Autoregressive Conditional Neural Processes

Wessel P. Bruinsma[*12], Stratis Markou[*2], James Requiema[*2],
Andrew Y. K. Foong[*1], Tom R. Andersson[3], Anna Vaughan[2],
Anthony Buonomo[2], J. Scott Hosking[34], Richard E. Turner[12]

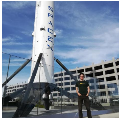
[*]Equal contribution

[1]Microsoft Research AI4Science, [2]University of Cambridge,
[3]British Antarctic Survey, [4]The Alan Turing Institute

The Eleventh International Conference on Learning Representations (ICLR 2023)

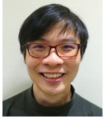
Collaborators

Wessel Bruinsma*[1,2]



Stratis Markou*[2]

James Requeima*[2]



Andrew Foong*[1]

Tom Andersson[3]

Anna Vaughan[2]

Anthony Buonomo[2]

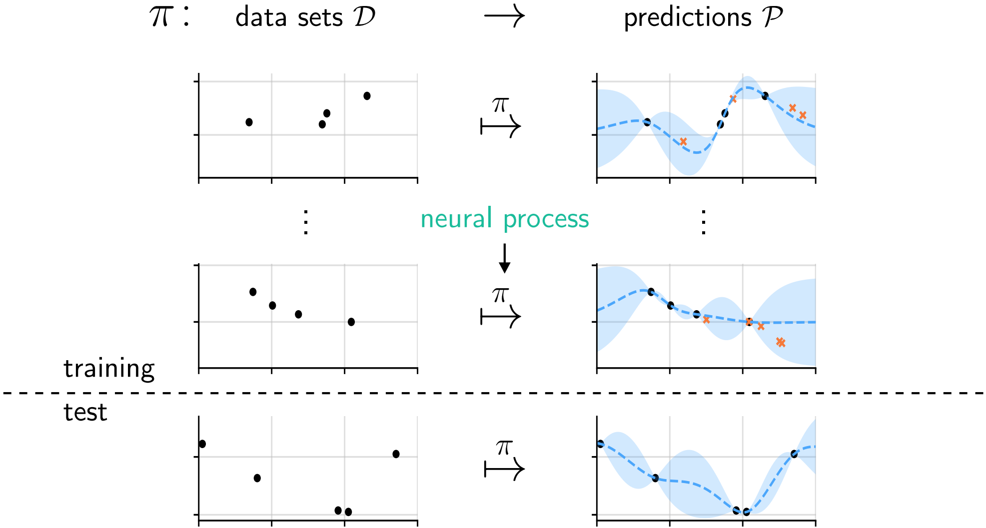
J. Scott Hosking[3,4]

Rich Turner*[1,2]

*Equal contribution

[1]Microsoft Research AI4Science, [2]University of Cambridge,
[3]British Antarctic Survey, [4]The Alan Turing Institute

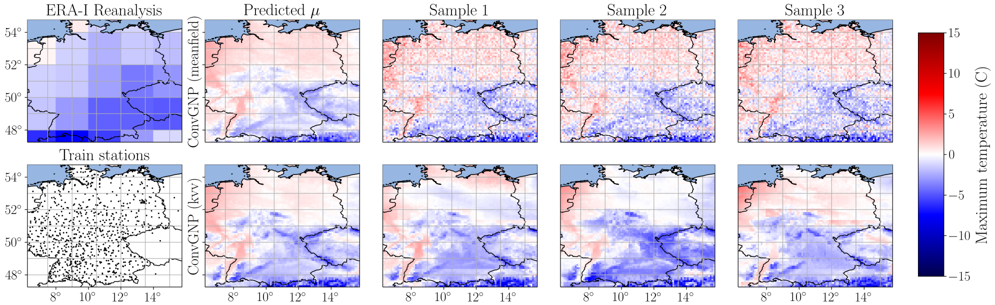
Meta-Learning and Neural Processes



The Appeal of Neural Processes

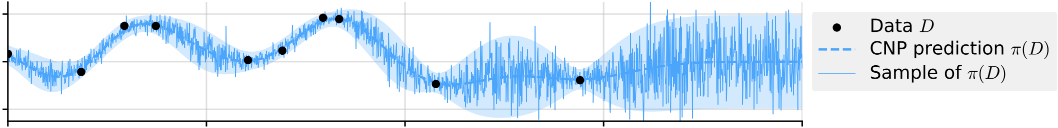
- ✓ Extremely versatile and flexible
- ✓ Fast, probabilistic predictions
- ✓ Simple to train
- ✓ Work well in practice

- Climate model downscaling (Markou et al., 2022):



But Neural Processes Are Not Without Challenges...

- Conditional neural process (CNP; Garnelo, Rosenbaum, et al., 2018):



	Correlated predictions	Non-Gaussian predictions	Exact training	Consistent predictions
CNPs (Garnelo, Rosenbaum, et al., 2018)	✗	✓	✓	✓
Gaussian NPs (Markou et al., 2022)	✓	✗	✓	✓
Latent-variable NPs (Garnelo, Schwarz, et al., 2018)	✓	✓	✗	✓
Autoregressive CNPs (AR CNPs; this work!)	✓	✓	✓	✗

Autoregressive Conditional Neural Processes

- Idea: feed output of CNP back into the model in an autoregressive fashion:

$$q^{(\text{AR CNP})}(\mathbf{y}_{1:3} \mid D) = q(y_1 \mid D)q(y_2 \mid y_1, D)q(y_3 \mid y_1, y_2, D).$$

CNP pred. of y_3 given y_1, y_2, and D

- AR modelling certainly not new, but not yet explored for NPs.

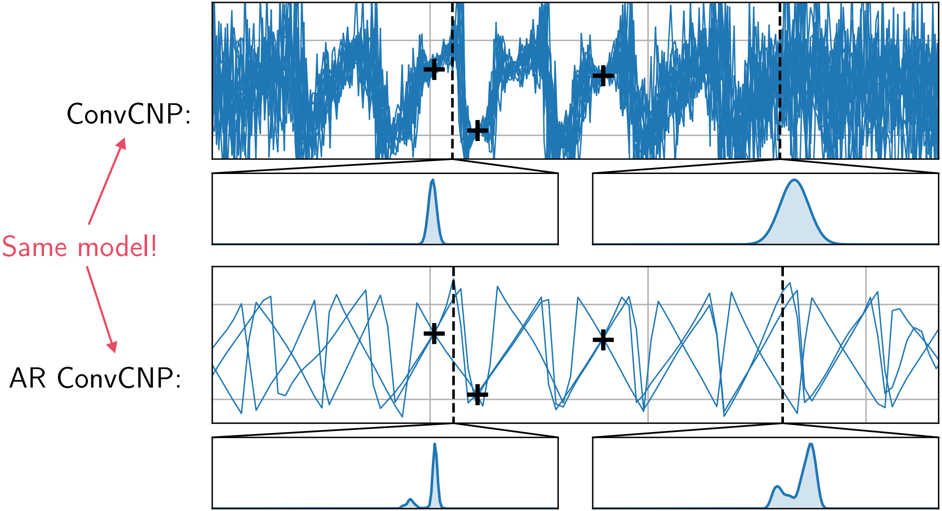
✓ Correlated and non-Gaussian predictions!

✓ No modifications to model or training procedure!

✗ Predictions depend on number and order of data (predictions no longer consistent)

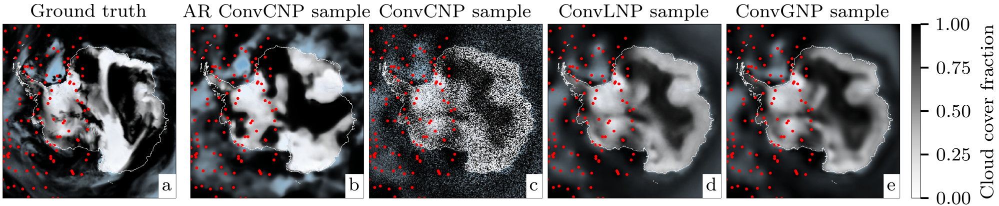
✗ Requires multiple forward passes of CNP (Prop. 2.2 offers a partial remedy!)

Example: ConvCNP (Gordon et al., 2020) Trained on Sawtooth Data



So What Else Is in the Paper?

- Prop. 2.1: In an idealised case, AR CNPs are guaranteed to perform better than GNPs.
- A detailed comparison of AR CNPs and neural density estimators (NDEs).
- Exceptional performance of the AR ConvCNP (Gordon et al., 2020) in 60 synthetic scenarios.
- A variety of real-world experiments, including a challenging cloud cover experiment:



Code: `https://github.com/wesselb/neuralprocesses`

Please come see us at the poster, or contact us at `wbruinsma@microsoft.com`! :)