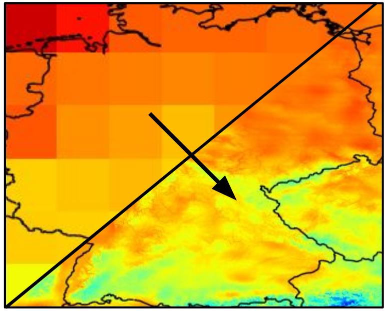
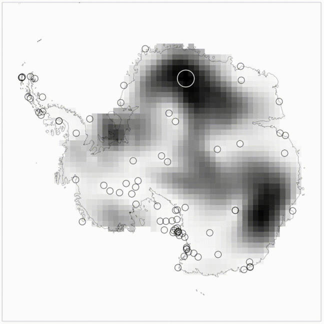
Tackling diverse environmental prediction tasks with neural processes

NERC AI in Environmental Science Webinar, Aug 4th 2023

Tom Andersson (BAS)





Collaborators


Wessel Bruinsma


Stratis Markou


James Requiema


Alejandro Coca-Castro


Anna Vaughan


Jonas Scholz


Paolo Pelucchi


Anna-Louise Ellis


Matthew Lazzara


Dani Jones


Scott Hosking


Rich Turner

British Antarctic Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

UNIVERSITY OF CAMBRIDGE

Microsoft

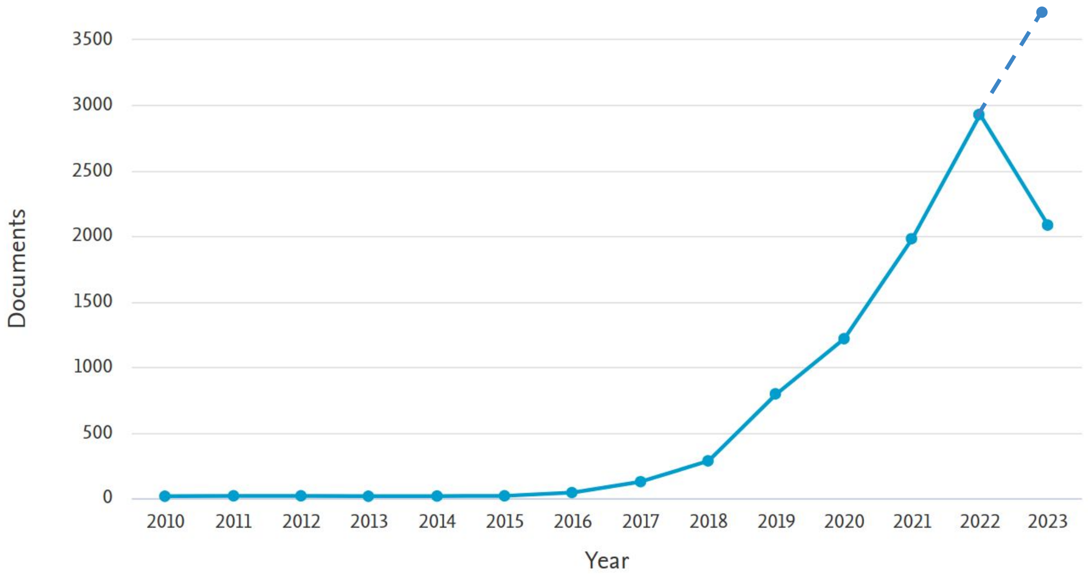
The Alan Turing Institute

Talk outline

1) **The state of play**
2) Modelling environmental observations
3) Convolutional neural processes
4) Data fusion experiments
5) DeepSensor
6) Closing thoughts

Deep learning: From the fringes to the frontiers of environmental science

Environmental science journal articles with “deep learning” in title or abstract



Source: Elsevier Scopus

Revolution in ML weather forecasting over the past year

Phys.org

t

Google DeepMind

[Submitted on 24 Dec 2022]

GraphCast: Learning skillful medium-range global weather forecasting

Microsoft Research

[Submitted on 24 Jan 2023 (v1), last revised 10 Jul 2023 (this version, v3)]

ClimaX: A foundation model for weather and climate

Shanghai academics

[Submitted on 6 Apr 2023]

FengWu: Pushing the Skillful Global Medium-range Weather Forecast beyond 10 Days Lead

Google DeepMind

[Submitted on 6 Jun 2023 (v1), last revised 6 Jul 2023 (this version, v3)]

Deep Learning for Day Forecasts from Sparse Observations

Huawei

Article | Open Access | Published: 05 July 2023

Accurate medium-range global weather forecasting with 3D neural networks

ECMWF

[Submitted on 19 Jul 2023]

The rise of data-driven weather forecasting

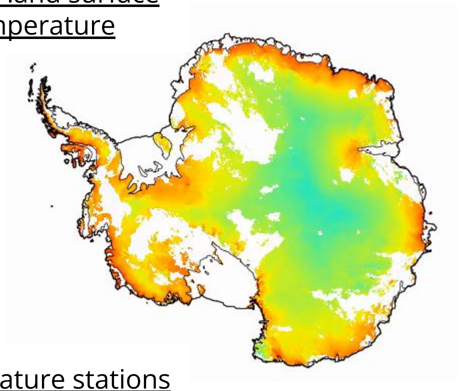
"A new NWP paradigm is emerging relying on inference from ML models..." — ECMWF

Talk outline

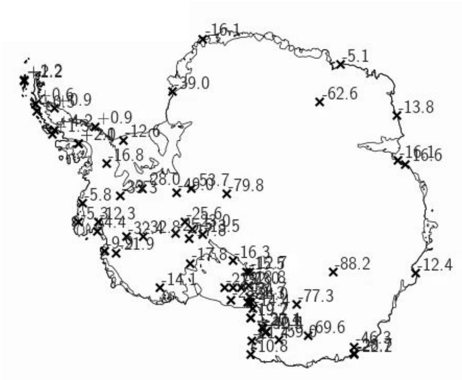
1) The state of play
2) **Modelling environmental observations**
3) Convolutional neural processes
4) Data fusion experiments
5) DeepSensor
6) Closing thoughts

Wait, what if we have missing data or irregularly sampled data?

MODIS land surface temperature

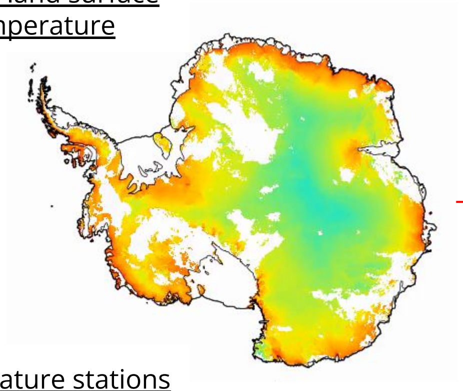


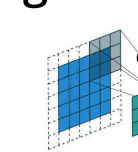
Temperature stations

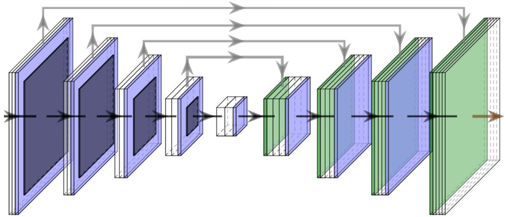


Wait, what if we have missing data or irregularly sampled data?

MODIS land surface temperature



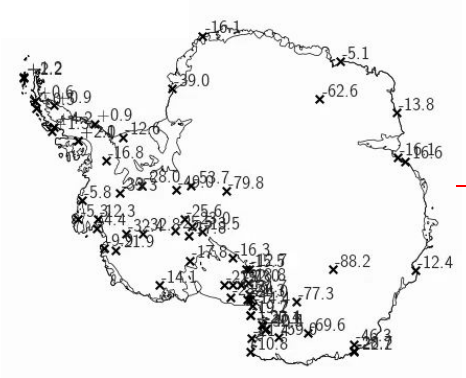




Convolutional neural network (CNN)

❌ Can't handle NaNs

Temperature stations

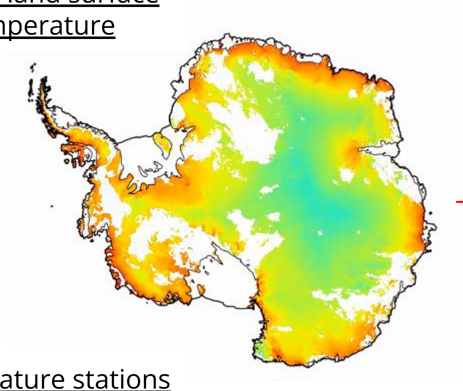


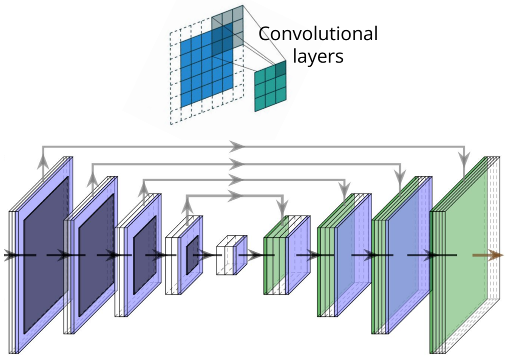
Gaussian process?

❌ Can't ingest multiple data streams
❌ Computationally expensive
❌ Not very expressive

Wait, what if we have missing data or irregularly sampled data?

MODIS land surface temperature

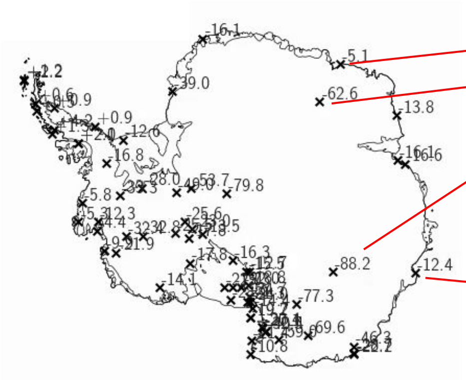


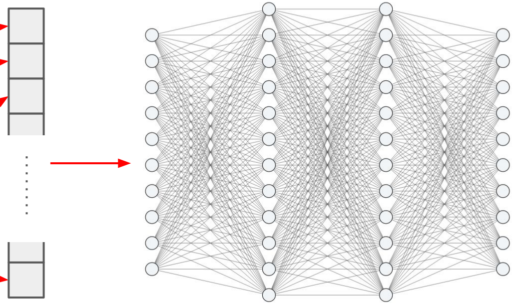


Convolutional neural network (CNN)

❌ Can’t handle NaNs

Temperature stations



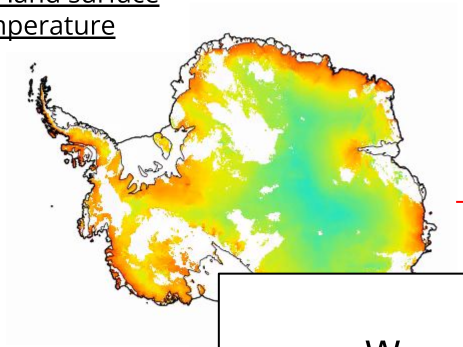


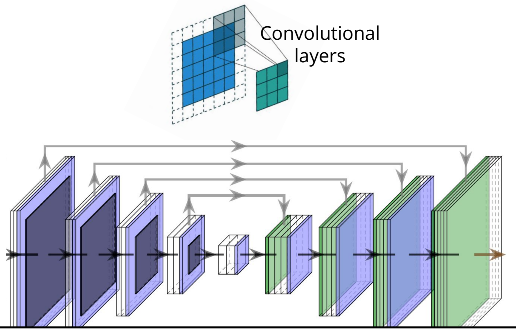
Multilayer perceptron (MLP)

❌ Can’t handle NaNs
❌ Can’t leverage new stations
❌ Depends on station ordering

Wait, what if we have missing data or irregularly sampled data?

MODIS land surface temperature

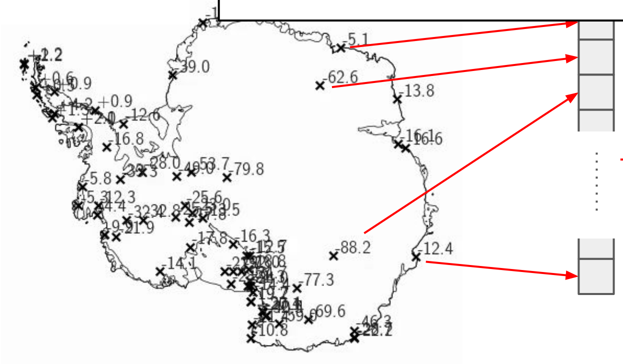


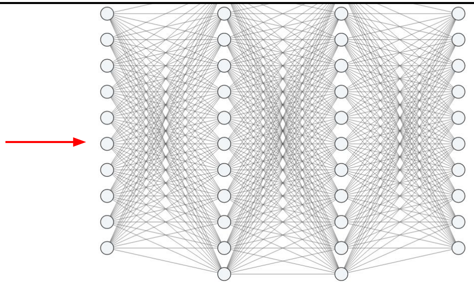


- ❌ Can't handle NaNs

We need to model these observations as *sets*!

Temperature stations





- ❌ Can't handle NaNs
- ❌ Can't leverage new stations
- ❌ Depends on station ordering

Talk outline

1) The state of play
2) Modelling environmental observations
3) **Convolutional neural processes**
4) Data fusion experiments
5) DeepSensor
6) Closing thoughts

Context sets, target sets, tasks, and convolutional neural processes

Context sets: $C_i = \{(\mathbf{x}_1^{(c)}, \mathbf{y}_1^{(c)}), (\mathbf{x}_2^{(c)}, \mathbf{y}_2^{(c)}), \ldots, (\mathbf{x}_{n_{c_i}}^{(c)}, \mathbf{y}_{n_{c_i}}^{(c)})\}_i$

Target sets: $T_i = \{(\mathbf{x}_1^{(t)}, \mathbf{y}_1^{(t)}), (\mathbf{x}_2^{(t)}, \mathbf{y}_2^{(t)}), \ldots, (\mathbf{x}_{n_{t_i}}^{(t)}, \mathbf{y}_{n_{t_i}}^{(t)})\}_i$

Task: $\mathcal{D} = \{C_{1:N_c}, T_{1:N_t}\}$

ConvNP: $p_\theta(\mathbf{y}^{(t)}; \mathbf{x}^{(t)}, C) = \mathcal{N}(\mathbf{y}^{(t)}; \mu(\mathbf{x}^{(t)}, \mathrm{Enc}(C)), \sigma^2(\mathbf{x}^{(t)}, \mathrm{Enc}(C)))$

Permutation invariant encoder

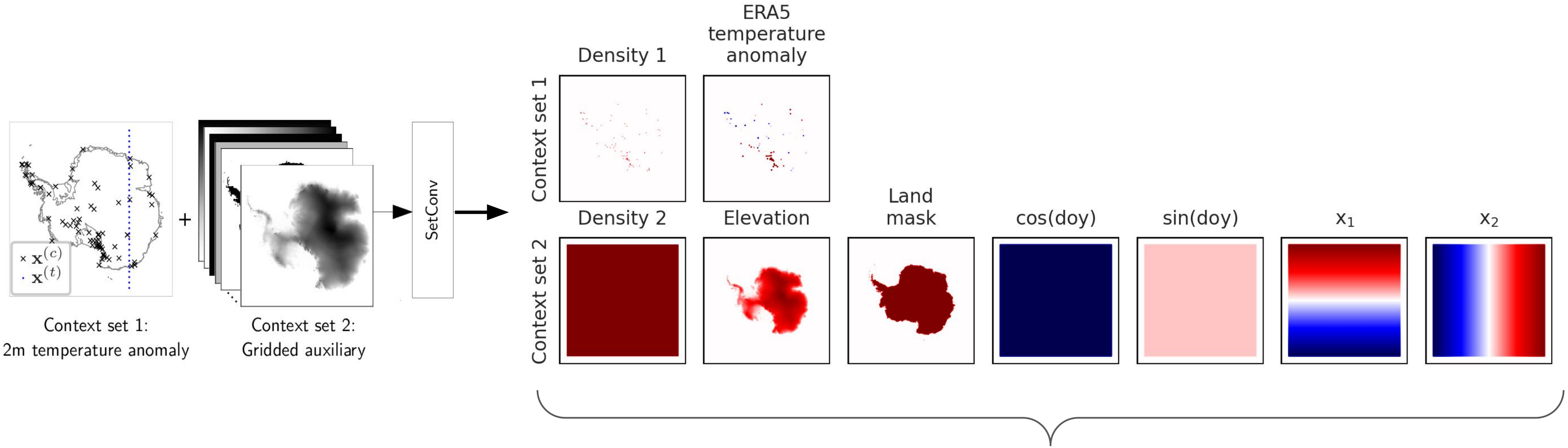
Model parameters trained using maximum likelihood

Parameterised by a CNN (translation equivariant)

<u>Disclaimer</u>: This is the 'conditional' neural process (ConvCNP) variant with a Gaussian output distribution.
I use 'ConvNP' in a variant- and distribution-agnostic way.

M. Garnelo et al., *ICML*, 2018
J. Gordon et al., *ICLR*, 2020

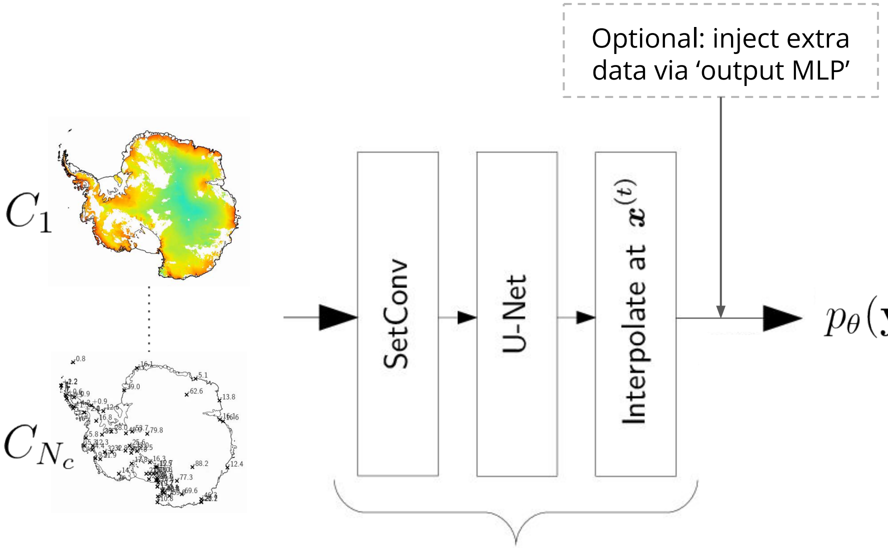
The encoder in ConvNPs: SetConv



3D tensor output by SetConv encoder

- Density channels: observation *locations*
- Data channels: observation *values*

Context sets, target sets, tasks, and convolutional neural processes



Convolutional neural process (ConvNP)

$$p_\theta(\mathbf{y}^{(t)}; \mathbf{x}^{(t)}, \phi(C))$$

Distribution parameters ϕ *depend on the context data* = 'meta-learning'

Key ConvNP capabilities for spatiotemporal modelling

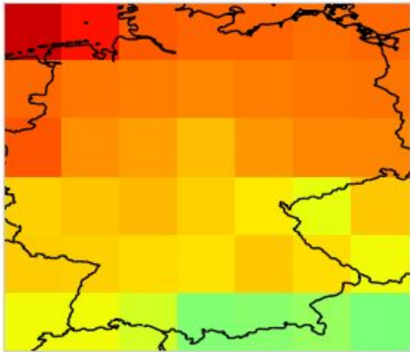
- ✔ Fuses multiple data streams
- ✔ Off-the-grid and gridded modalities
- ✔ Multi-resolution data
- ✔ Missing data
- ✔ Predict at arbitrary target locations
- ✔ Quantify uncertainty
- ✔ $\mathcal{O}(N)$ inference cost (generally)

Talk outline

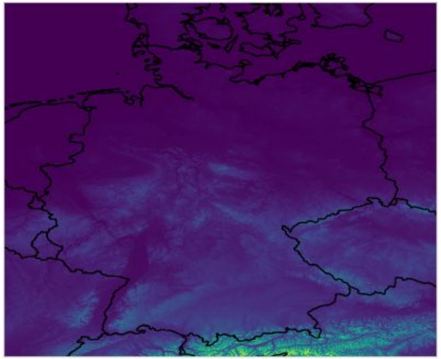
1) The state of play
2) Modelling environmental observations
3) Convolutional neural processes
4) **Data fusion experiments**
5) DeepSensor
6) Closing thoughts

Statistical downscaling with a ConvNP

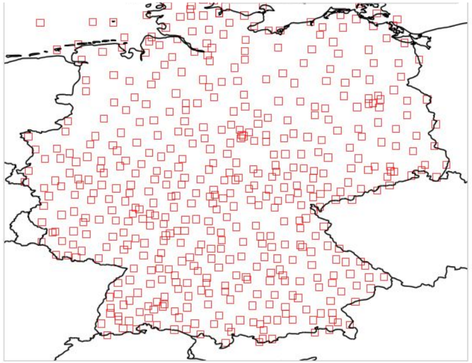
Gridded reanalysis



Auxiliary/background information



In-situ stations



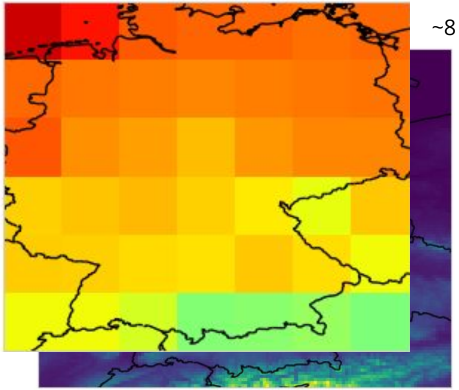
Question: How can we combine these three variables?

Statistical downscaling with a ConvNP

(Coarsened to 1.25° x 1.25° ~= 120 km)

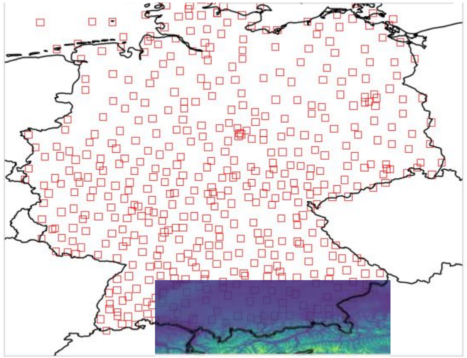
ERA5

~8 km elev



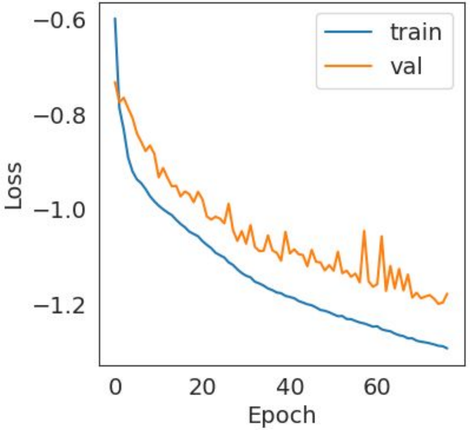
ConvNP

German temperature stations



Inject high-res (500 m) auxiliary data just before prediction

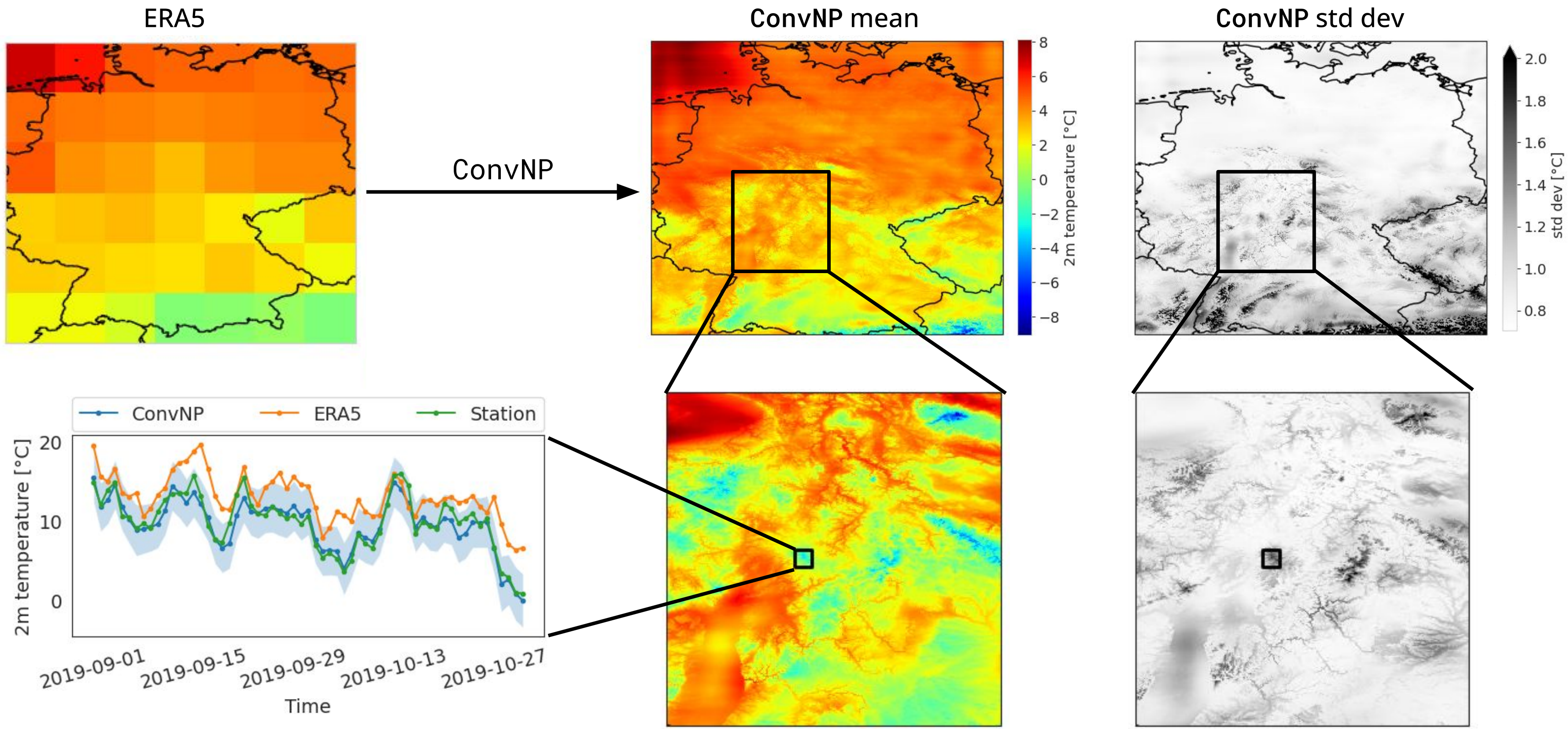
- Train: 2006—2017
- Val: 2018



Vaughan et al., 2022, *GMD*

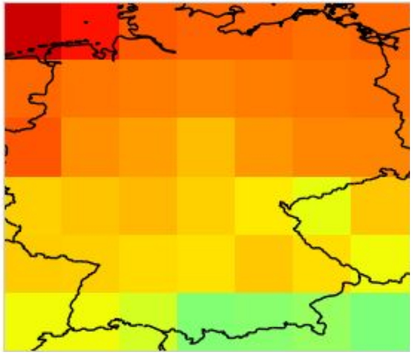
ConvCNP outperforms a suite of statistical downscaling methods.

Statistical downscaling with a ConvNP

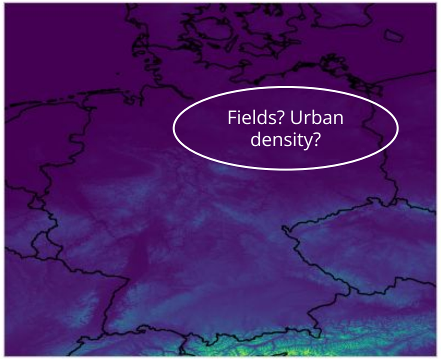


Statistical downscaling caveats

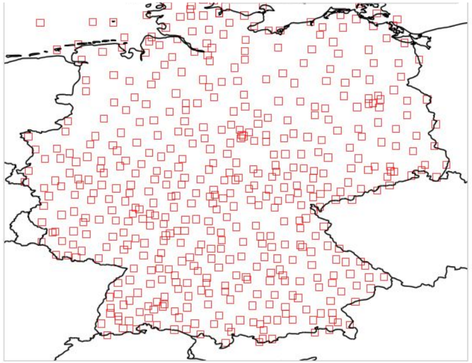
Gridded reanalysis



Auxiliary/background information



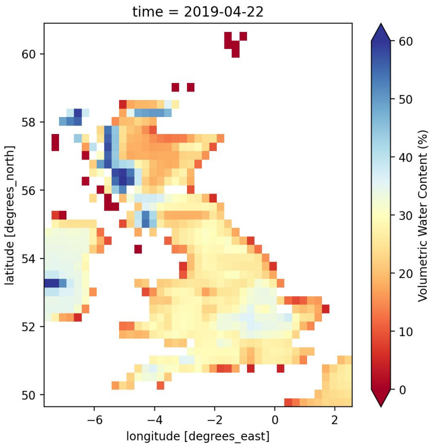
In-situ stations



- Reanalysis sufficiently correlated with station obs
- Availability of auxiliary variables that reveal station microclimate
- Sufficient coverage of stations in auxiliary space

Statistical downscaling caveats

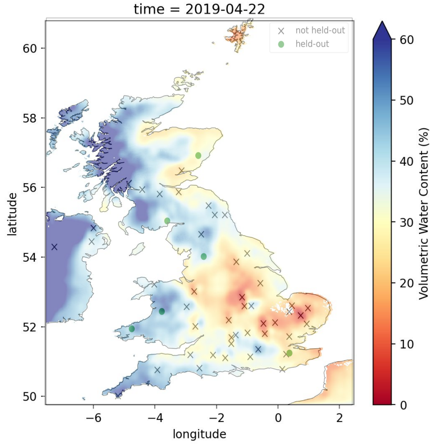
ERA5-Land soil moisture



+

High-res topographic aux data

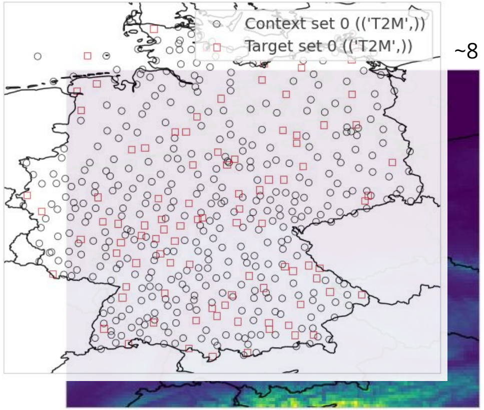
ConvNP



- Training stations are sparse
- ConvNP spatial patterns don't appear realistic
- **Open questions**: Transfer learning? Multi-task learning?

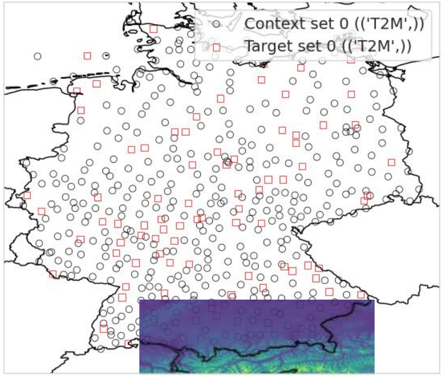
High-resolution station interpolation with ConvNPs

Temperature context stations



ConvNP

Temperature target stations

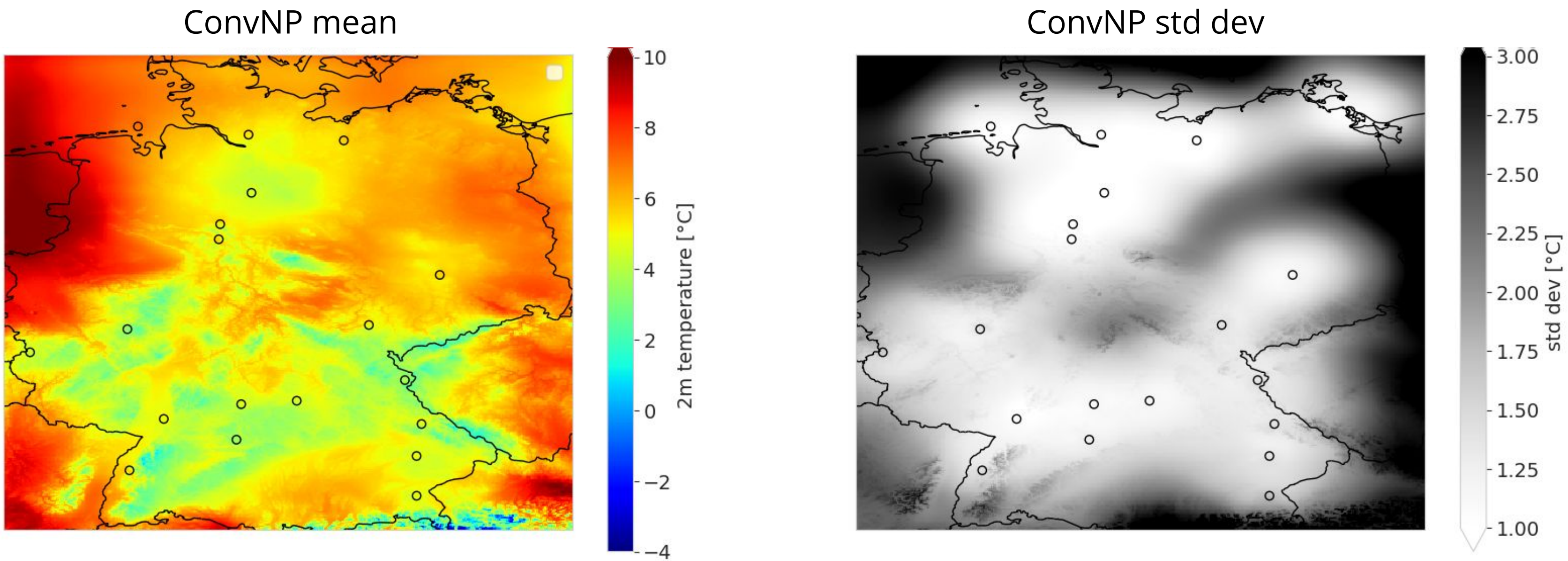


○ = context points

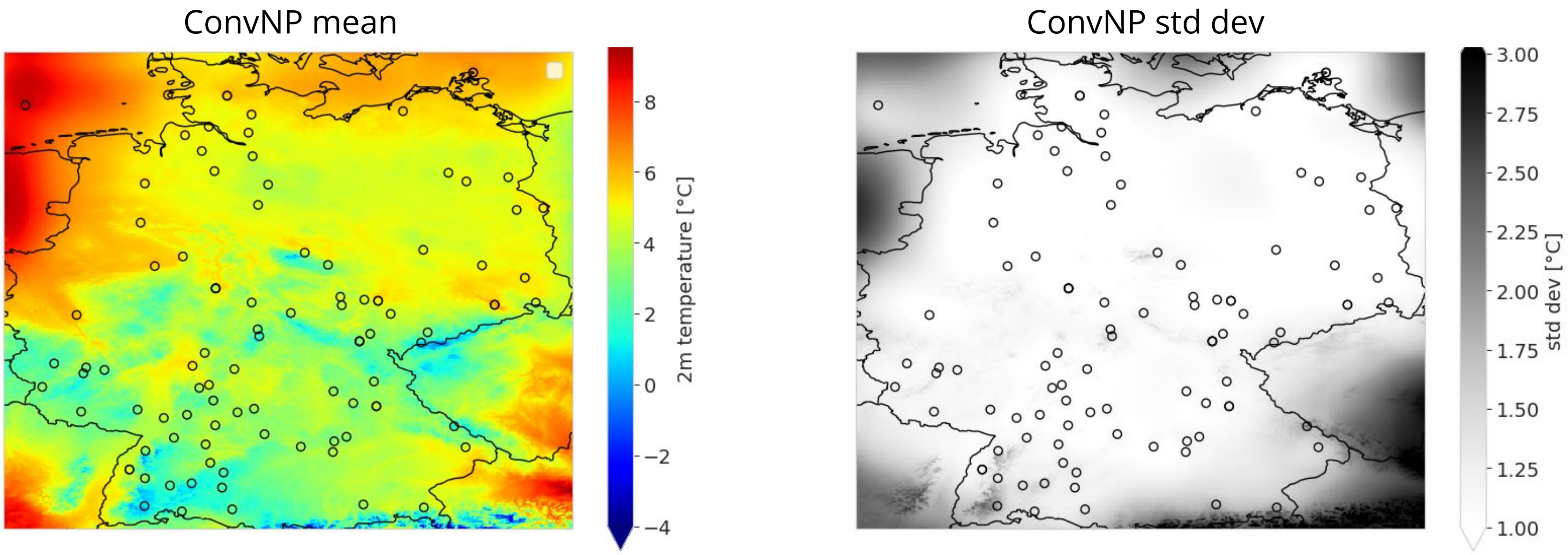
□ = target points

Inject high-res (500 m) auxiliary data just before prediction

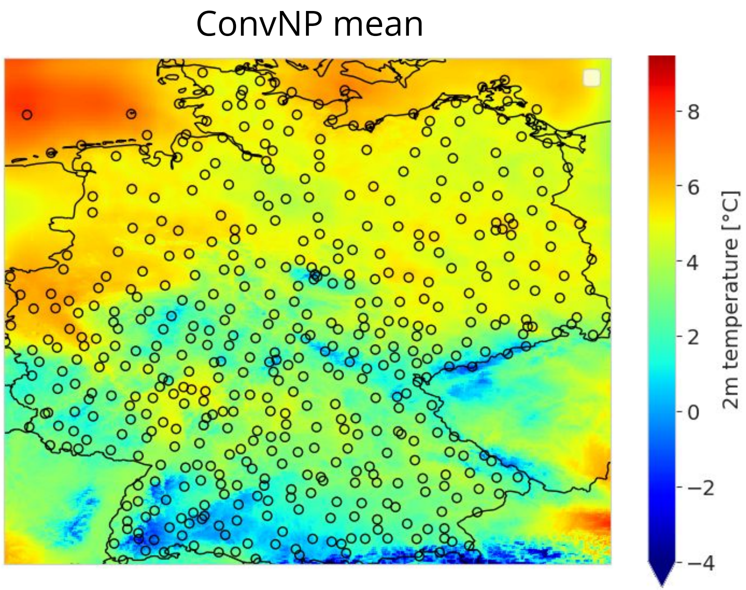
High-resolution station interpolation with ConvNPs

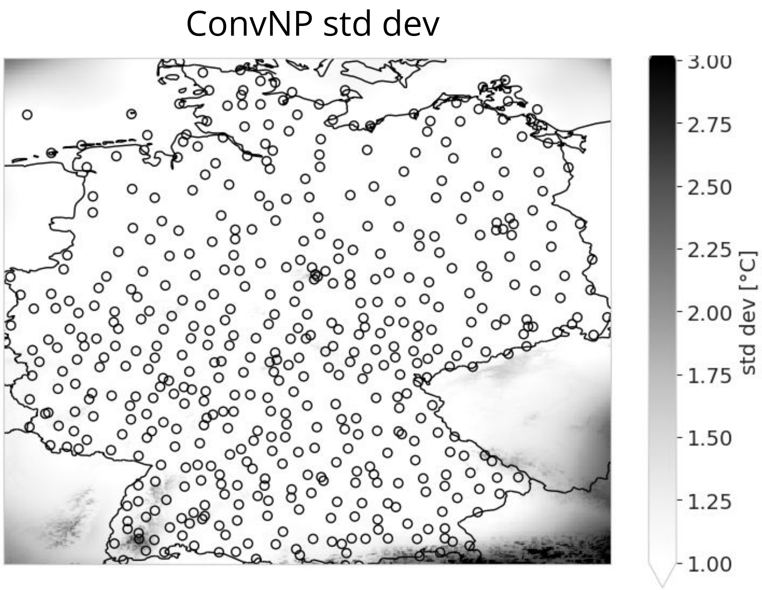


High-resolution station interpolation with ConvNPs

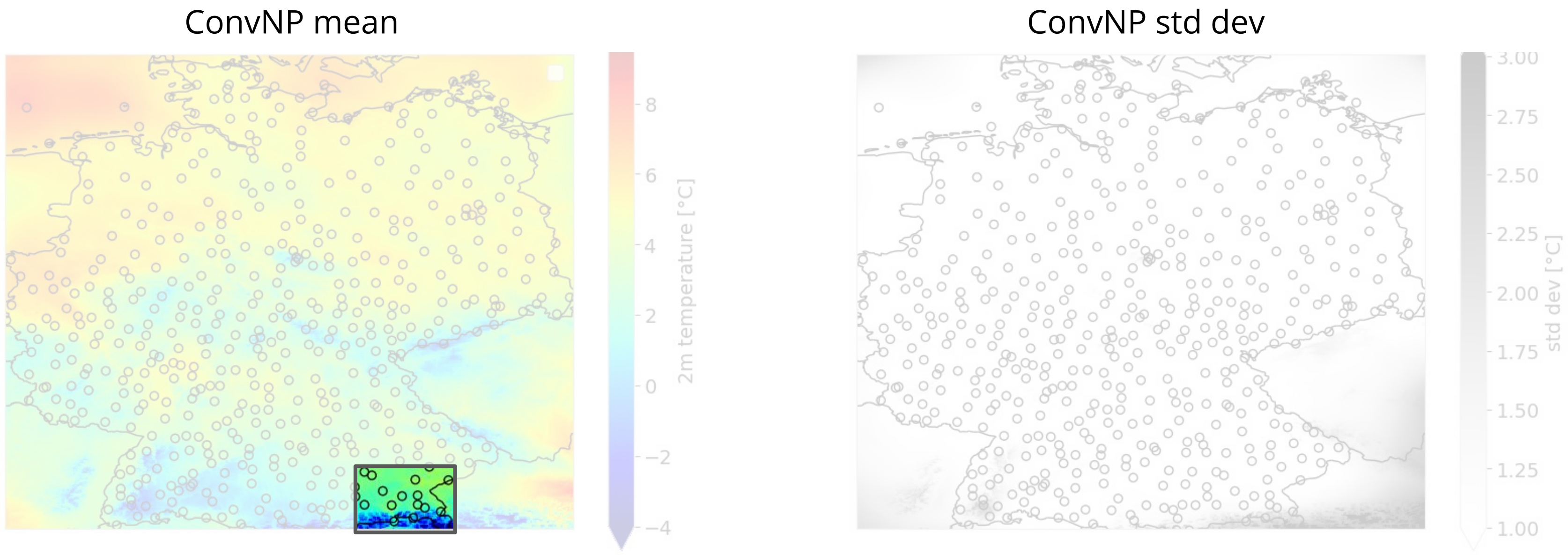


High-resolution station interpolation with ConvNPs

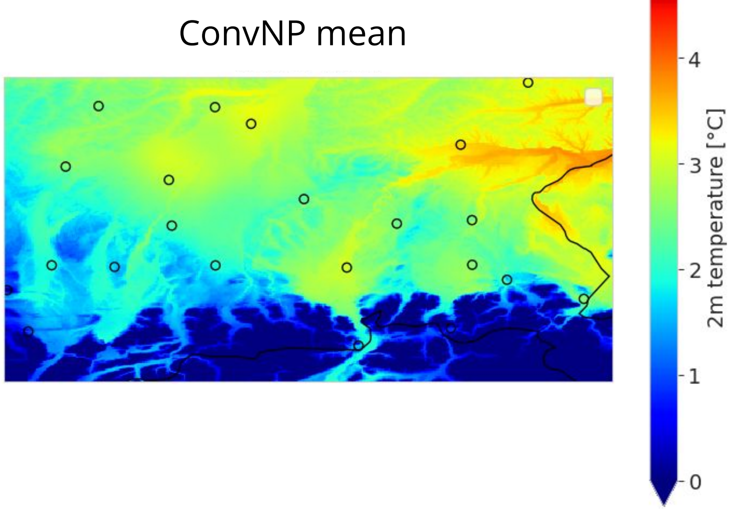


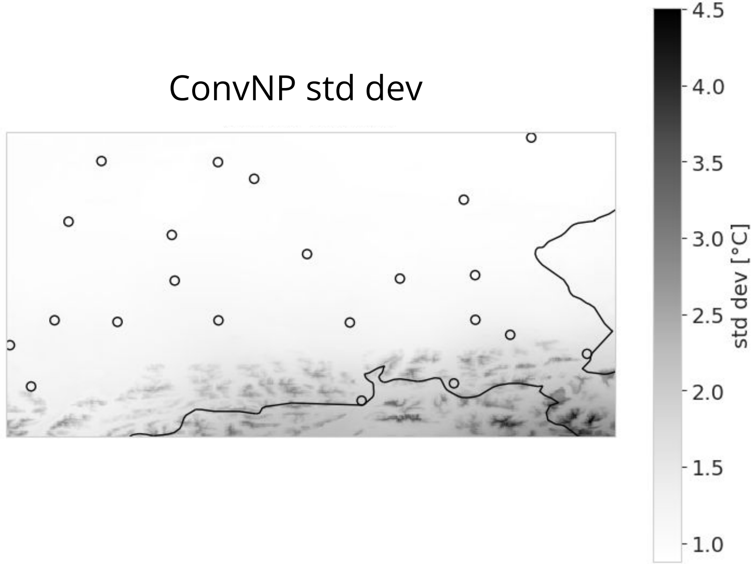


High-resolution station interpolation with ConvNPs



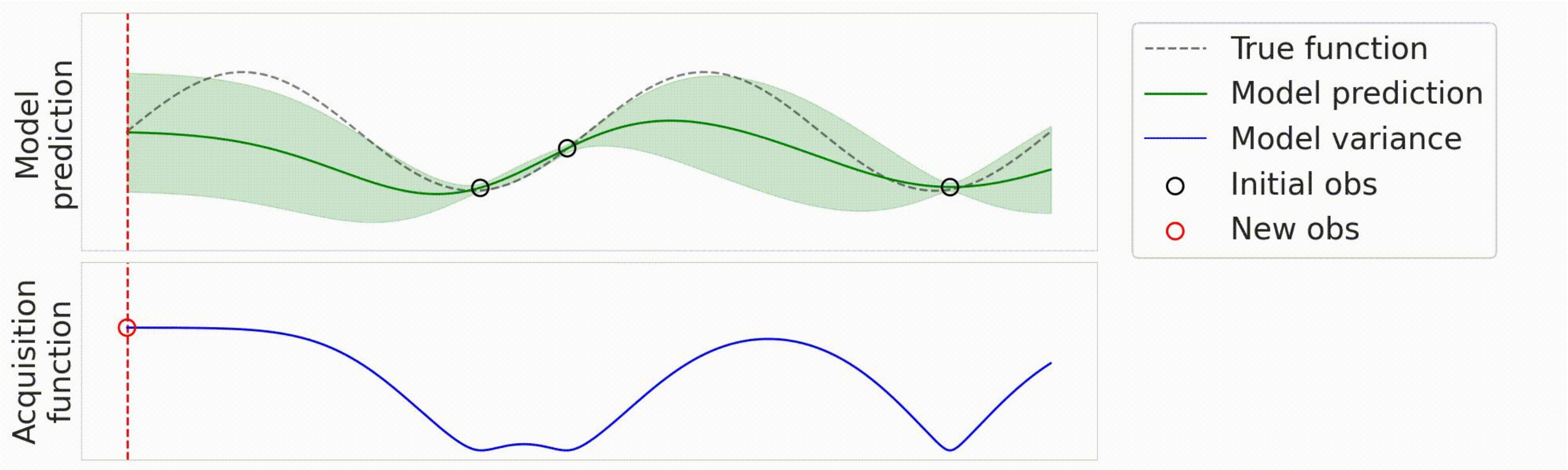
High-resolution station interpolation with ConvNPs





Sensor placement using active learning

- Two ingredients: model $p(\mathbf{y}|\text{data})$ + acquisition function $\alpha(x) \longrightarrow$ placements $\mathbf{x}^*$



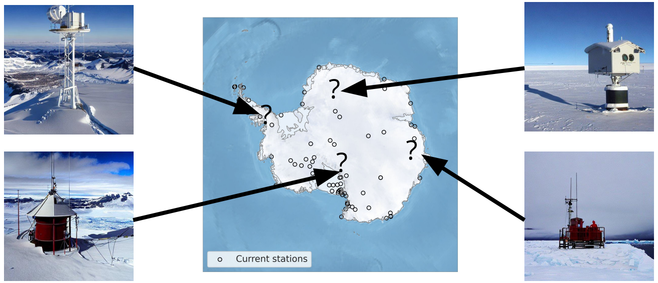
Sensor placement with a ConvGNP

ENVIRONMENTAL DATA SCIENCE

Environmental sensor placement with convolutional Gaussian neural processes

Published online by Cambridge University Press: 03 August 2023 !

Tom R. Andersson, Wessel P. Bruinsma, Stratis Markou, James Requeima, Alejandro Coca-Castro, Anna Vaughan, Anna-Louise Ellis, Matthew A. Lazzara, Dani Jones, Scott Hosking and Richard E. Turner

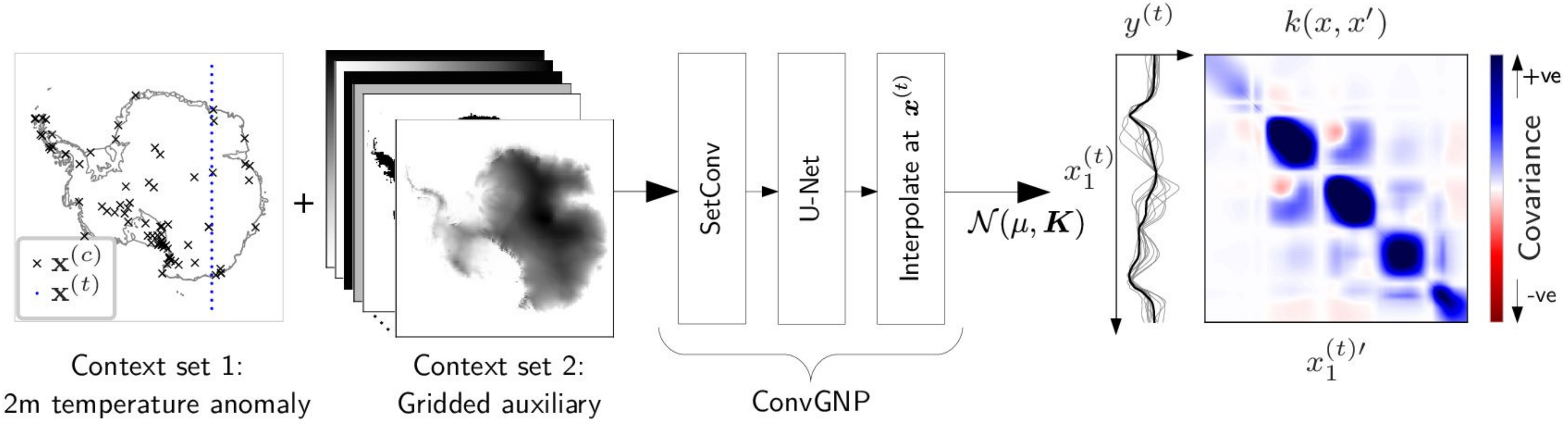


Paper

Conference talk

Sensor placement with a ConvGNP: The model architecture

<u>Convolutional **G**aussian neural process</u>

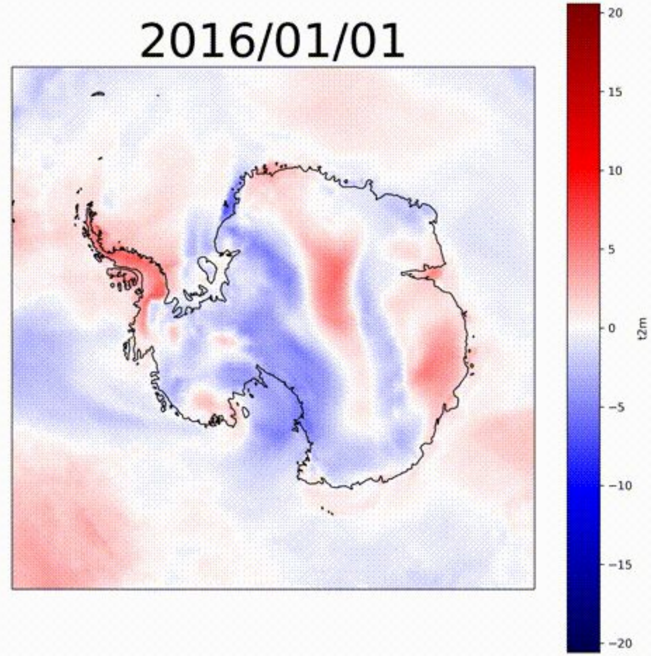


- ✔ Learns to output arbitrary mean & covariance functions given context data
- ✔ Inference scale linearly with dataset size

Sensor placement with a ConvGNP: Training

Paper

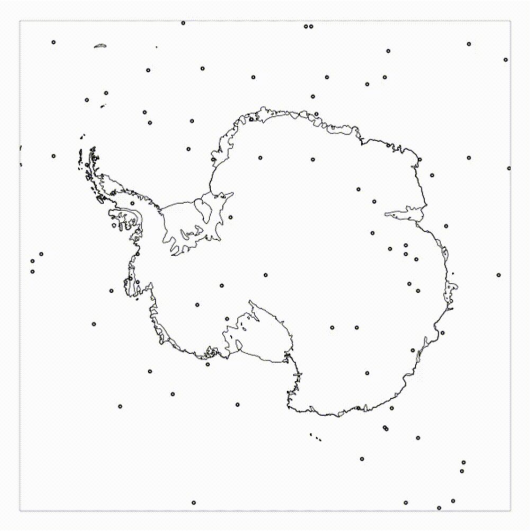
We train a ConvGNP to spatially interpolate ERA5 daily-average 2-metre temperature anomaly



Sensor placement with a ConvGNP: Training

Paper

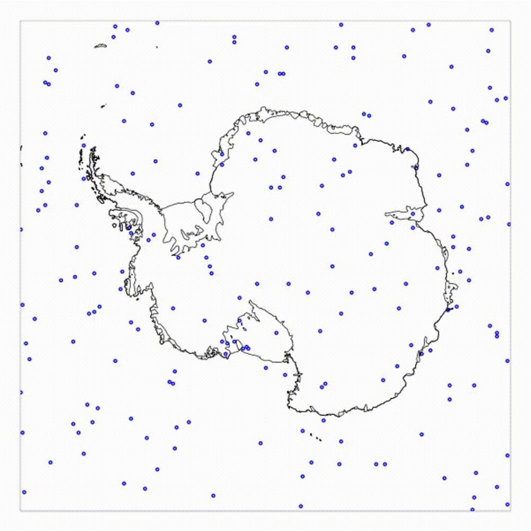
ERA5 2m temperature anomaly **context** points



Train ~4 million parameters to maximise target probability

ConvGNP

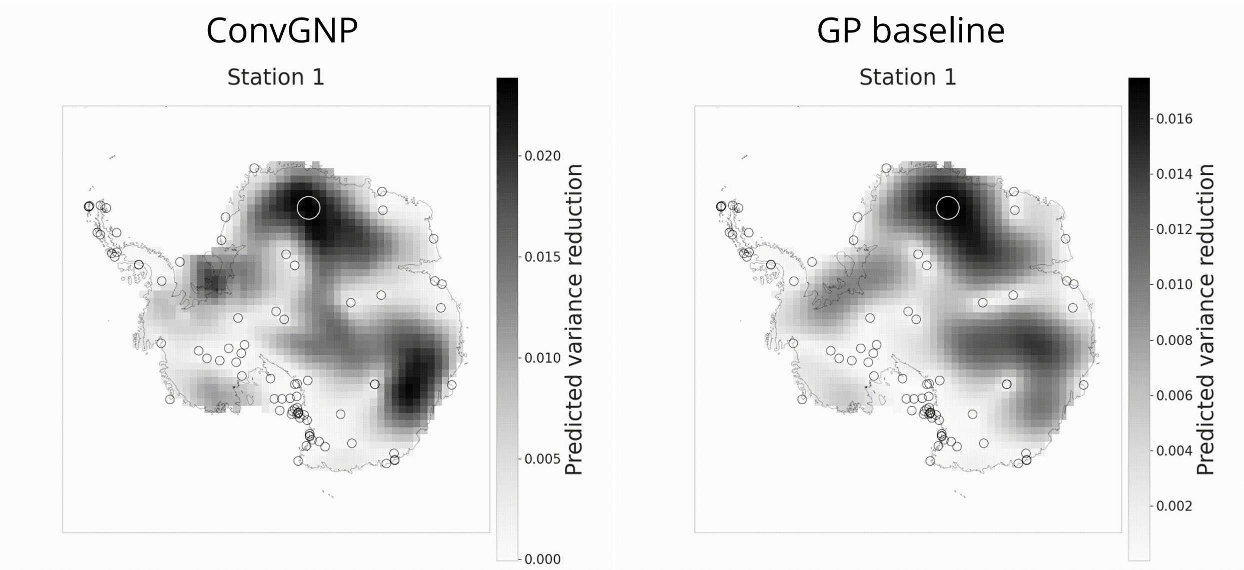
ERA5 2m temperature anomaly **target** points



The ConvGNP finds highly informative sensor placements

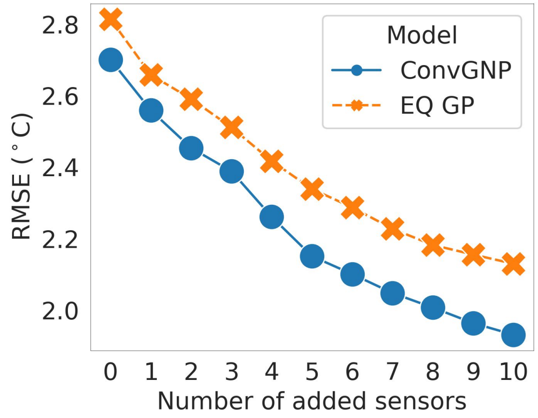
Paper

Sensor placements



Limitation: Model trained on reanalysis, not observations

Results



The ConvGNP:

✅ starts off with better RMSE
✅ reduces its error faster
✅ finds better sensor placements

British Antarctic Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL
The Alan Turing Institute
POLAR SCIENCE FOR PLANET EARTH

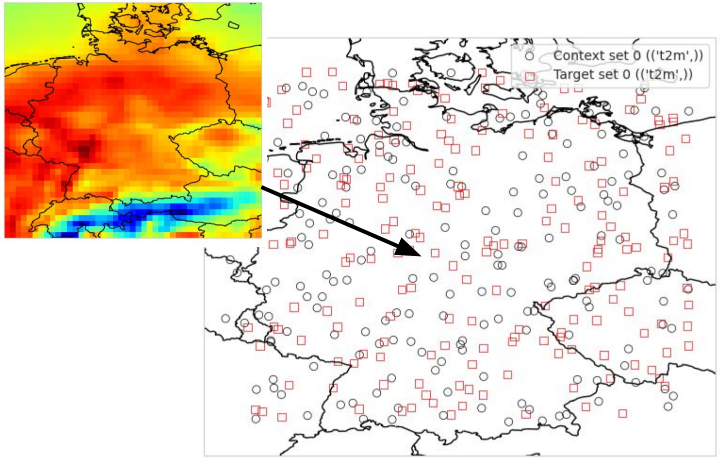
Sim2Real for ConvNP temperature interpolation in Germany

Jonas Scholz

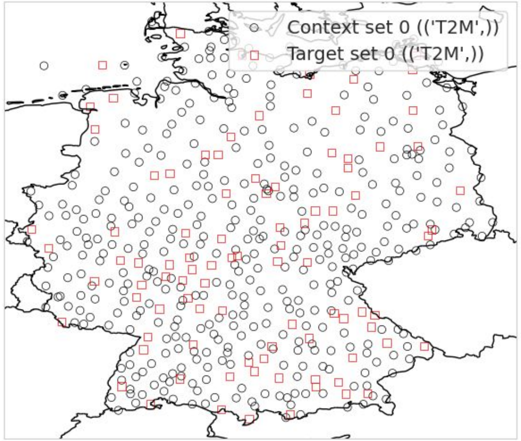
'Sim' phase: ERA5 pre-training

'Real' phase: Station fine-tuning

ERA5 2m temp



Sim2Real

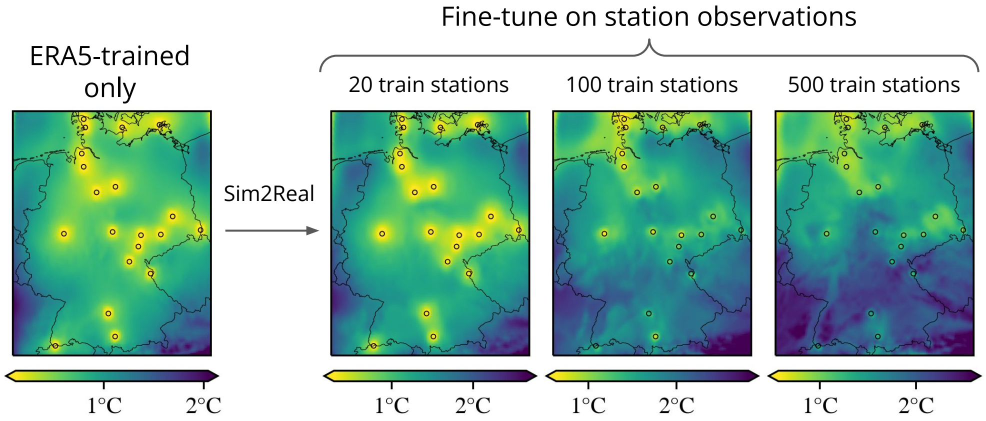


○ = context points

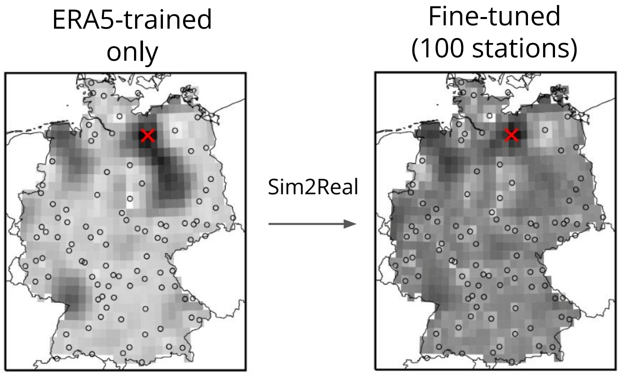
□ = target points

Sim2Real for ConvNP temperature interpolation in Germany

Model uncertainty



Acquisition function



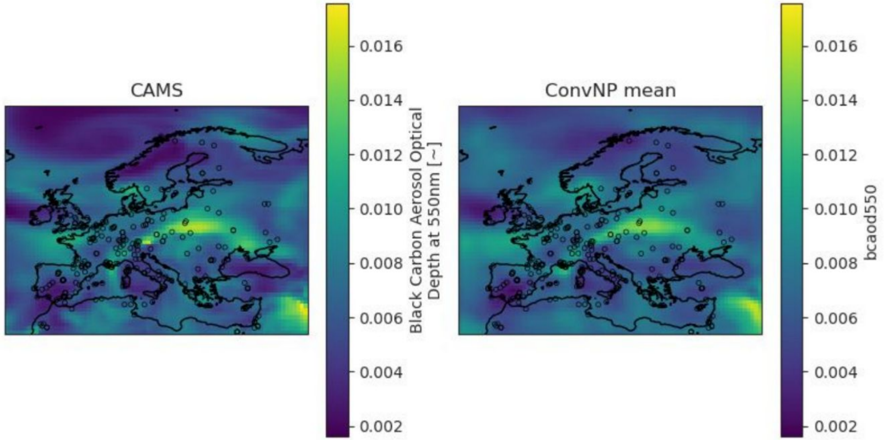
Findings: Pre-training on ERA5 improves performance on observational station data but becomes less important as more observational data becomes available

Aerosol sensor placement

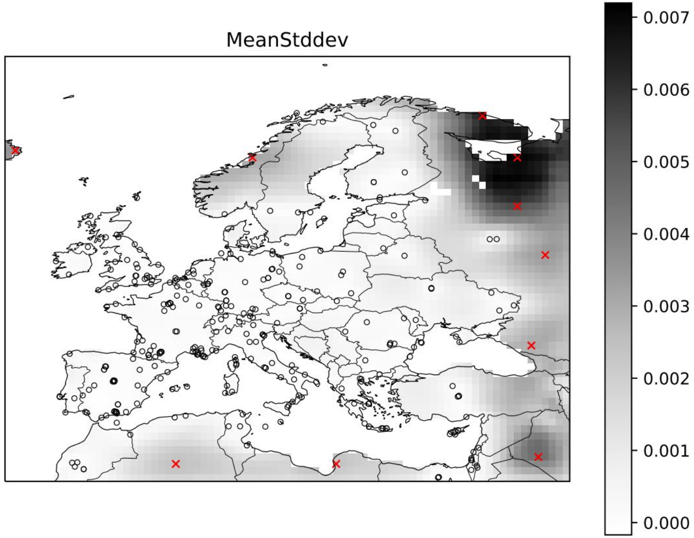

Paolo Pelucchi

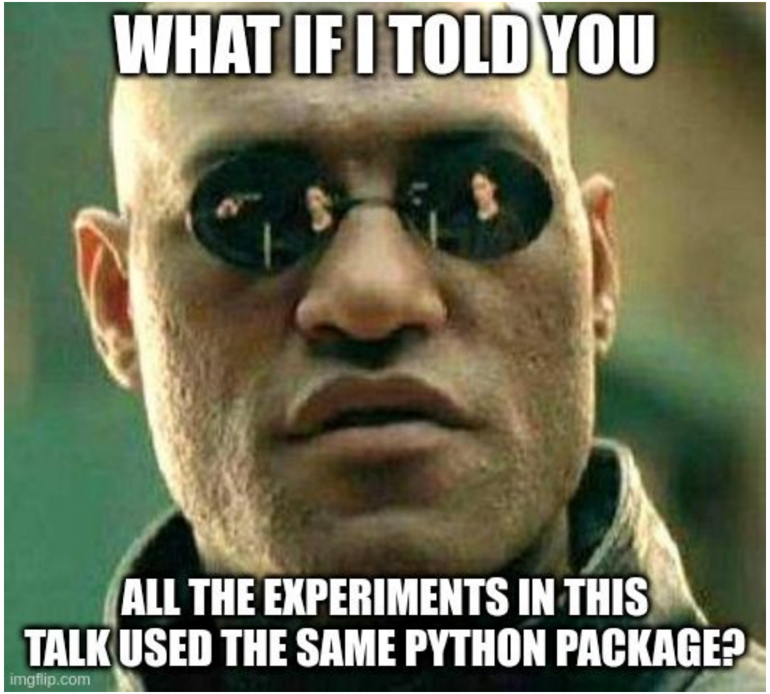
- Target var: black carbon aerosol optical depth (highly uncertain effect on warming)
- Training data: CAMS reanalysis (no fine-tuning yet)
- Sensor network: Aeronet

ConvNP predictions over Europe



Sensor placement acquisition function



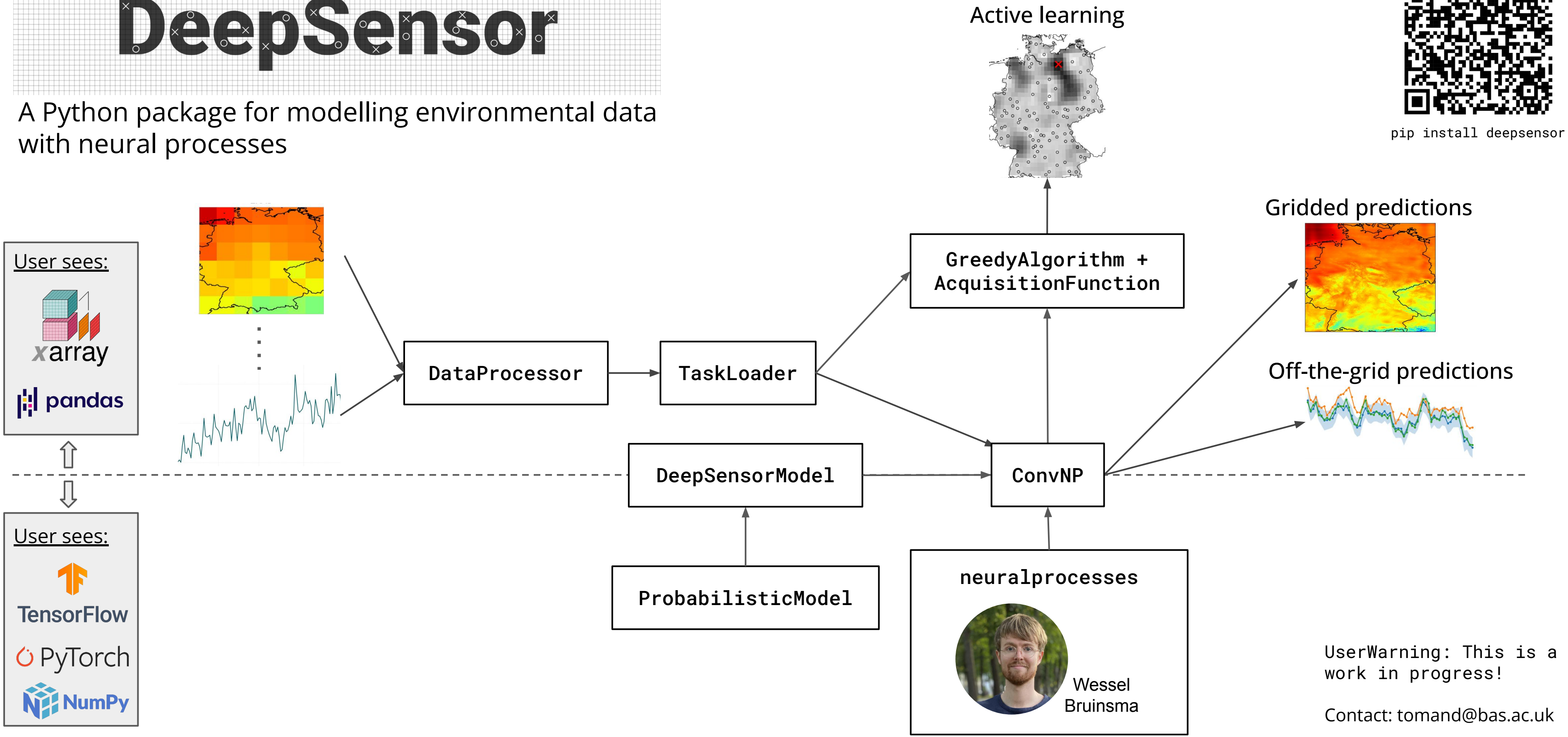

WHAT IF I TOLD YOU
ALL THE EXPERIMENTS IN THIS TALK USED THE SAME PYTHON PACKAGE?
imgflip.com

Talk outline

1) The state of play
2) Modelling environmental observations
3) Convolutional neural processes
4) Data fusion experiments
5) **DeepSensor**
6) Closing thoughts

DeepSensor

A Python package for modelling environmental data with neural processes

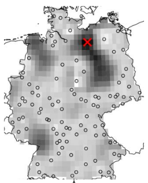


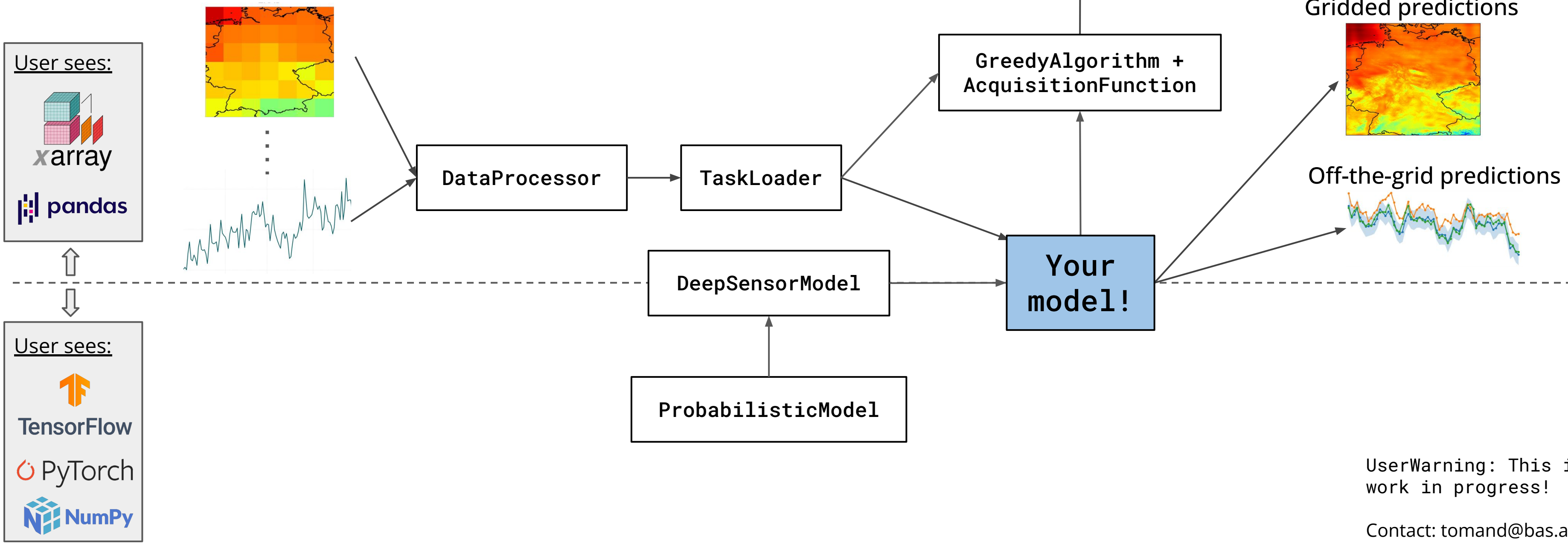
DeepSensor

A Python package for modelling environmental data with neural processes

`pip install deepsensor`

Active learning





UserWarning: This is a work in progress!

Contact: tomand@bas.ac.uk

British Antarctic Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

The Alan Turing Institute

POLAR SCIENCE FOR PLANET EARTH

DeepSensor

A Python package for modelling environmental data with neural processes

Hopes for DeepSensor:

- Lower the barrier to entry for both environmental scientists and machine learners
- Galvanise research progress by building an open-source software community
  - Positive feedback loop between research and software
- Blue sky thinking: Become leading software for the latest environmental ML paradigms

DeepSensor testimonial!

patel-zeel commented 8 hours ago | Author

"`DeepSensor` has an easy-to-use interface similar to `sklearn` and its seamless integration with `xarray` saves a lot of time and energy to preprocess and post-process the data."

Talk outline

1) The state of play
2) Modelling environmental observations
3) Convolutional neural processes
4) Data fusion experiments
5) DeepSensor
6) **Closing thoughts**

Closing thoughts

- **Conclusions:**
  - Entering a new era of environmental ML with versatile modelling capabilities
  - ConvNPs are one such model, and can tackle a range of environmental prediction tasks:
    - Sensor placement, downscaling, data fusion, forecasting, filling missing data
  - DeepSensor: a package for the next generation of environmental ML models(?)
  - ConvNPs are not a panacea:
    - Active research area: limitations and improvements will be found
    - Other novel architectures like GNNs or transformers have lots of potential too

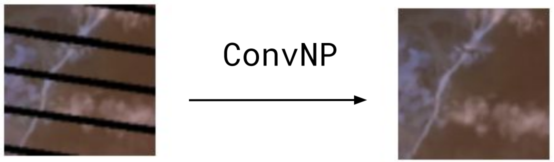
Closing thoughts

- **Challenges**:
  - ConvNPs must *learn how to condition on data* — data hungry
  - Chicken-and-egg problem for sensor placement
    - Pre-train on simulation data
  - Scaling to high dimensions (e.g. depth/altitude, 100s/1000s of variables...)

Closing thoughts

- **Challenges**:
  - ConvNPs must *learn how to condition on data* — data hungry
  - Chicken-and-egg problem for sensor placement
    - Pre-train on simulation data
  - Scaling to high dimensions (e.g. depth/altitude, 100s/1000s of variables...)

- **Future work**:
  - Transfer learning from densely- to sparsely- monitored regions
  - Infilling missing satellite data
  - Sensor trajectories for Digital Twinning



Pondaven et al., 2022

Neural process timeline

2018: M. Garnelo et al., "Conditional Neural Processes." *ICML*

2020: J. Gordon et al., "Convolutional Conditional Neural Processes." *ICLR*

2021: W.P. Bruinsma et al., "The Gaussian Neural Process." *AABI*

2022: S. Markou et al., "Practical Conditional Neural Processes via Tractable Dependent Predictions." *ICLR*

2022: A. Vaughan et al., "Convolutional Conditional Neural Processes for Local Climate Downscaling." *GMD*

2023: W.P. Bruinsma et al., "Autoregressive Conditional Neural Processes." *ICLR*

2023: T.R. Andersson et al., "Environmental Sensor Placement with Convolutional Gaussian Neural Processes." *EDS*

Sensor placement paper

Thanks for listening!

tomand@bas.ac.uk

tom_r_andersson

DeepSensor

British Antarctic Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

UNIVERSITY OF CAMBRIDGE

Microsoft

The Alan Turing Institute

British Antarctic Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

The Alan Turing Institute