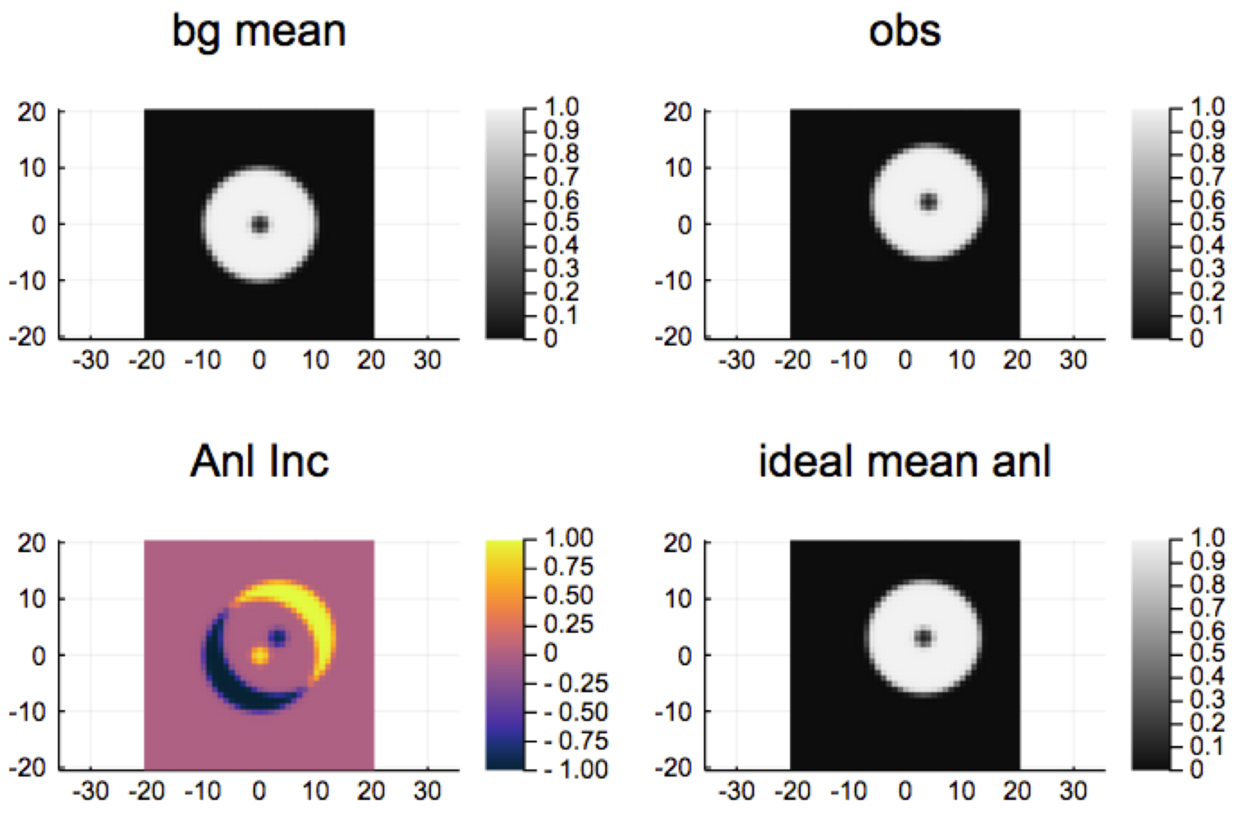



**Problem:**

- Assume that the true latent variables, their PDF, and the rule that generates the images from the the latent variables, are all unknown, but climatology of observed images are available.
- Under such a condition, can we correct “tropical cyclone” position errors by assimilating images?
- N.B.: This problem may seem trivial to human eyes, but is actually very difficult for classical DA methods given the strong non-Gaussianity of the pixel data (superposition of two delta functions centered at 0 and 1 in this case).

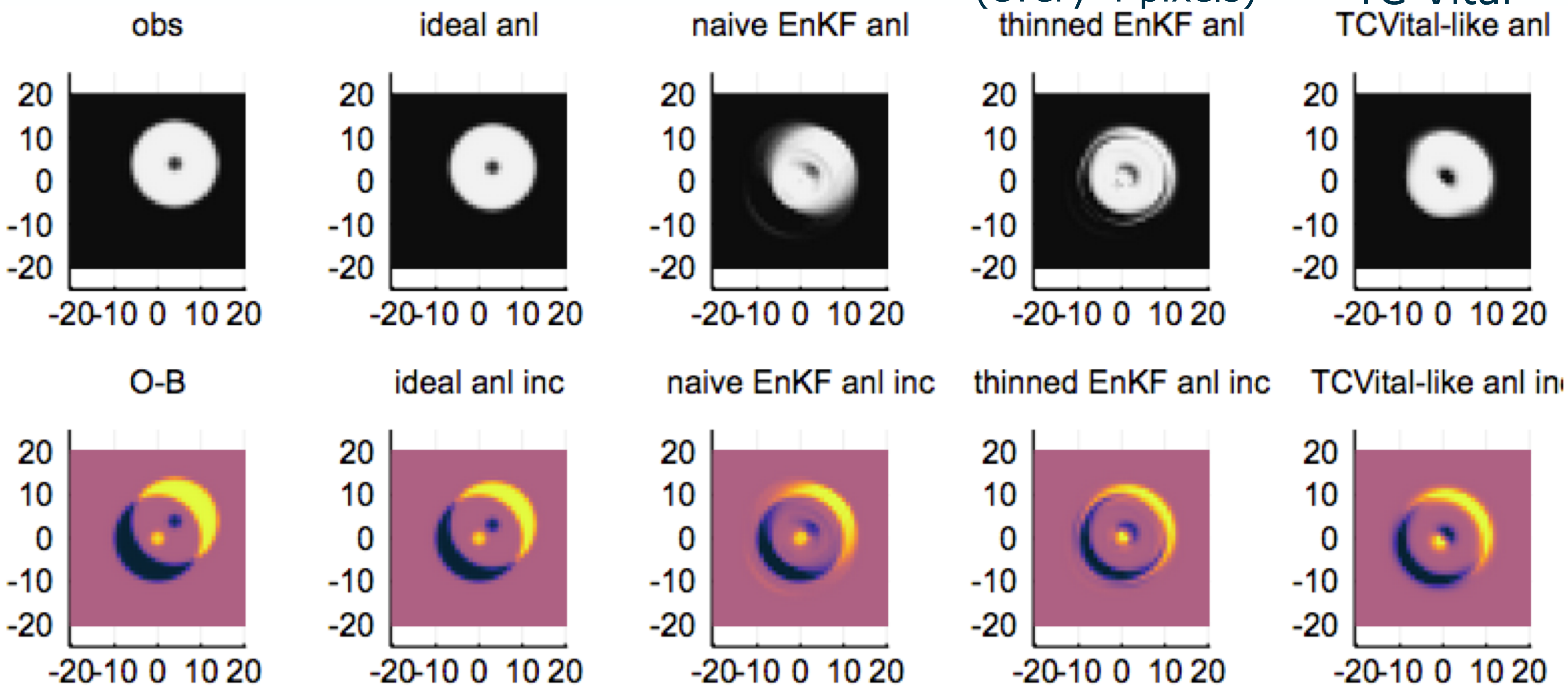
# Idealized Experiment: ideal analysis (used as reference hereafter)

- In this problem, the center position is Gaussian distributed, so the optimal solution is obtained by Kalman filter assimilation of the center position.
- Analysis thus obtained is hereafter used as the a reference solution. Performance of various DA methods are evaluated by how close their analyses are to this reference solution.



# Idealized Experiment: analyses from existing methods (10,000 member EnKF)

| | reference | naïve EnKF (Perturbed Obs, diagonal R) | EnKF with thinning (every 4 pixels) | EnKF assimilating TC-Vital |
|---|---|---|---|---|



# Idealized Experiment: VAE setup

- Draw 5,000 instances of the center position from the climatological distribution and generate 5,000 samples of $\mathbf{x}$ that follow climatological distribution.
- Train VAE on such 5,000 climatological dataset to obtain encoder $q_\varphi(\mathbf{z}|\mathbf{x})$ and decoder $p_\theta(\mathbf{x}|\mathbf{z})$

- Encoder $q_\varphi(\mathbf{z}|\mathbf{x})$ configuration:
  - Multilayer perceptron with only one hidden layer
  - Input layer: 41x41-dimensional $\mathbf{x}$ → (dense layer with ReLU activation) → hidden layer, 500 neurons → (dense layer with identity activation) → output layer: mean $\mu_\varphi$ and log-variance $\log\sigma^2_\varphi$ , 10 neurons each
- Decoder $p_\theta(x|z)$ configuration:
  - a mirror of encoder
  - Input layer: 10-dimensional $\mathbf{z}$ → (dense layer with ReLU activation) → hidden layer ,500 neurons → (dense layer, sigmoid activation) → output layer: the mean $f_\theta(\mathbf{z})$ of $p_\theta(\mathbf{x}|\mathbf{z})$
  - $p_\theta(\mathbf{x}|\mathbf{z}) = \mathcal{N}(f_\theta(\mathbf{z}), \sigma^2_x \mathbf{I})$, with the variance $\sigma^2_x = 0.01^2$ assumed independent of $\mathbf{z}$

# Idealized Experiment: Prerequisite for VAE-based assimilation to work well

- Trainning on VAE should be successful, which means that:
  1. The reconstruction error should be small enough: $\| f_\theta(\mu_\varphi(\mathbf{x})) - \mathbf{x} \| \ll \| \mathbf{x} \|$
  2. The distribution of encoded data should be close to Gaussian

- The above prerequisites should hold not for the training data but also for independent test data.

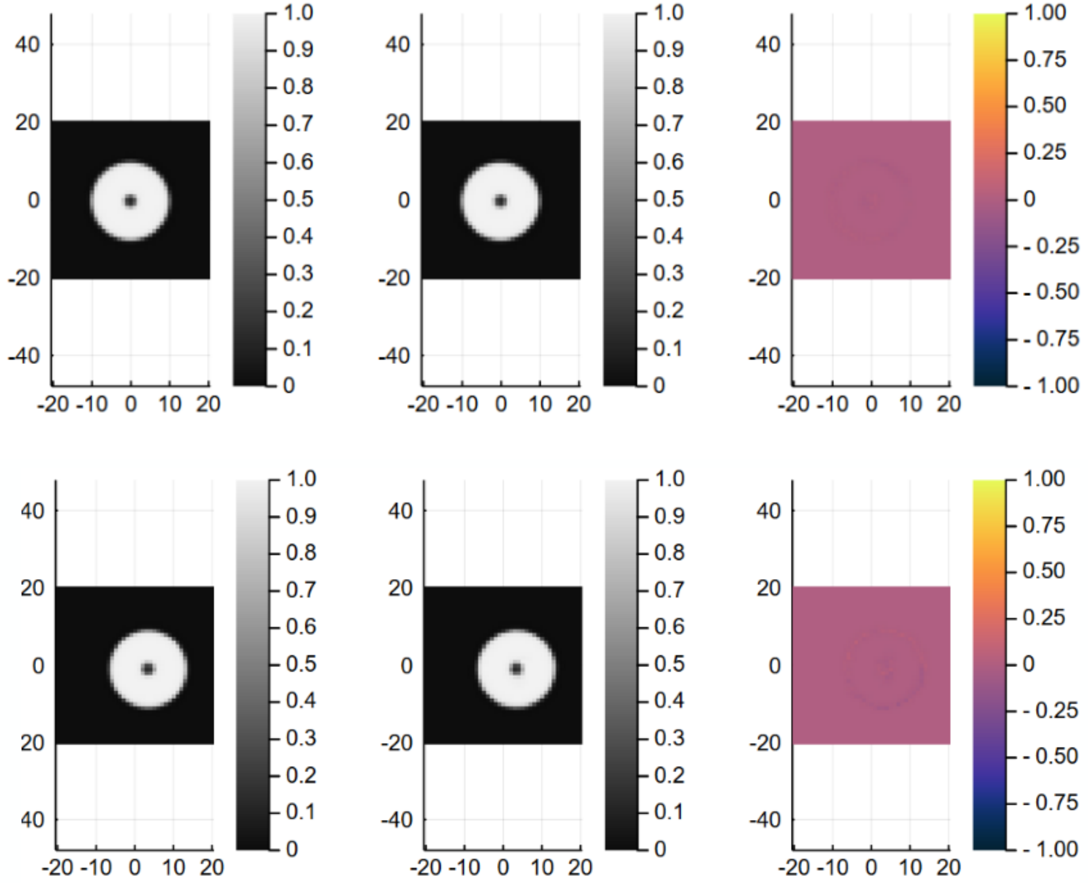
- We proposed to train VAE on climatological samples.
- The above prerequisites should hold not only for climatological samples, but also for the instantaneous flow-dependent distribution that reflect "errors of the day" (← non-trivial assumption)

# Idealized Experiment: Reconstruction error of the trained VAE

original x

reconstructed x $f_\theta(\mu_\phi(x))$

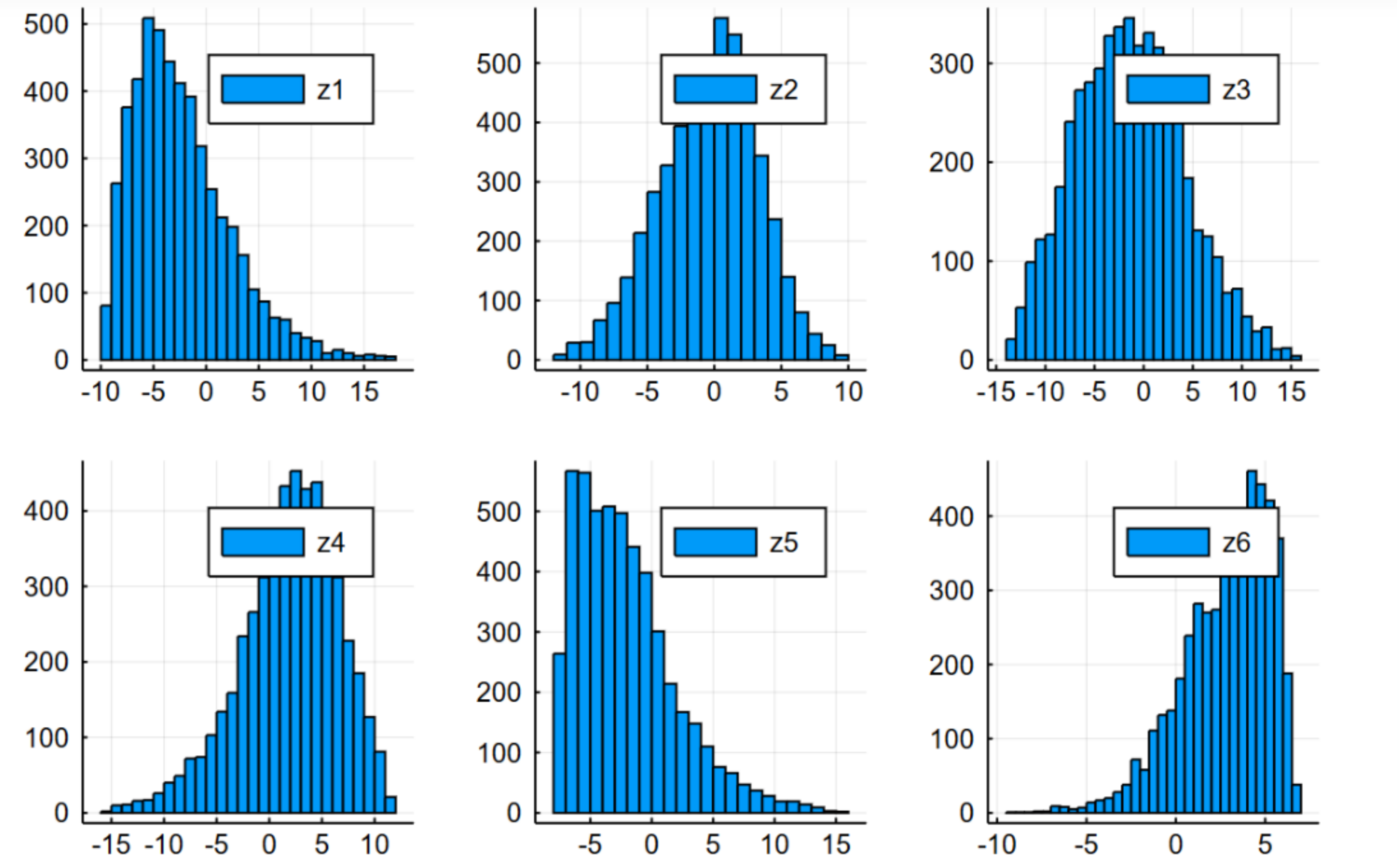
reconstruction error




random draw of x from $p^{clim}(x)$

random draw of x from prior $p(x)$

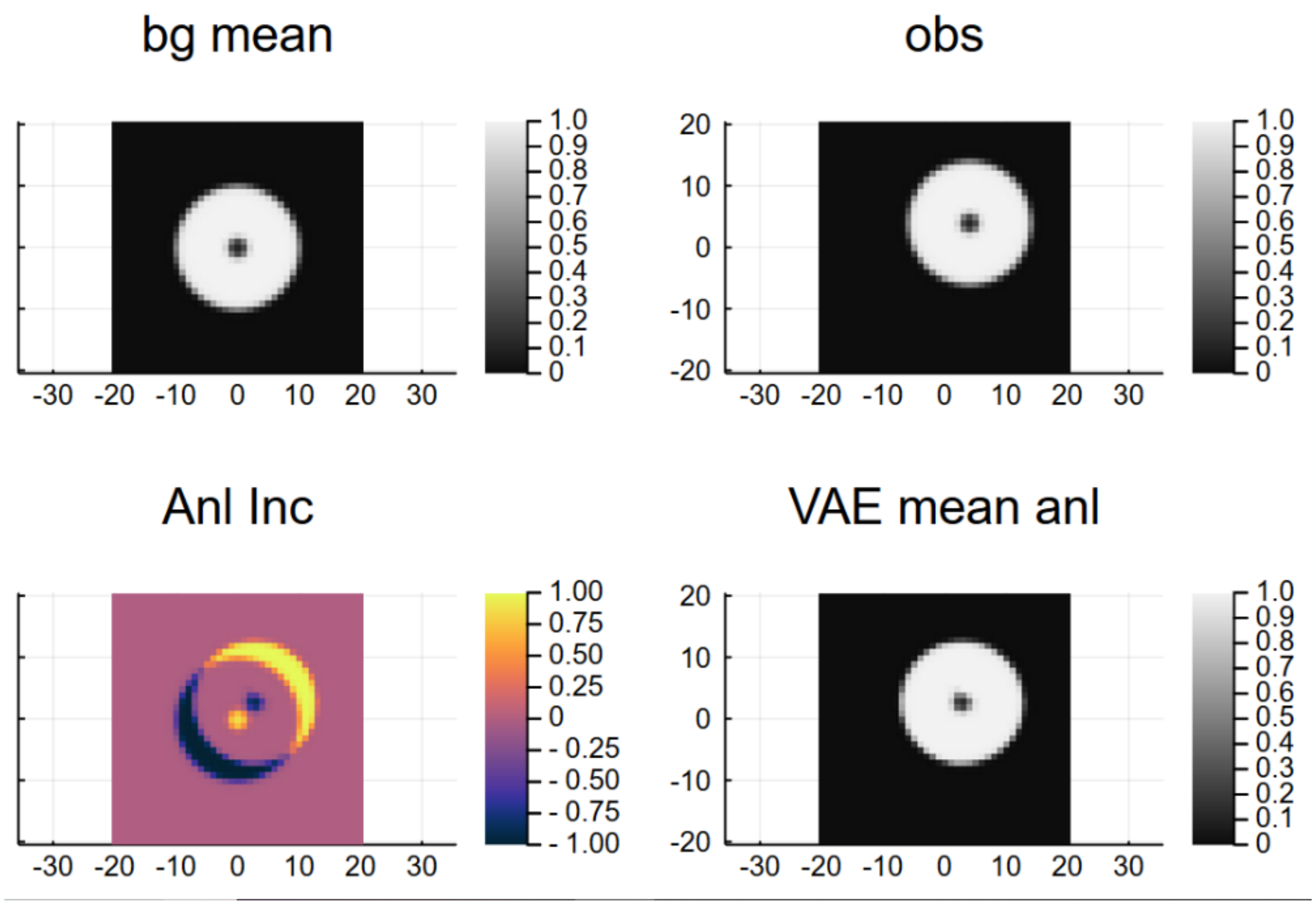
# Idealized Experiment: Gaussianity of encoded data




Marginal histograms of each component of the encoder mean $\{\mu_\varphi(\mathbf{x}) | \mathbf{x} \sim p^{clim}(\mathbf{X})\}$

Not perfectly Gaussian, but good enough considering that pixel-wise marginal distribution of the original data (before encoding) was a superposition of two delta functions $a\,\delta(x-0)+(1-a)\delta(x-1)$

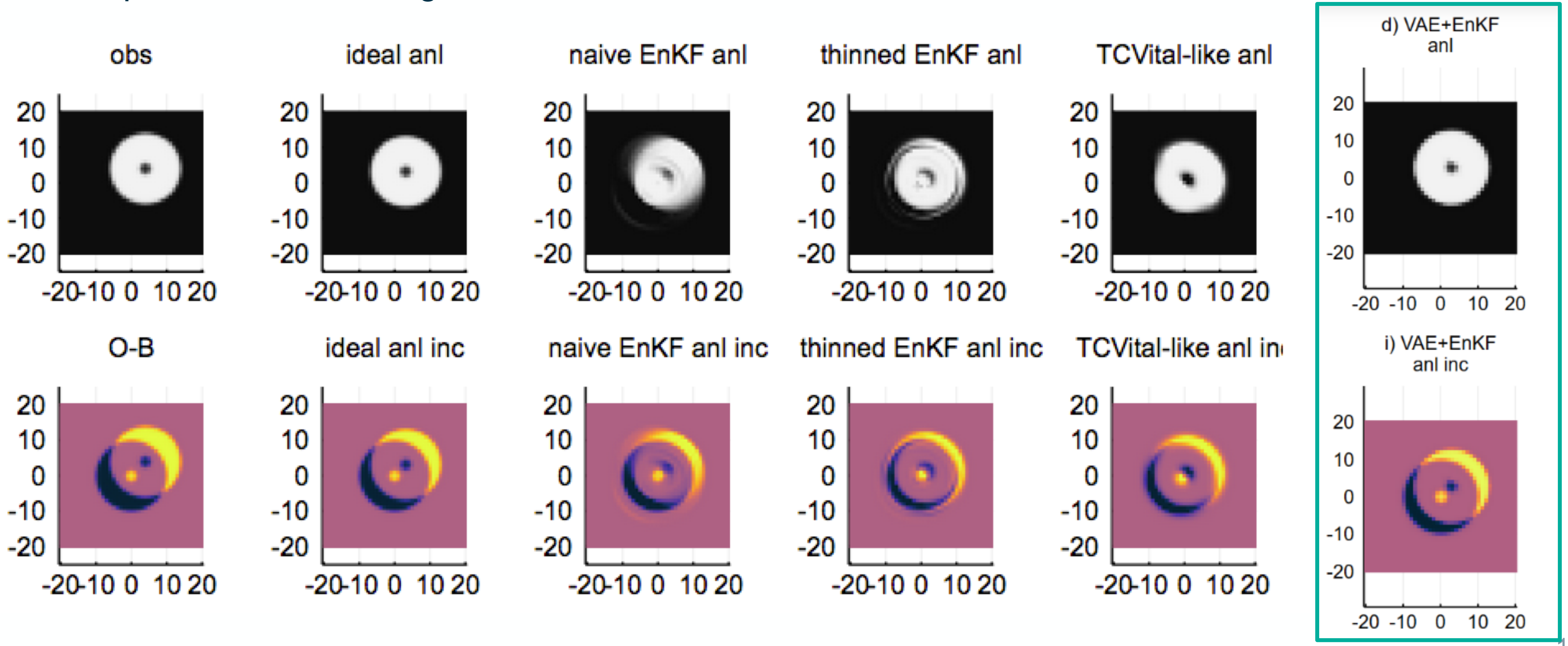
# Idealized Experiment: Result of VAE-aided assimilation




- Very close to the optimal reference analysis
- without any visible distortion to the “disk-with-an-eye” structure

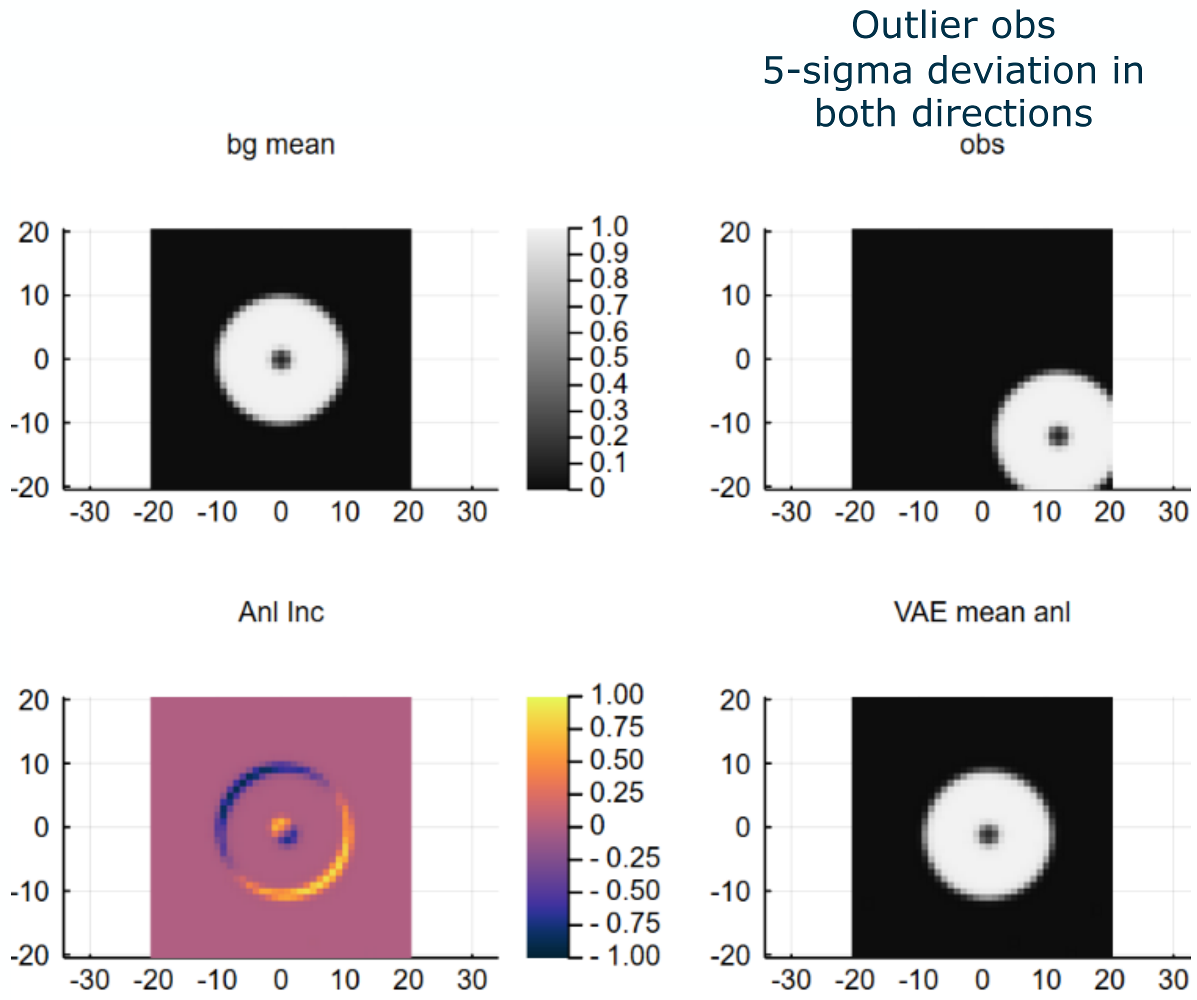
# Idealized Experiment: Result of VAE-aided assimilation

- Comparison with existing methods



# Idealized Experiment: Built-in QC of VAE-aided assimilation

- Machine learning is often blamed for being incapable to handle events not contained in training dataset.
- However, with VAE, such “outlier events” can be detected by increase in encoder variance
- → Adding encoder variance to **R** allows to automatically ignore outlier observations.



# Summary and ideas toward more realistic situation

- VAE can be used as a multidimensional extension to Gaussian Anamorphosis to effectively assimilate non-Gaussian image data. Promising results have been obtained for a simplistic case where all state variables are observed.

- Challenge for application to more realistic situations:
  - highly nonlinear observation operator. When obs op $H \neq \mathrm{id}$, VAE has to be trained separately for state vector **x** and observable **y**, resulting in different **z**-space for each, so that obs op in **z**-space is a composition of VAE decoder for **x** (which is a nonlinear neural net), obs op in physical space *H*, and VAE encoder for **y** (again a nonlinear neural net).
  - → VAE may resolve non-Gaussianity, but it comes at the cost of introducing nonlinearity. How can the latter be handled?
  - Promising approach is perhaps to combine VAE with iterative ensemble Kalman smoother (IEnKS; Evensen 2018; Bocquet and Sakov 2014)
    - VAE expected to handle non-Gaussianity. IEnKS will likely overcome non-linearity in observation operator.