The Forty Second International Conference on Machine Learning (ICML 2025)

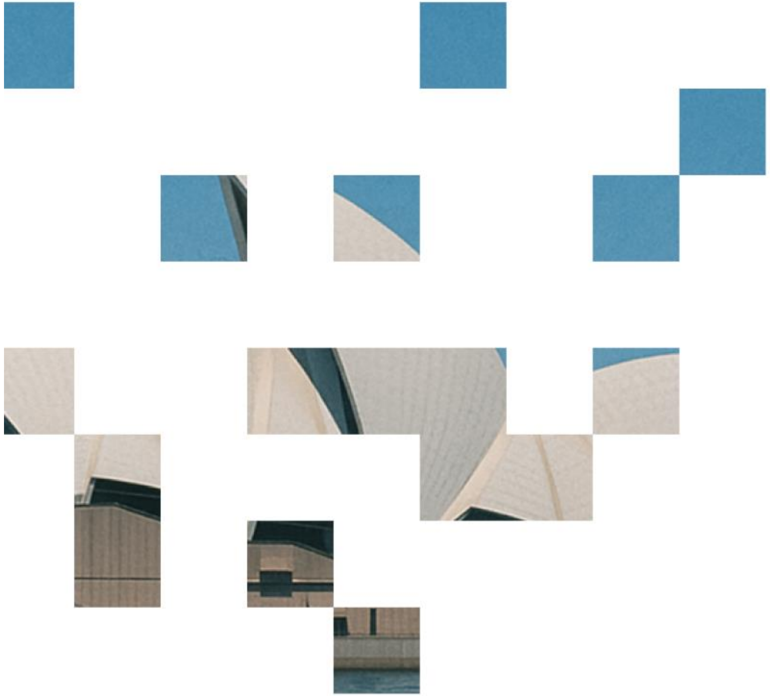
Renyi Neural Processes

Xuesong Wang

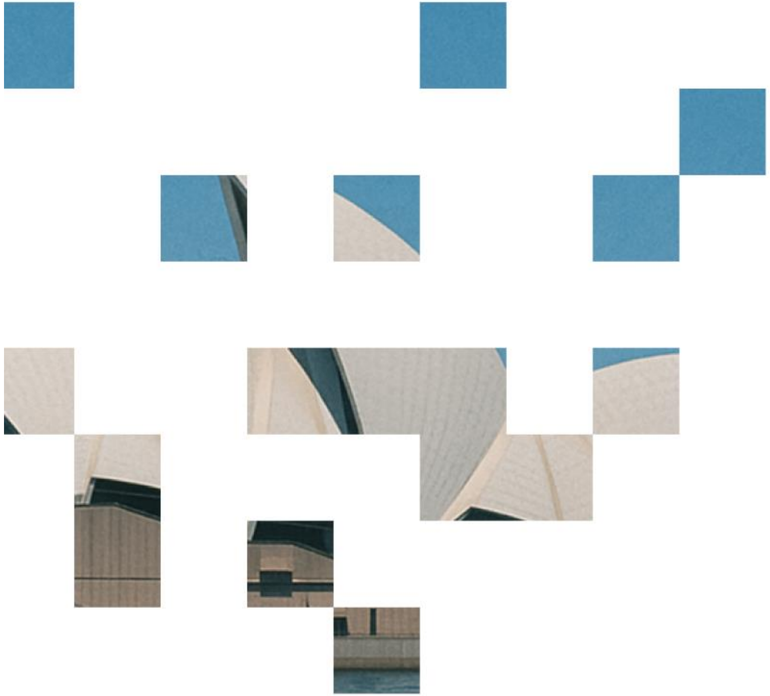
He Zhao

Edwin V. Bonilla

Where is this ?



Where is this ?

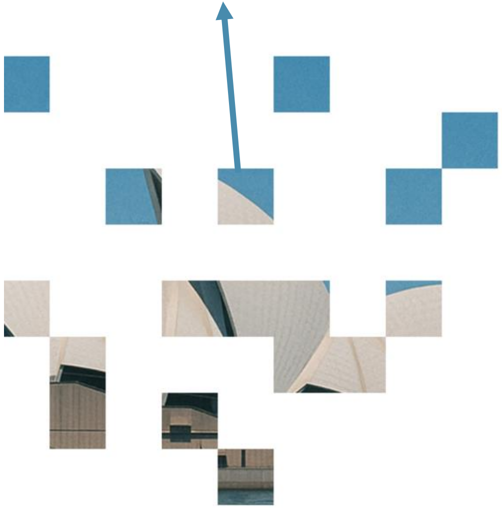


Sydney Opera House !

Neural Processes: a predictive model based on contexts

Context set

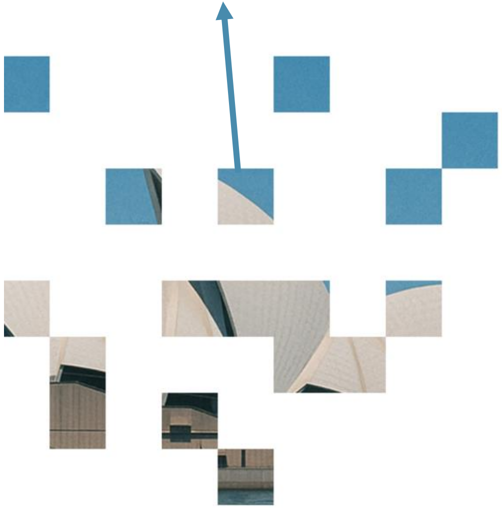
$(x_C, y_C) \in C$



Neural Processes: a predictive model based on contexts

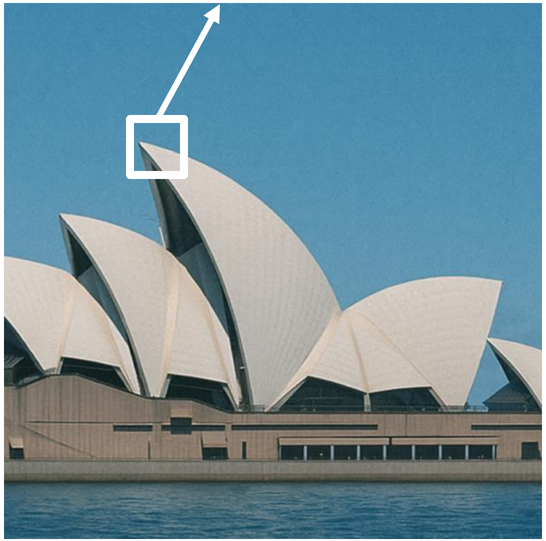
Context set

$(x_C, y_C) \in C$



Target set

$(x_T, y_T) \in T$



$$p(y_T | x_T, C) = ?$$

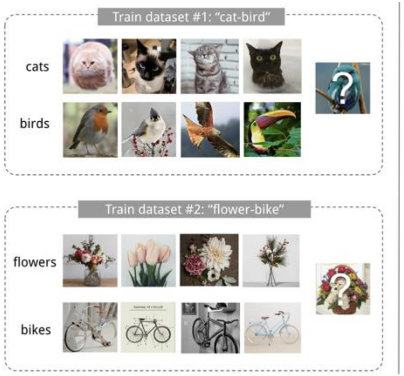
Why is this important ?

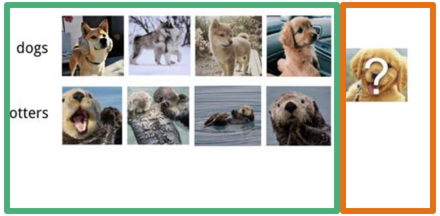
fast adaptation by conditioning on a new context set

Why is this important ?

fast adaptation by conditioning on a new context set

- Meta learning & multitask learning





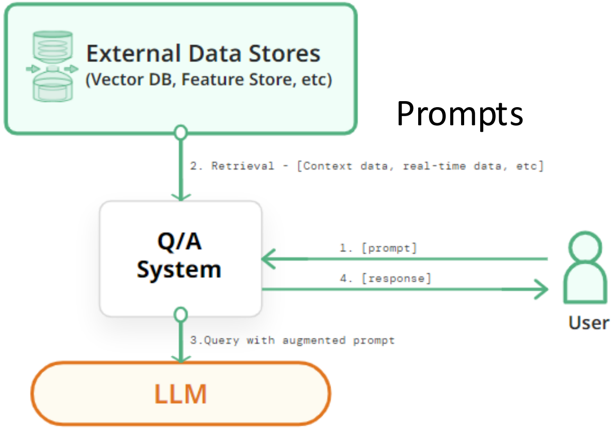
new task

Why is this important ?

fast adaptation by conditioning on a new context set

- Meta learning & multitask learning
- Retrieval-augmented generation (RAG) for LLMs

Prompts

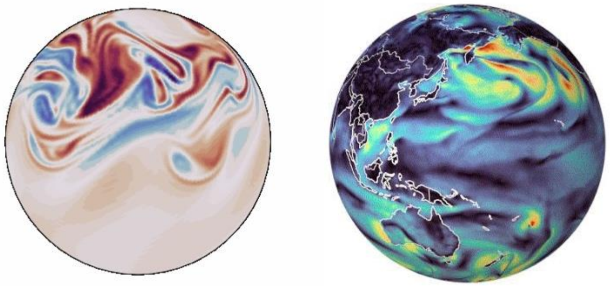


 https://www.hopsworks.ai/dictionary/retrieval-augmented-generation-llm

Why is this important ?

fast adaptation by conditioning on a new context set

- Meta learning & multitask learning
- Retrieval-augmented generation (RAG) for LLMs
- Sim2real in scientific computing
- …



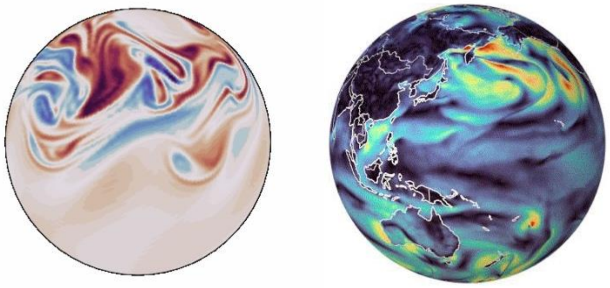
simulation physical

 “Spherical Fourier Neural Operators”, Bonev et al, ICML 2023

Why is this important ?

fast adaptation by conditioning on a new context set

- Meta learning & multitask learning
- Retrieval-augmented generation (RAG) for LLMs
- Sim2real in scientific computing
- …

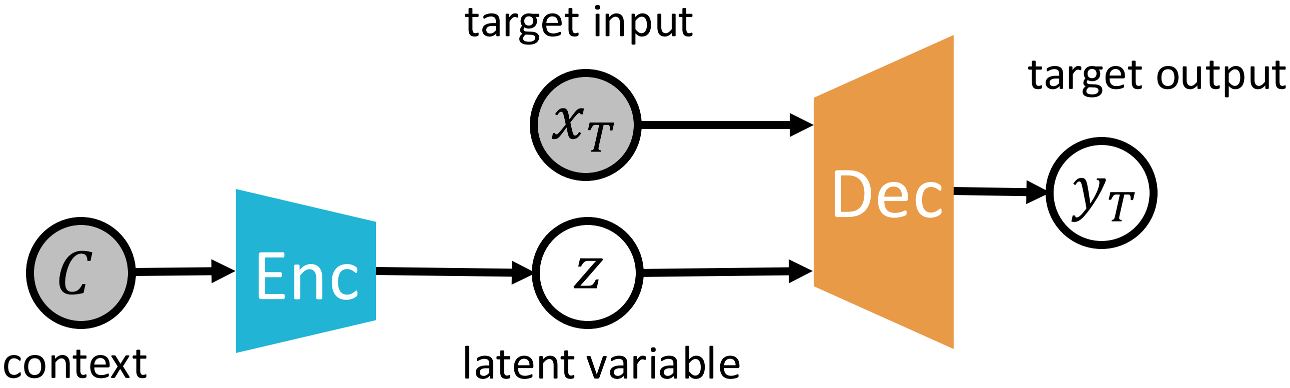


simulation physical

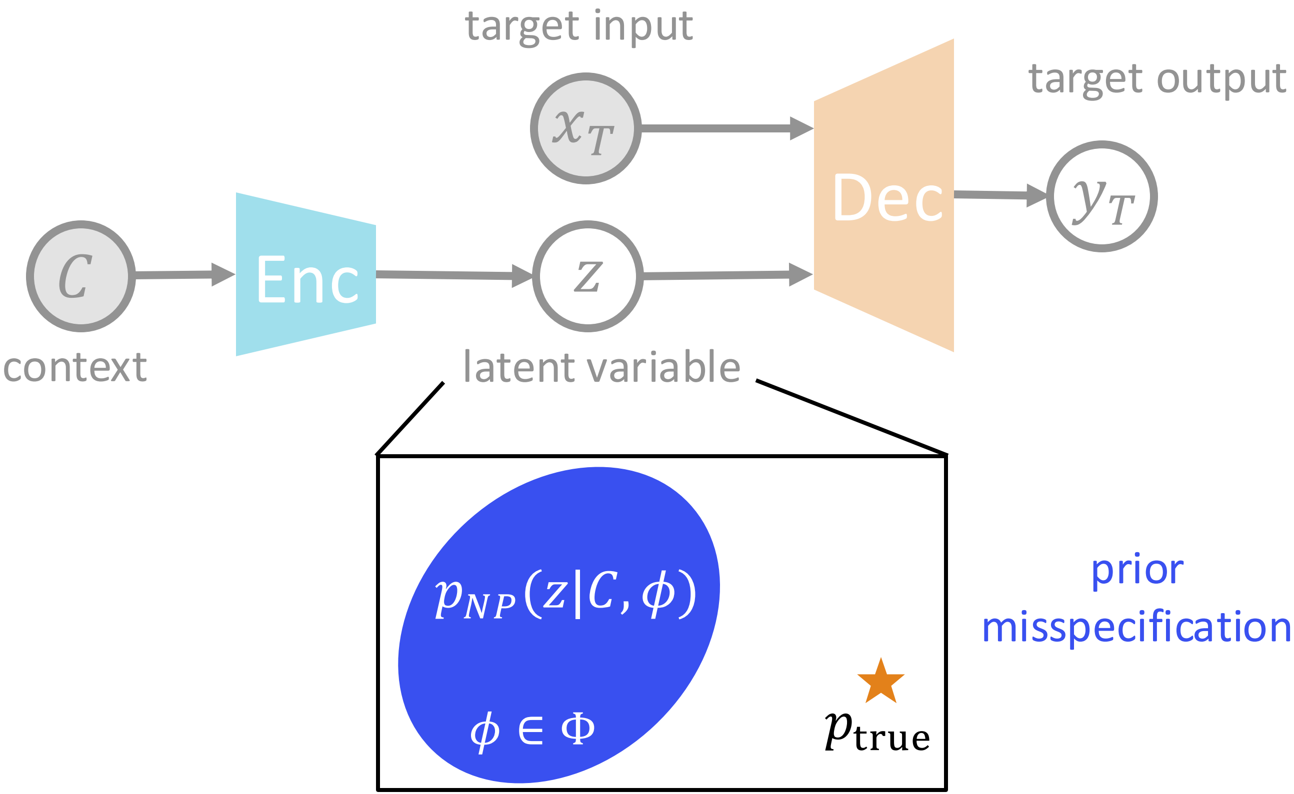
NPs empowers an interpretable and robust decision-making process.

 "Spherical Fourier Neural Operators", Bonev et al, ICML 2023

Challenge with encoding the context information & Prior misspecification

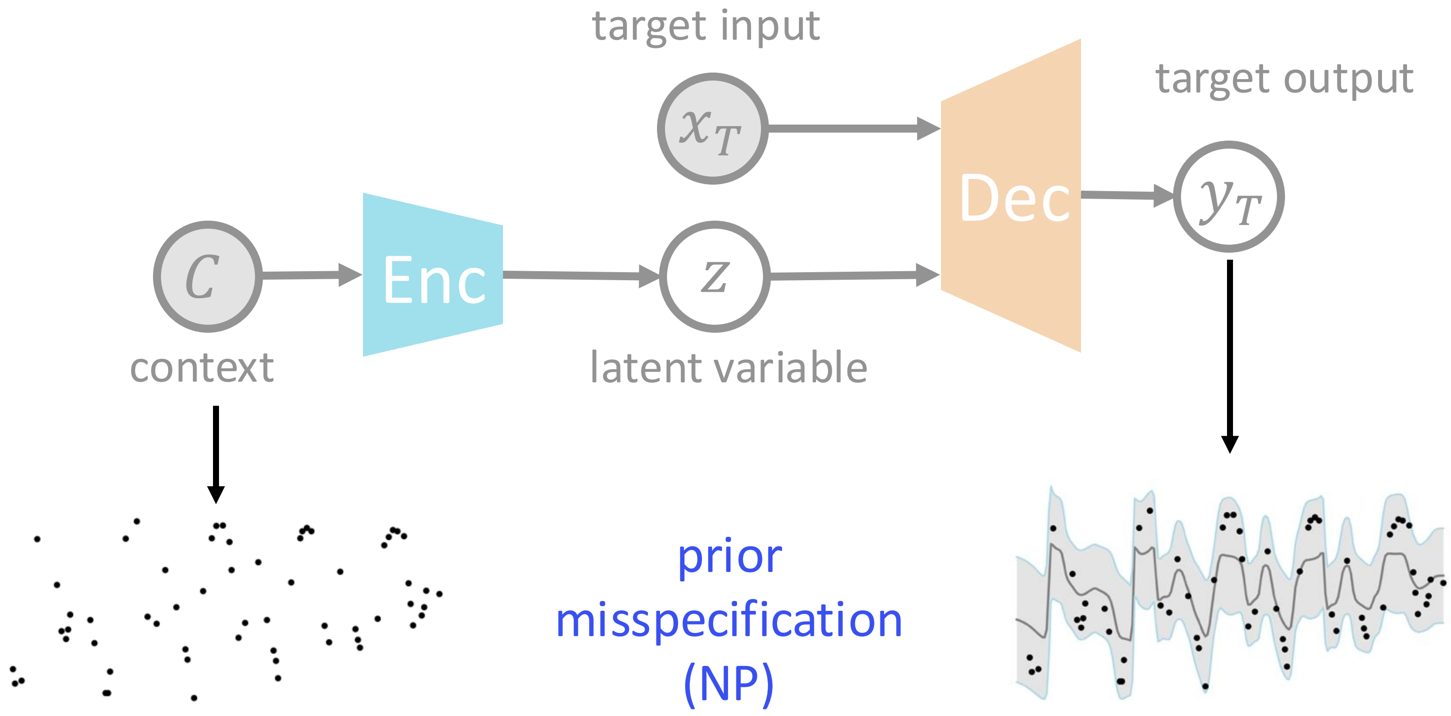


Challenge with encoding the context information & Prior misspecification

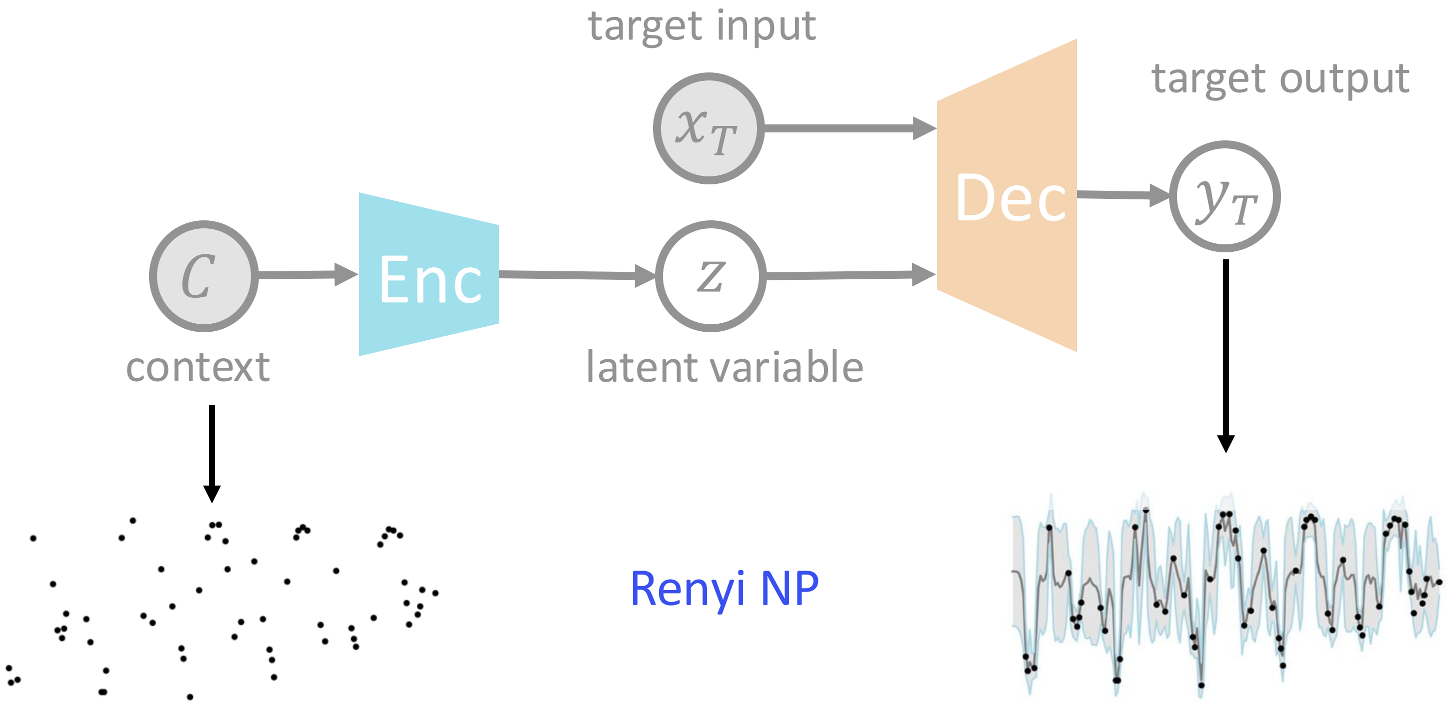


prior
misspecification

Challenge with encoding the context information & Prior misspecification



Challenge with encoding the context information & Prior misspecification



Renyi NP

Cause for prior misspecification

$$-\mathcal{L}_{VI}(\theta, \phi) = \mathbb{E}_{q_\phi(\boldsymbol{z}|C,T)} \log p_\theta(Y_T|X_T, \boldsymbol{z}) - D_{KL}(q_\phi(\boldsymbol{z}|C,T) \| p(\boldsymbol{z}|C))$$

Cause for prior misspecification

Likelihood　　Posterior　　Prior

$$-\mathcal{L}_{VI}(\theta,\phi) = \mathbb{E}_{q_\phi(\boldsymbol{z}|C,T)} \log p_\theta(Y_T|X_T,\boldsymbol{z}) - D_{KL}(q_\phi(\boldsymbol{z}|C,T)||p(\boldsymbol{z}|C))$$

Cause for prior misspecification

Likelihood — Posterior — Prior

$$-\mathcal{L}_{VI}(\theta,\phi) = \mathbb{E}_{q_\phi(\boldsymbol{z}|C,T)} \log p_\theta(Y_T|X_T,\boldsymbol{z}) - D_{KL}(q_\phi(\boldsymbol{z}|C,T)||p(\boldsymbol{z}|C))$$

$q_\phi(\boldsymbol{z}|C)$

approximated prior
(parameter coupling)

Cause for prior misspecification

Likelihood   Posterior   Prior

$$-\mathcal{L}_{VI}(\theta,\phi) = \mathbb{E}_{q_\phi(\boldsymbol{z}|C,T)} \log p_\theta(Y_T|X_T,\boldsymbol{z}) - D_{KL}(q_\phi(\boldsymbol{z}|C,T)||p(\boldsymbol{z}|C))$$

$q_\phi(\boldsymbol{z}|C)$

approximated prior
(parameter coupling)

Prior misspecification: $q_\phi(\boldsymbol{z}|C) \neq p(\boldsymbol{z}|C),\ \forall\, \phi \in \Phi$

What if we don’t have to trust this approximated prior?

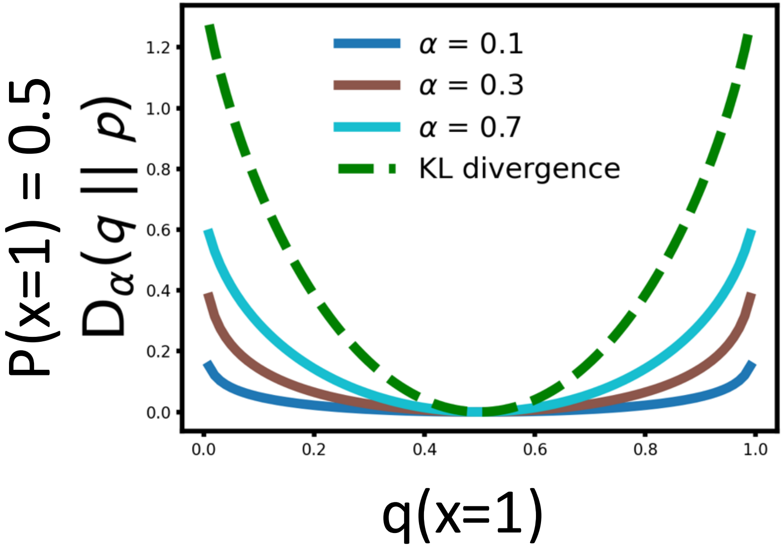
Renyi divergence (1961):

$$D_\alpha\big(q(\boldsymbol{z})||p(\boldsymbol{z})\big) = \frac{1}{\alpha - 1}\log \int q(\boldsymbol{z})^{\alpha} p(\boldsymbol{z})^{1-\alpha}\, d\boldsymbol{z}$$

- $D_\alpha \to D_{KL}$ as $\alpha \to 1$

Renyi divergence (1961):

$$D_\alpha\big(q(\boldsymbol{z})||p(\boldsymbol{z})\big) = \frac{1}{\alpha - 1}\log \int q(\boldsymbol{z})^\alpha p(\boldsymbol{z})^{1-\alpha}\, d\boldsymbol{z}$$



- Smaller α, less trust on prior
- Larger α, more trust on prior

Renyi neural processes (RNP)

$$-\mathcal{L}_{RNP}(\theta,\phi)=\frac{1}{1-\alpha}\log\mathbb{E}_{q_\phi(\boldsymbol{z}|C,T)}\left[\frac{p_\theta(Y_T|X_{\boldsymbol{T}},\boldsymbol{z})q_\phi(\boldsymbol{z}|C)}{q_\phi(\boldsymbol{z}|C,T)}\right]^{1-\alpha}$$

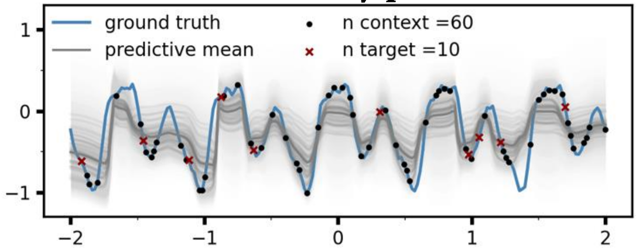
Renyi neural processes (RNP)

$$-\mathcal{L}_{RNP}(\theta, \phi) = \frac{1}{1-\alpha} \log \mathbb{E}_{q_\phi(\boldsymbol{z}|C,T)} \left[\frac{p_\theta(Y_T|X_{\boldsymbol{T}}, \boldsymbol{z}) q_\phi(\boldsymbol{z}|C)}{q_\phi(\boldsymbol{z}|C,T)} \right]^{1-\alpha}$$

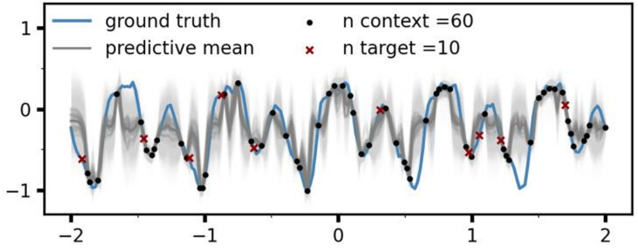
- $\alpha = 0,\quad \mathcal{L}_{RNP} = \mathcal{L}_{ML}$ (maximizing likelihood estimation)
- $\alpha \to 1,\quad \mathcal{L}_{RNP} = \mathcal{L}_{VI}$ (variational inference)
- $\alpha \in (0, 1)$, RNP mitigates the prior misspecification via α

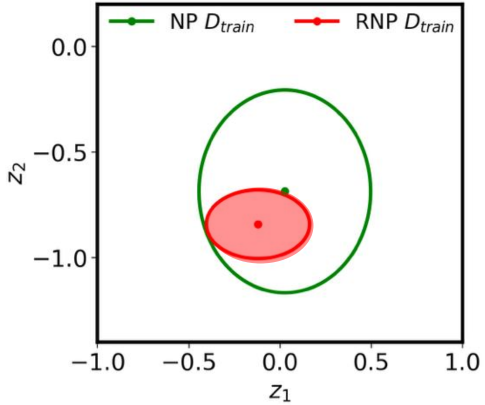
Vanilla NP vs RNP on a GP-regression benchmark

NP ($\mathcal{L}_{VI}$)



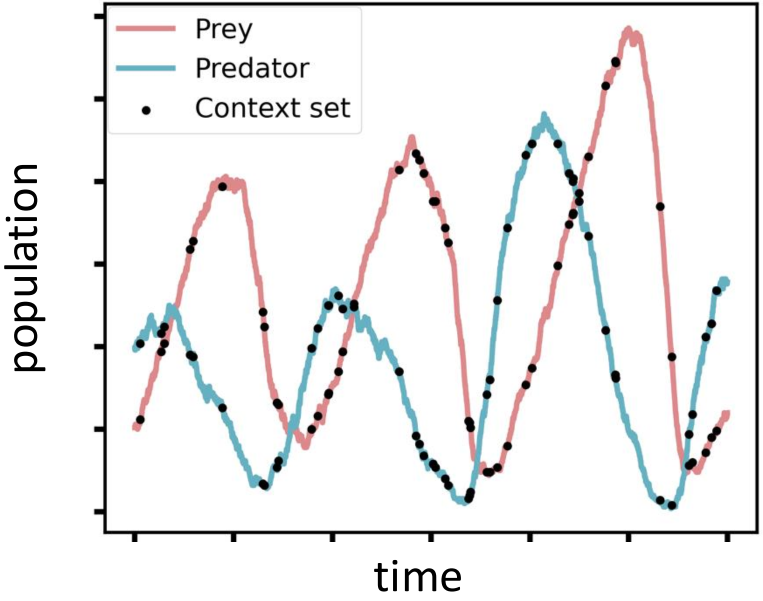
RNP ($\alpha = 0.7$)





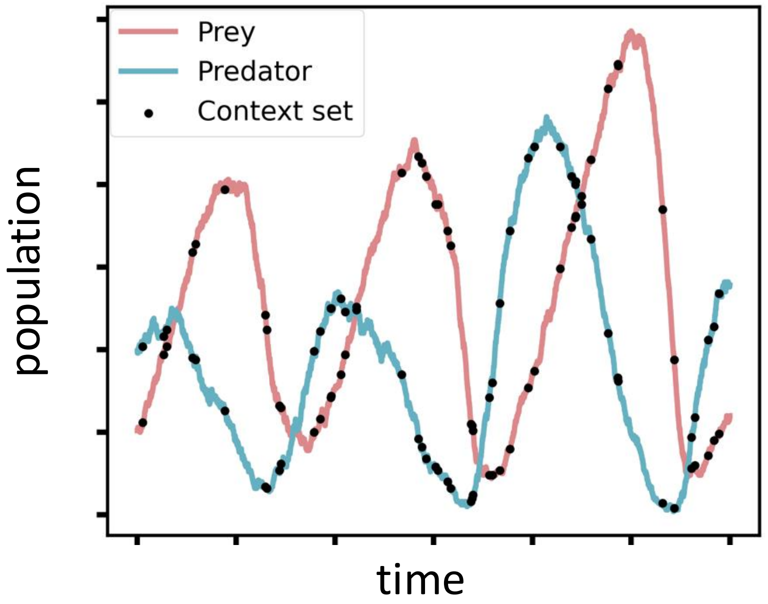
Better variance estimation

Simulation to real-world example

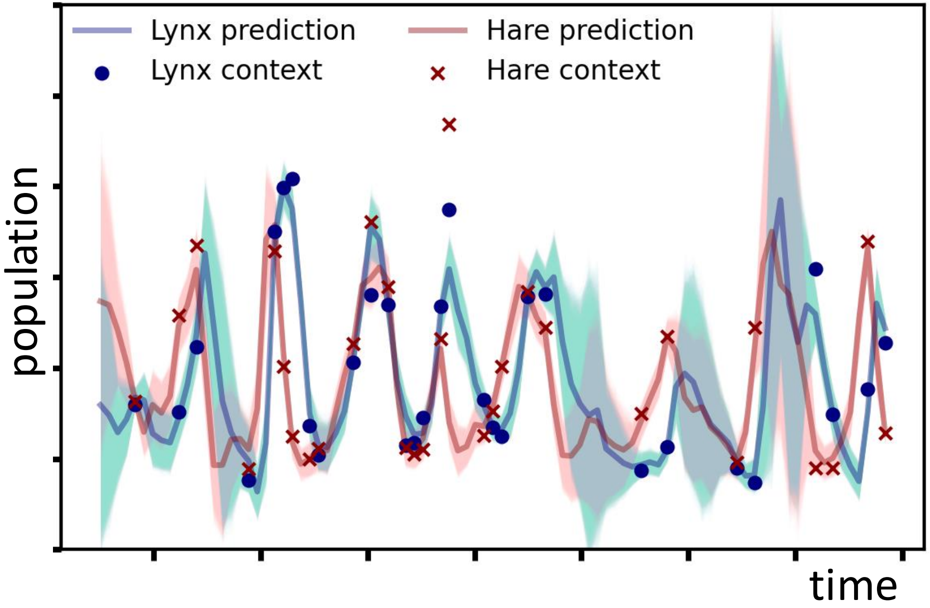


Sim: Prey-predator data D_{train}

Simulation to real-world example



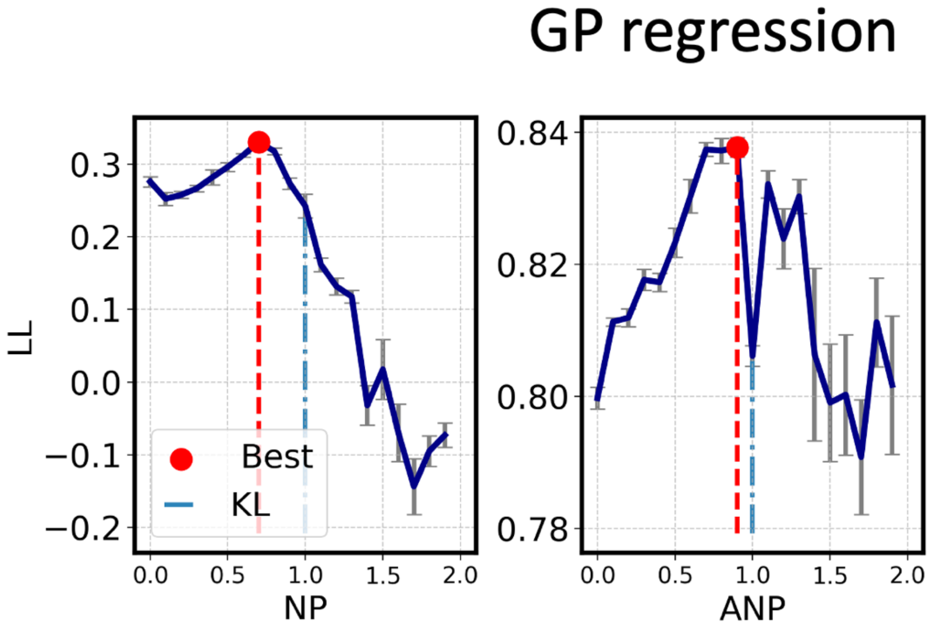
Sim: Prey-predator data D_{train}



Real: Hare - lynx data D_{test}

$$LL_{RNP} = \mathbf{-3.63 \pm 0.09} \uparrow$$
$$LL_{NP} = -4.44 \pm 0.41$$

α tuning



- $\alpha = 0.7$ empirically generalizes well
- Optimal α can be found using cross-validation
- Heuristic: annealing α from 1 to 0

Renyi neural processes:

Our contributions:

- Identification of prior misspecification in neural processes
- Renyi divergence to mitigate prior misspecification
- Unification of two NP objectives with RNP

Poster session 5 (ID 43943) (11 a.m. — 1:30 p.m. PDT)

Email: xuesong.wang@data61.csiro.au

Paper

Github