Artificial intelligence reconstructs global temperature from scarce local data

In Review: Communications Earth & Environment

UNIVERSITÄT BERN

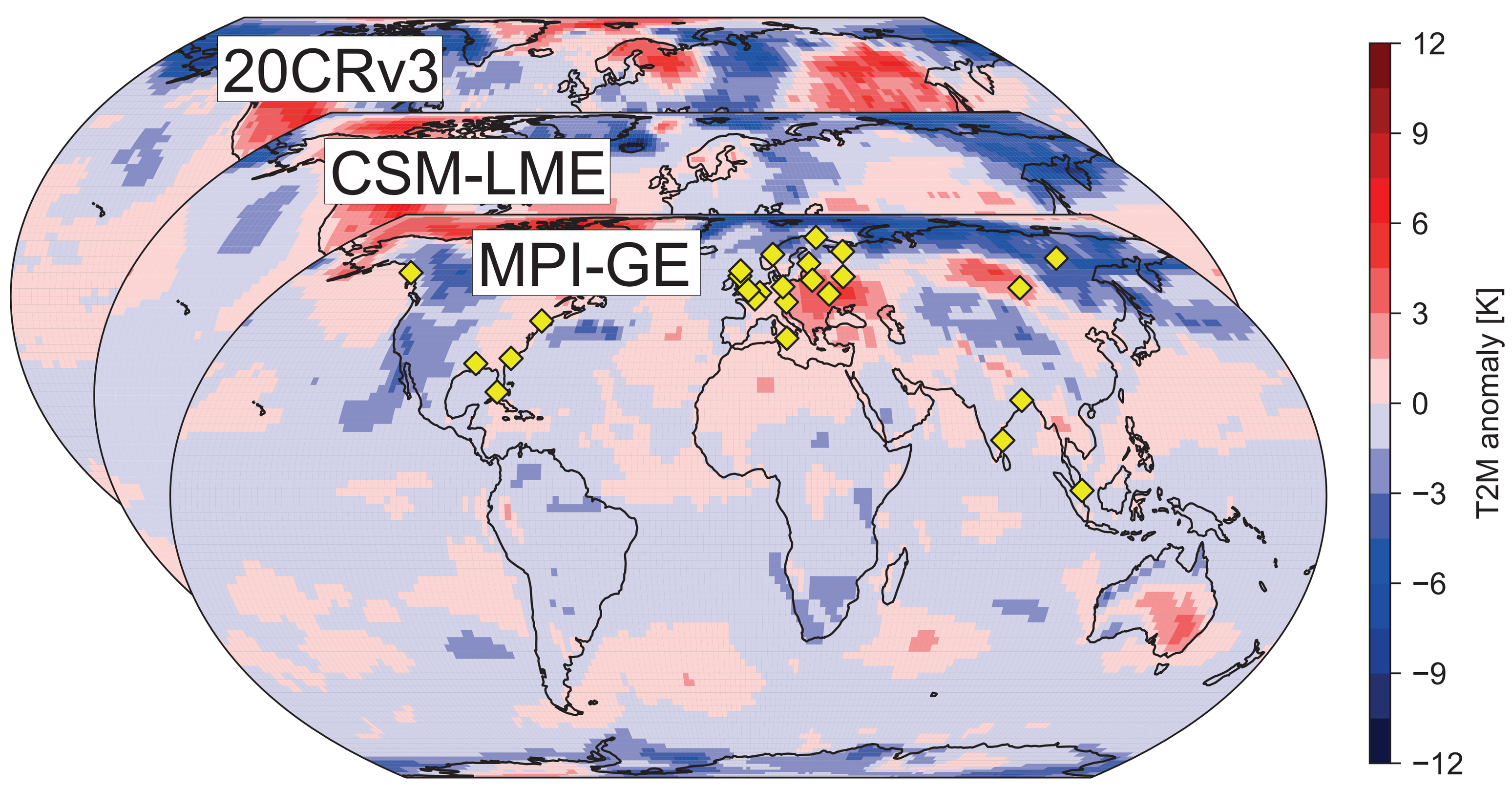
OESCHGER CENTRE CLIMATE CHANGE RESEARCH

Martin Wegmann[1,2] (martinwegmann@pm.me & @martin_wegmann) **& Fernando Jaume Santero**[3,4]

1 Institute of Geography, University of Bern, 2 Oeschger Centre for Climate Change Research, University of Bern
3 Department of Radiology and Medical Informatics, University of Geneva, 4 Haute École de Gestion de Genève, Haute École Spécialisée de Suisse Occidentale

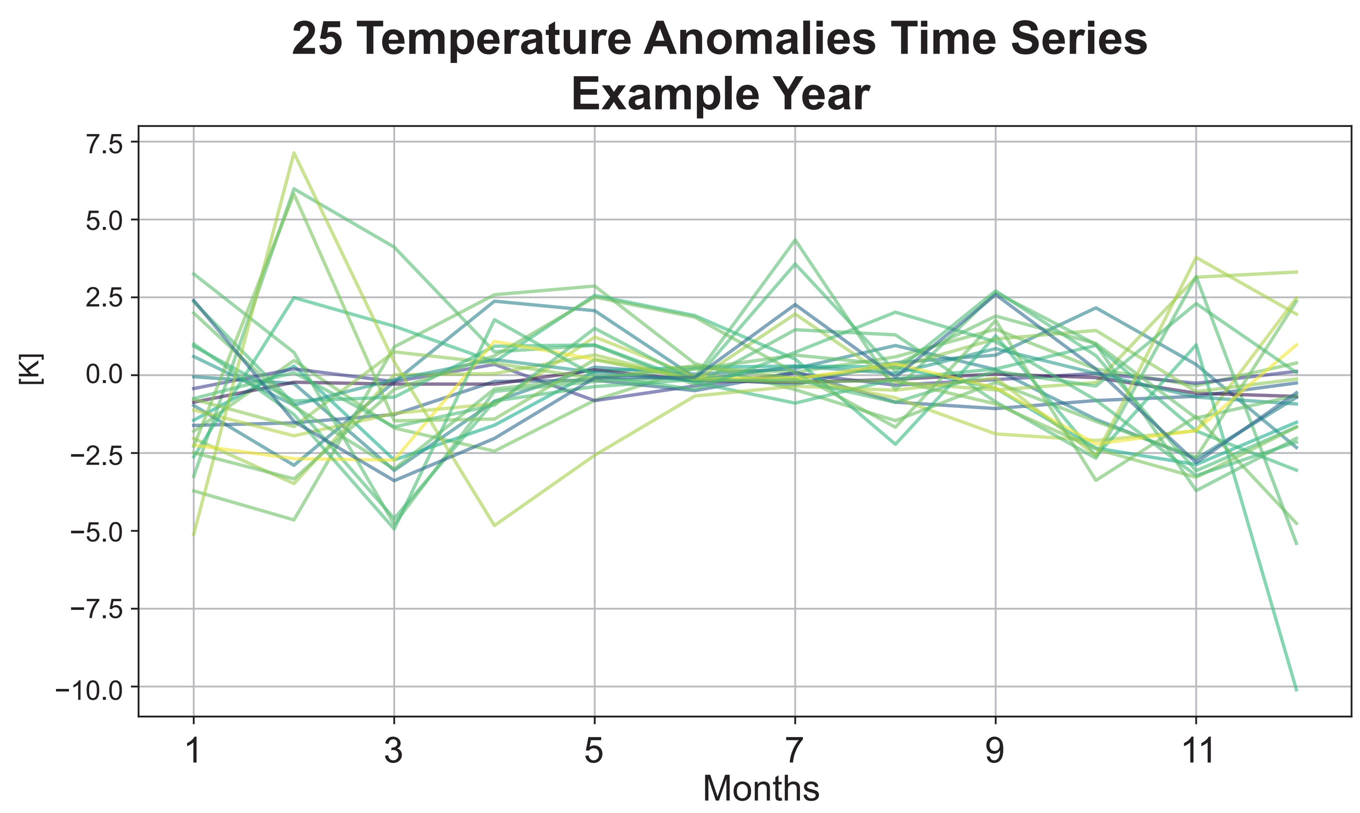
Goals

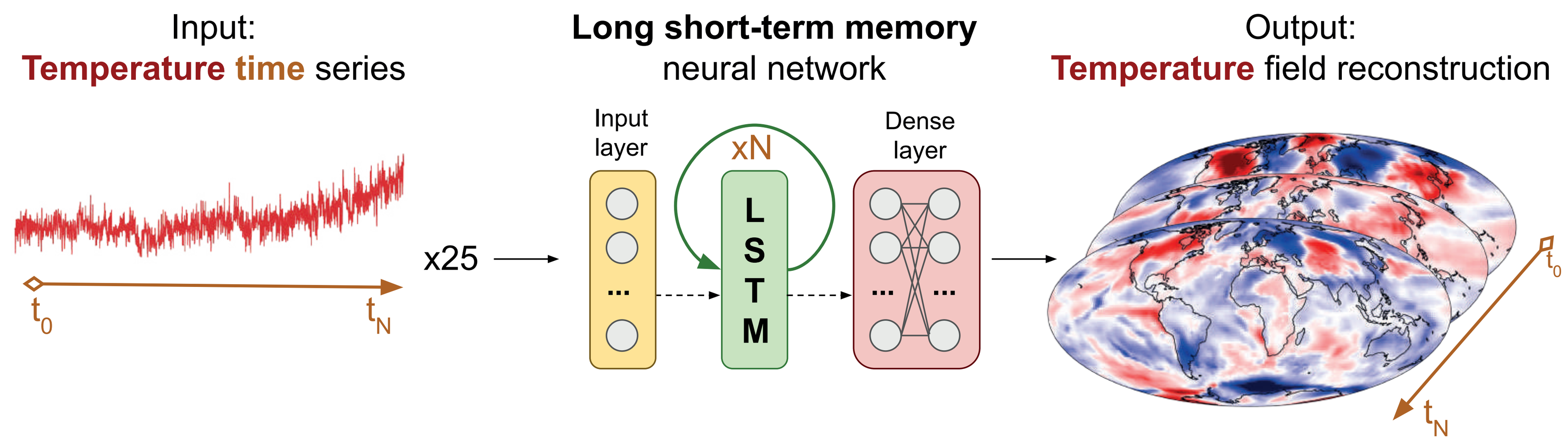
- Understanding **monthly-to-annual climate variability** is essential for adapting to future climate extremes.
- We need **cheap, performant reconstructions** of global, gridded, monthly temperature anomalies given only **few local data.**
- As proof-of-concept, we **want to reconstruct 4,824 months** (1602–2003) of 2m mean temperature anomalies to compare with independent data and to highlight a **low cost, high flexibility approach.**



Procedure

- As training data set we use monthly 2m temperature anomalies with respect to the period 1951-1980 CE from **three different gridded products**, one reanalysis (20CRv3), two coupled climate models (CSM-LME & MPI-GE).
- We extract the nearest neighbour information from realistic **25 locations**, displayed as yellow diamonds. Those are our "**pseudo local data**" which are located roughly where we can expect the longest temperature time series in reality.
- 25 monthly 2m temperature anomaly times series are then used as input to reconstruct a global field time series via a **simple long short-term memory neural network.**
- We utilized a **small LSTM** with an output dimensionality of 50 units and the hyperbolic tangent as activation function, followed by a dense layer of 18,432 parameters that were reshaped into a grid of 96 × 192 temperature points. The LSTM was trained with **three layers (i.e., latitude, longitude, and 2m temperature anomaly)**. Training data size is N=1980 and for comparison N=20000.

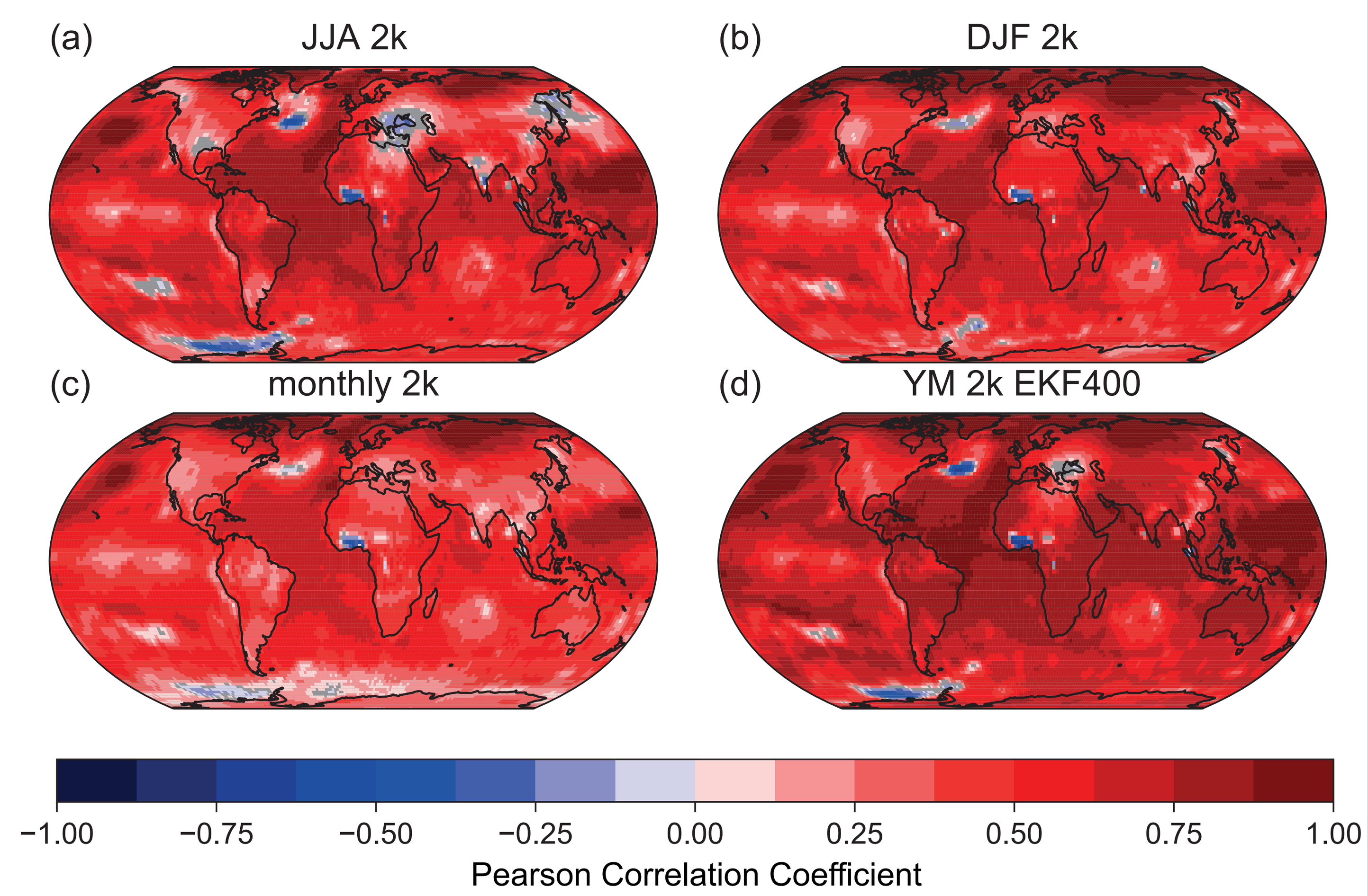




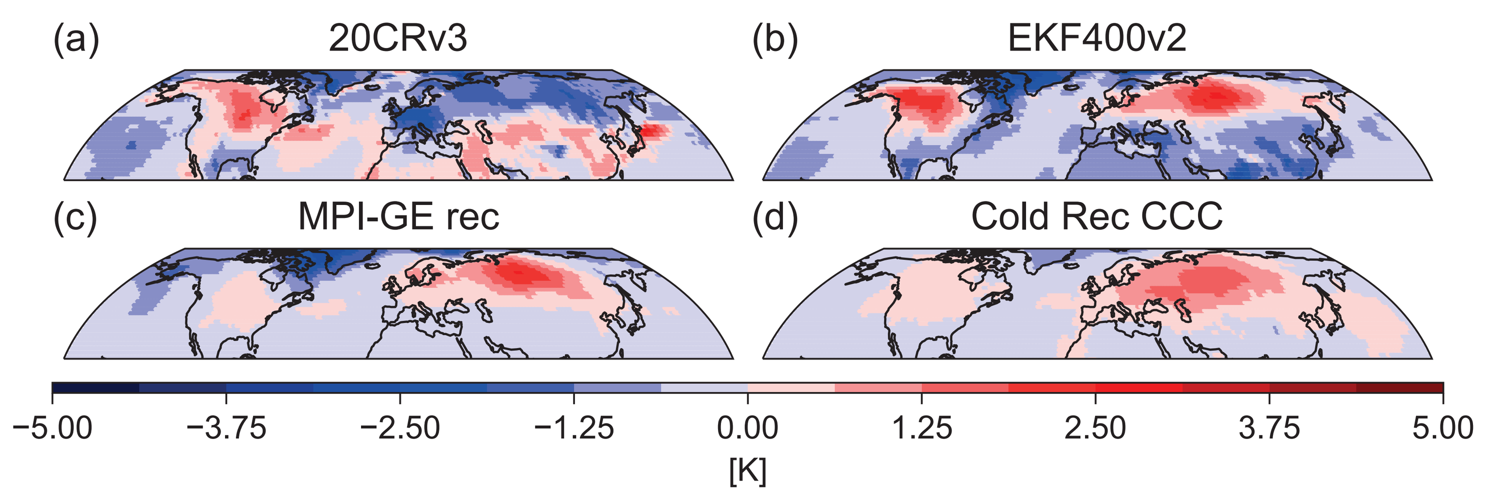
Concept of the reconstruction process. (Upper) Example of gridded data resolution and location of pseudo-station data. (Middle) example of one year of pseudo-station temperature anomaly time series. (Lower) Schematic of training flow for the LSTM.

Key Results

- **Correlation between independent data and LSTM are very high,** even when training sample size is small (N=1980). Highest correlations close to local data, but **teleconnections are captured.**
- Seasonal and regional differences make sense and highlight challenges intrinsic to the climate system.
- **Some features can be artifacts,** such as negative correlation over equatorial Western Africa. It is specific to MPI-GE as a training data set and do not occur for 20CRv3 nor CESM-LME.
- Highlighting the boreal cold season of 1834/1835 shows the method`s **capability to reproduce** not only mean statistics, but also **specific climate events.**



Maps of Pearson Correlation Coefficients between EKF400v2 ensemble mean and the MPI-GE based (N=1980 training months) LSTM temperature anomaly reconstruction for the period 1602–2003 CE. a) Correlation Coefficients for boreal summer (JJA) seasons (N=402). b) Correlation Coefficients for boreal winter (DJF) seasons (N=401). c) Correlation Coefficients for all months (N=4824). d) Correlation Coefficients for yearly means (N=402).



Cold season (ONDJFMAM) 2m temperature anomalies for 1834/35 CE. c) is our LSTM reconstruction, a,b,d) are independent reconstructions that take into account much more input data.

Take-home message

- We focus here on a **grid reconstruction problem**, a typical issue in the (paleo-) climate community. By using a very conservative approach with small sample sizes (especially for deep learning), we could produce a **realistic, robust global temperature reconstruction**.
- Our approach allows for the reconstruction of **different climate variables from (paleo) climate archives such as tree rings, coral and ice cores.** Needless to say, reconstructions do not need to be global, but **can focus on a region of interest.**
- By utilising existing data, storage and **energy consumption can be kept at a minimum,** while at the same time contributing to the United Nations Sustainable Development Goals. The RNN reconstructions **can be created in less than an hour on a average-priced laptop, operating solely with open-access, open-source software and data.**