# Meta-Learning Stationary Stochastic Process Prediction with Convolutional Neural Processes


Andrew Y. K. Foong[*1], Wessel P. Bruinsma[*1 2], Jonathan Gordon[*1], Yann Dubois[1], James Requeima[1 2], and Richard E. Turner[1 3]

[1]University of Cambridge, [2]Invenia Labs, [3]Microsoft Research Cambridge

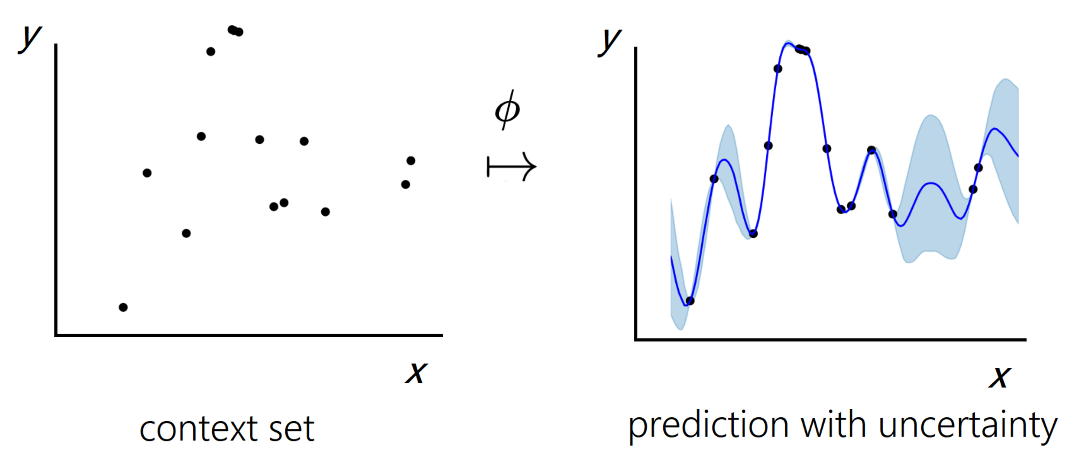

UNIVERSITY OF CAMBRIDGE

## What are Neural Processes (NPs)?

NPs **meta-learn** a map from datasets to **stochastic processes**.

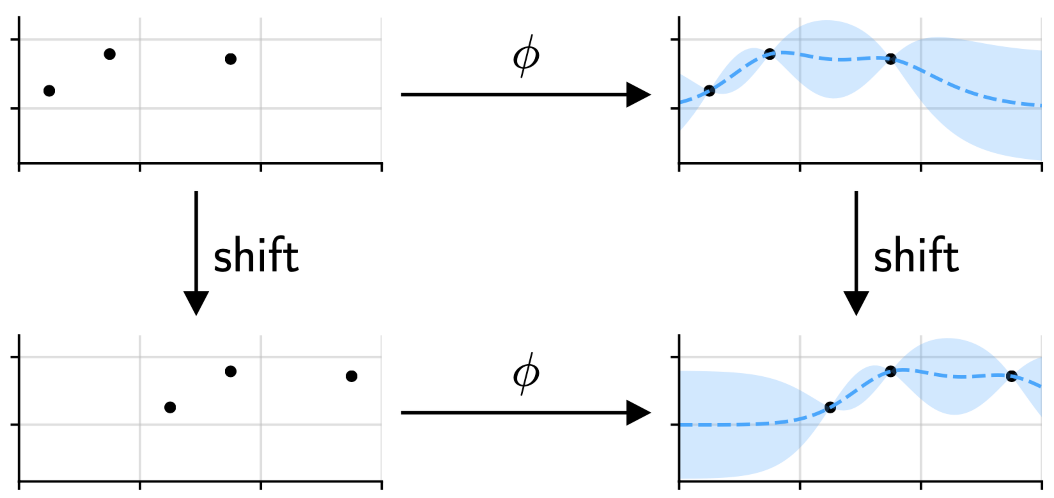
- Given an observed **context set** $D_\mathcal{C} = \{(x_c, y_c)\}_{c=1}^{C}$.
- Make predictions at a **target set** $\mathbf{x}_\mathcal{T} = \{x_t\}_{t=1}^{T}$.
- NPs use NNs to **directly parameterise** the prediction map:

$$\phi\colon \text{datasets } \mathcal{D} \to \text{predictions } \mathcal{P}, \quad \phi(D_\mathcal{C}) = p(\mathbf{y}_\mathcal{T}|\mathbf{x}_\mathcal{T}, D_\mathcal{C}).$$




## Translation Equivariance

If the prior stochastic process we are learning is **stationary**, then the prediction map $\phi$ is **translation equivariant**:




We propose the **Convolutional Neural Process** (ConvNP):

- Uses **convolutional neural networks** (CNNs) for $\phi$.
- More parameter-efficient than MLP-based neural processes.
- Is translation equivariant.
- Can perform **spatial generalisation**.

## Proposed Training Objective

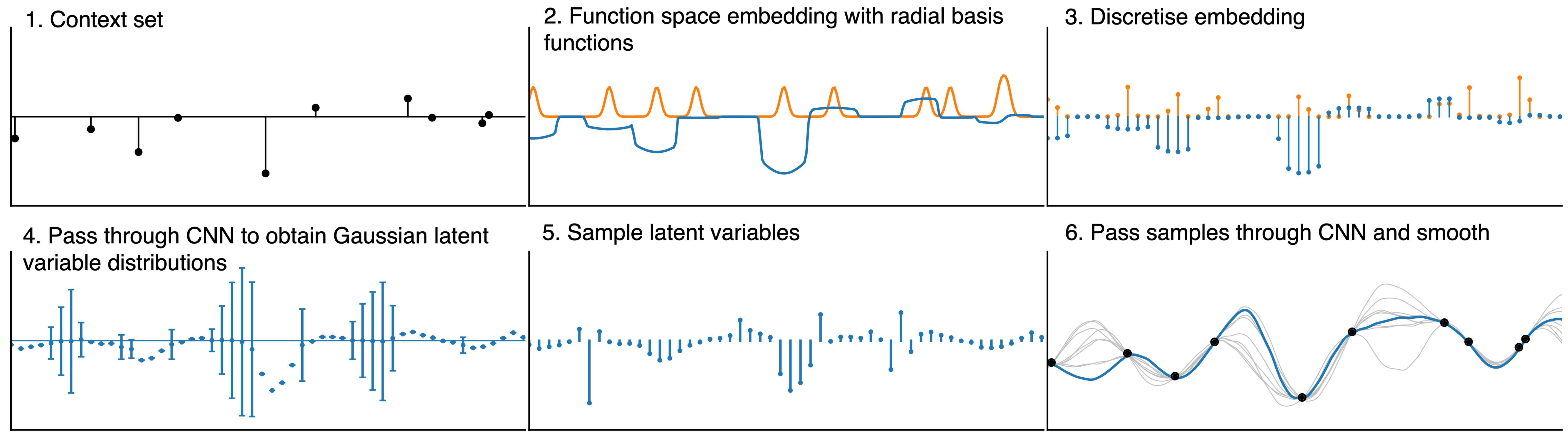
- The ConvNP likelihood is intractable due to its latent variable.
- We use a **Monte Carlo approximation to maximum likelihood**, instead of the usual variational inference objective:

$$\log p(\mathbf{y}_\mathcal{T}|\mathbf{x}_\mathcal{T}, D_\mathcal{C}) = \log \int p(\mathbf{y}_\mathcal{T}|\mathbf{x}_\mathcal{T}, z) p(z|D_\mathcal{C})\,\mathrm{d}z$$

$$\approx \log \frac{1}{L}\sum_{\ell=1}^{L} p(\mathbf{y}_\mathcal{T}|\mathbf{x}_\mathcal{T}, z_\ell), \quad z_\ell \sim p(z|D_\mathcal{C}).$$

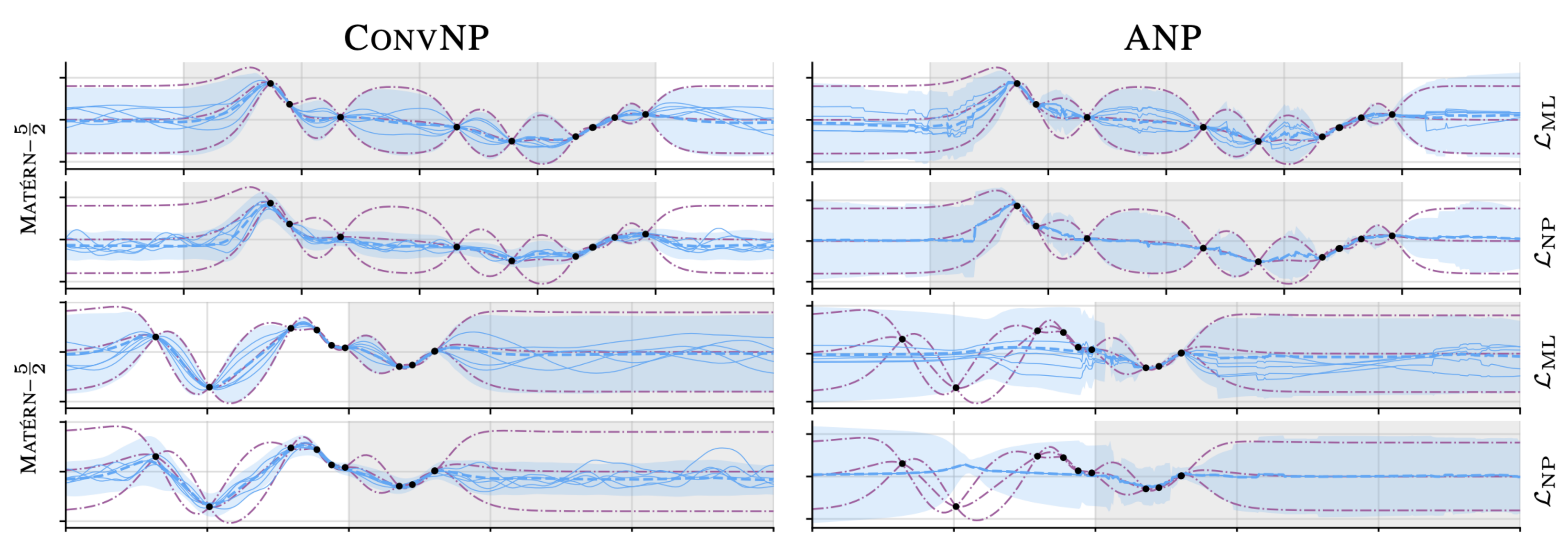
## The Convolutional Neural Process (ConvNP)

- The ConvNP builds on the previously published Convolutional *Conditional* Neural Process (ConvCNP).[a]
- The ConvCNP is translation equivariant but has factorised predictives: $p(\mathbf{y}_\mathcal{T}|\mathbf{x}_\mathcal{T}, D_\mathcal{C}) = \prod_{t=1}^{T} p(y_t|x_t, D_\mathcal{C})$.
- Our proposed ConvNP allows for **dependencies in the predictive distribution** and hence **coherent sampling**. Forward pass:

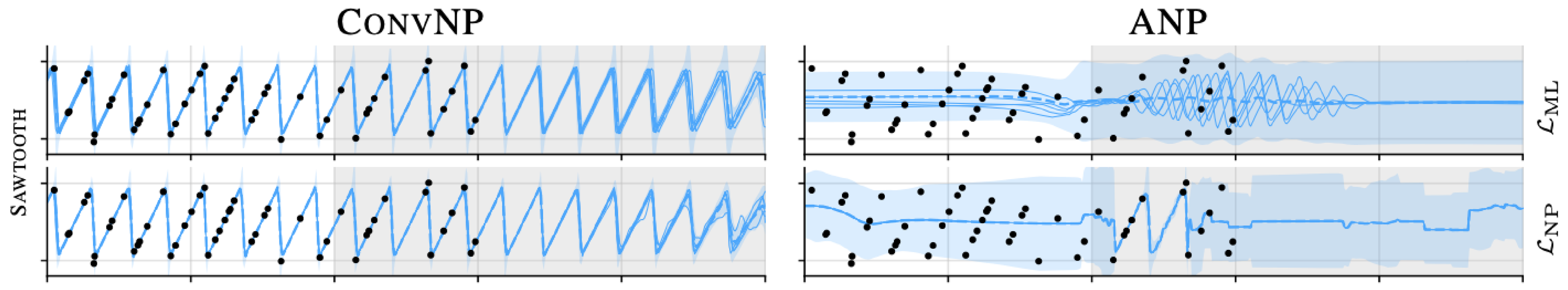



[a]Gordon, J., Bruinsma, W.P., Foong, A.Y.K., Requeima, J., Dubois, Y. and Turner, R.E., Convolutional Conditional Neural Processes. ICLR 2020.

## 1D Regression Experiments

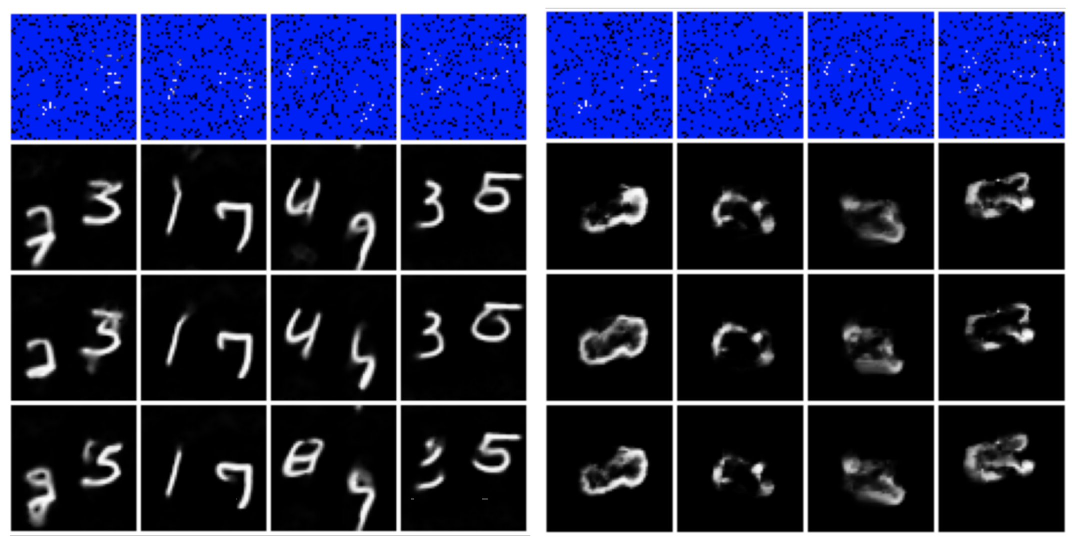



- ConvNP and Attentive NP (ANP) trained in gray region, on both our new $\mathcal{L}_{\text{ML}}$ objective and the old $\mathcal{L}_{\text{NP}}$ objective.
- **ANP fails to generalise spatially**.
- ConvNP with our new $\mathcal{L}_{\text{ML}}$ objective performs best.



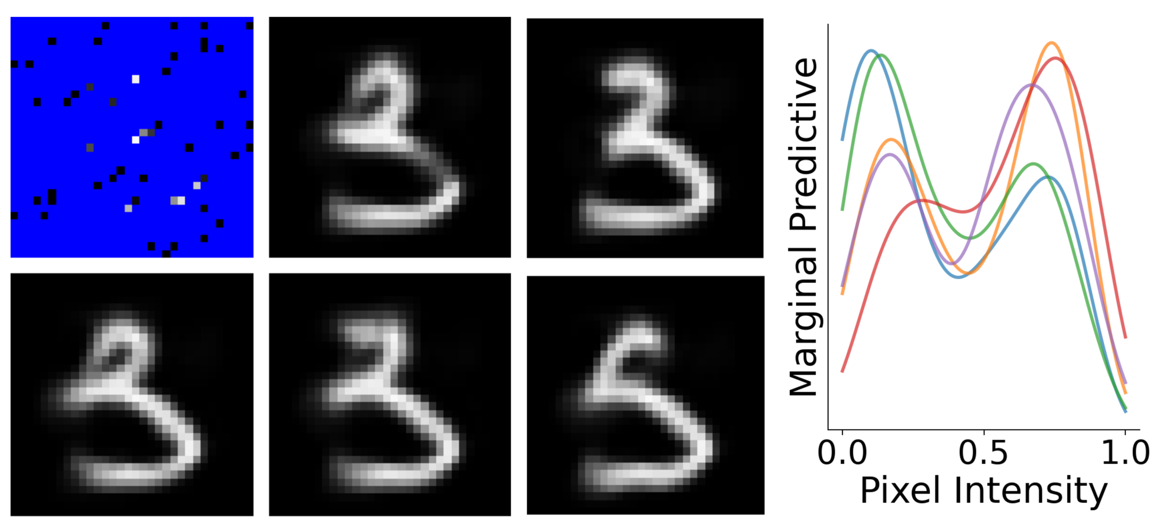

- ConvNP also successfully learns to extrapolate a challenging, non-Gaussian sawtooth process.

## Image (2D Regression) Experiments



(a) ConvNP (b) ANP

- Train on single, centered MNIST digits.
- Test on **multiple, non-centered** MNIST digits. Requires **spatial generalisation** — only ConvNP succeeds.
- ConvNP has **multimodal** predictive, unlike ConvCNP:




**See paper for CelebA, SVHN, and environmental data!**