Reproducing

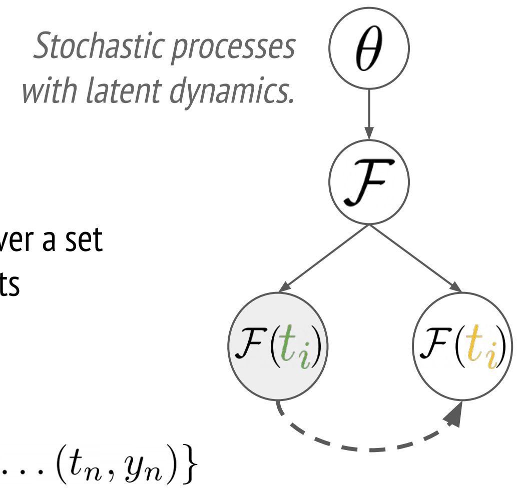
Neural ODE Processes

Alexander Norcliffe, Cristian Bodnar, Ben Day, Jacob Moss, Pietro Liò

(to appear in ICLR 2021)

Task

- Time-series governed by **latent dynamics**
- Estimate the **conditional distribution** $p(y \mid C)$ over a set T of target points, given a set C of context points
- Learn **latent dynamics** from the context C



$$T = \{\underbrace{(t_1, y_1) \dots (t_m, y_m)}_{C}, \underbrace{(t_{m+1}, y_{m+1}) \dots (t_n, y_n)}_{\text{extra target points}}\}$$

Model

NDP combines

Neural Processes framework :

$$p(y_{1:n}, z \mid t_{1:n}, C) = \overbrace{p(z \mid C)}^{\text{encoder}} \prod_{i=1}^{n} \overbrace{p(y_i \mid l_i, t_i)}^{\text{decoder}}$$

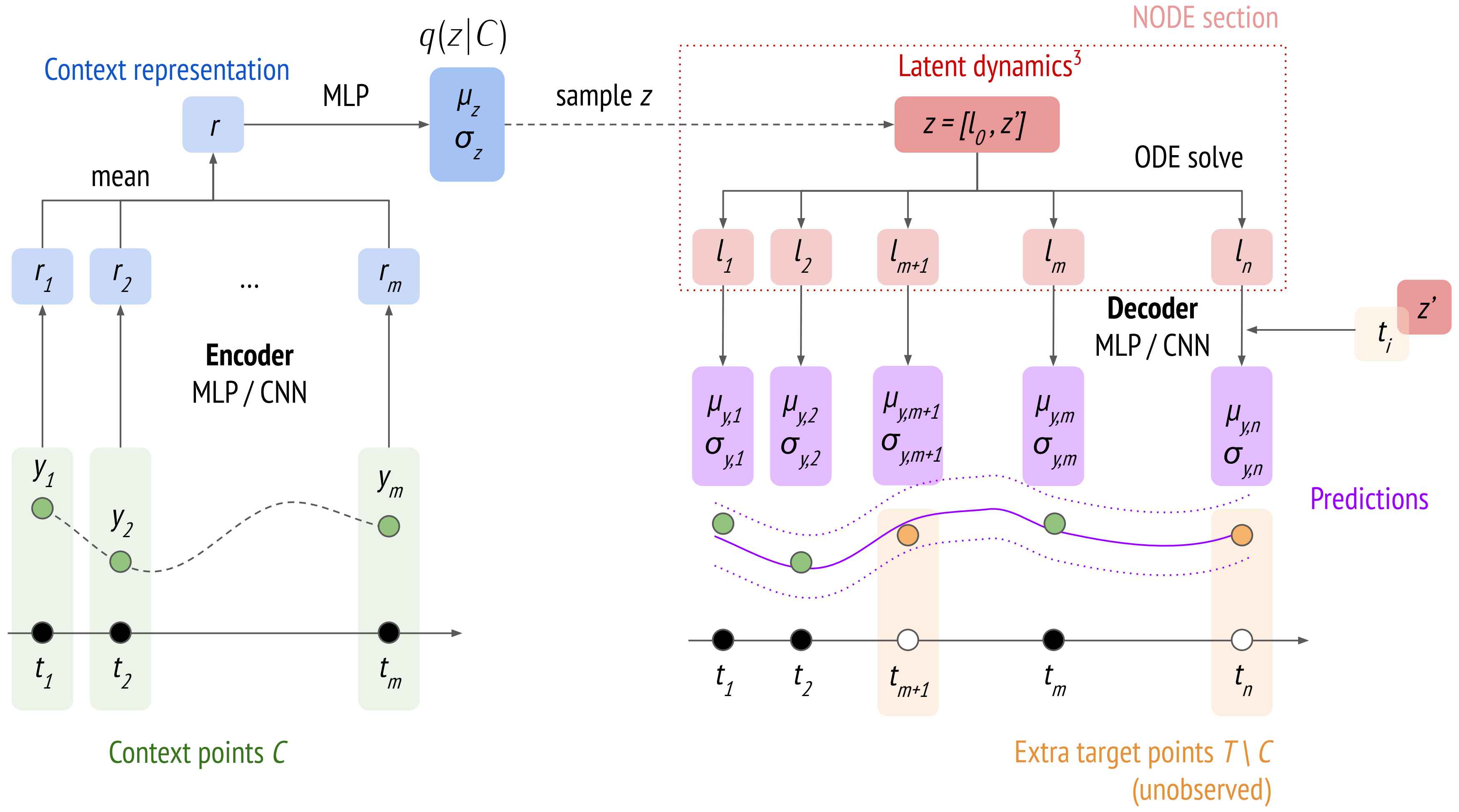
NODE for decoding :

$$l(t_i) = l_0 + \int_{t_0}^{t_i} f_\phi(l(t), t, z')dt$$

Training : maximize *log* $p(T \mid C)$ using ELBO loss:

$$\log p(y_{1:n} \mid t_{1:n}, C) \geq \mathbb{E}_{q(z \mid T)}\left[\sum_{i=1}^{n} \log p(y_i \mid l_i, t_i) + \log \frac{q(z \mid C)}{q(z \mid T)}\right]$$

Architecture



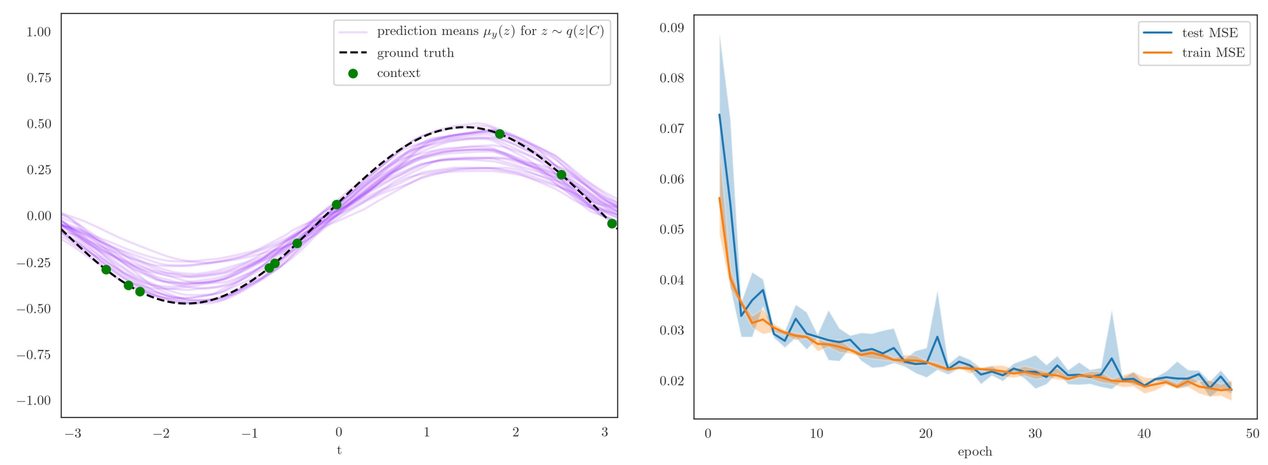
Visualizing model predictions

Sine Dataset

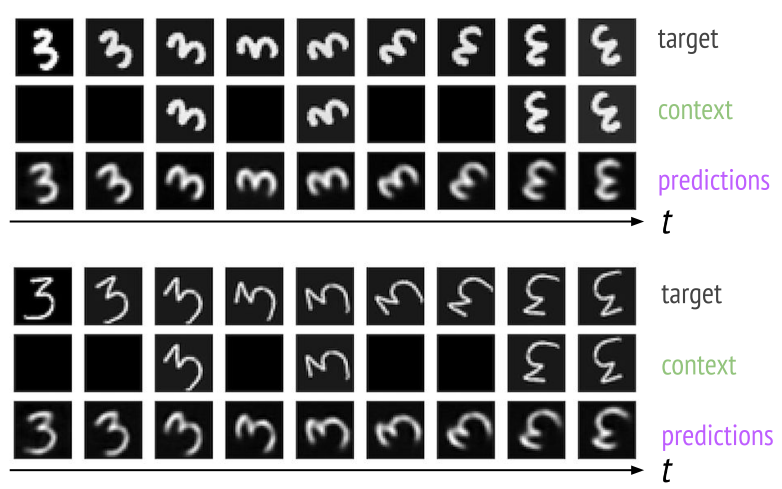
$t \longrightarrow$ **A** sin(t + **b**)

With latent dynamics

-1<**A**<1 and -0.5<**b**<0.5

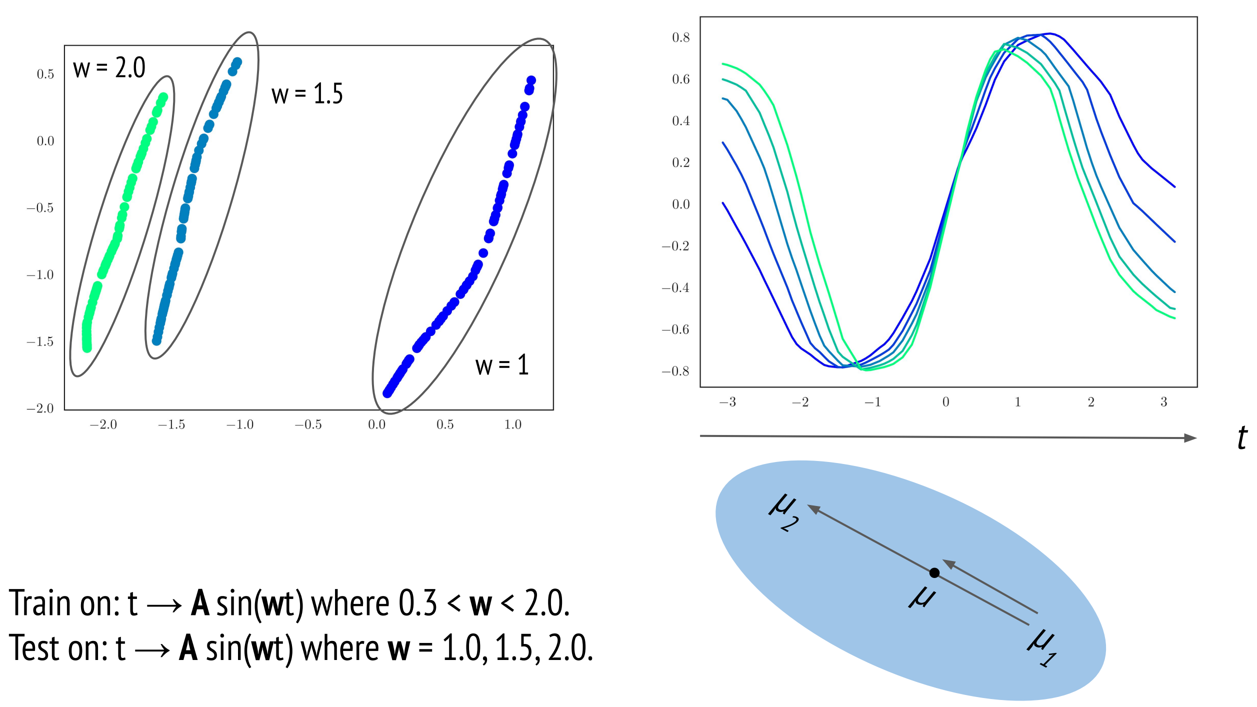


Rotnist Dataset[1]



Learning latent dynamics

Do context points suffice to represent different dynamics?



Train on: $t \longrightarrow$ **A** sin(**w**t) where 0.3 < **w** < 2.0.

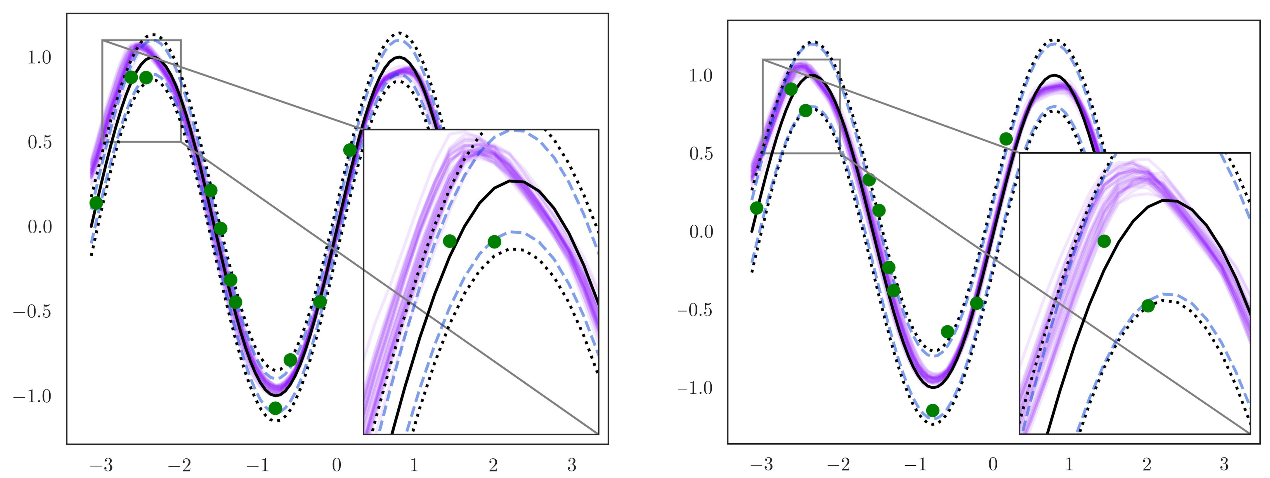
Test on: $t \longrightarrow$ **A** sin(**w**t) where **w** = 1.0, 1.5, 2.0.

Left: PCA (k=2) illustrates μ_z of **test time series context points**, with clear separation. Different colors have different latent dynamics (oscillator frequency).

Right: In latent space, we can interpolate between w = 1 and w = 2. Note the continuous transformation of the mean decoding.

Adaptation to noise level

Does the learned $q(z \mid C)$ represent noise in the training data?



--- : true uncertainty level σ

.... : estimated uncertainty level σ_y

● : noisy context points

Train/test on: $t \longrightarrow \sin(wt) + \sigma\,\mathbf{dB}(t)$, where w = 2. Deterministic dynamics, noisy samples.

Left: σ = 0.1. **Right:** σ = 0.2.

Note that the estimated and true uncertainty levels are close.

Summary

- NDP = NODE + NP
- Handles **irregular sampling** (alternative to binning).
- Captures **distribution over dynamics** & uncertainty.
- Scales to **high-dimensional data.**

Key references

[1] Casale *et al. Gaussian Process Prior Variational Autoencoders*. NeurIPS (2018)

[2] Garnelo *et al. Neural Processes*. ICML workshop (2018)

[3] Chen *et al. Neural Ordinary Differential Equations. NeurIPS* (2018)

[4] Le *et al. Empirical Evaluation of NP Objectives. NeurIPS workshop* (2018)

SORBONNE UNIVERSITÉ

Poster by:

Théophile Cantelobre

Matthias Crémieux

Alexandre Thomas

https://github.com/theophilec/Reproducing-Neural-ODE-Processes