Reconstructing historical daily weather fields using a deep-learning variational auto-encoder

Conall E. Ruth[1,2], Yannis Schmutz[3], Stefan Brönnimann[1,2]

[1]Oeschger Centre for Climate Change Research, University of Bern, [2]Institute of Geography, University of Bern, [3]Applied Machine Intelligence, Bern University of Applied Sciences

conall.ruth@unibe.ch

1 Motivation

- Daily gridded weather reconstructions are crucial for studying historical extreme events and their impacts as well as the long-term evolution of weather variability.
- Traditional spatial-interpolation techniques struggle with historically sparse data availabilities (Fig. 1) while re-analysis NWP models have high computational costs.
- Deep learning models already represent a promising alternative, therefore investigating the use of various possible architectures continues to be valuable.

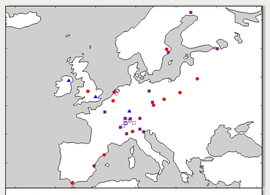
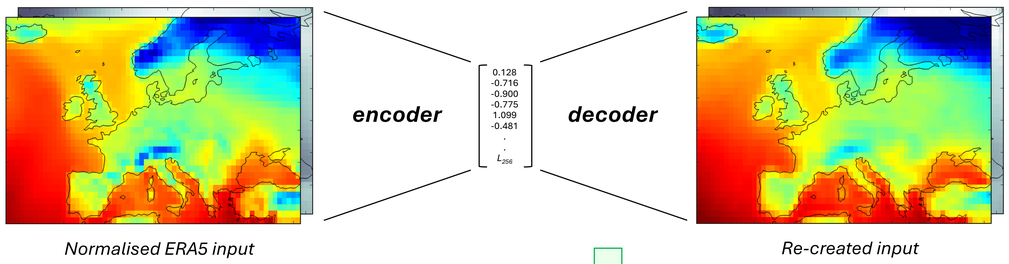


Fig. 1: Study domain showing the station locations of 1807 (red: temperature, blue: pressure, purple: both; filled: model input, unfilled: validation)

2 Study overview

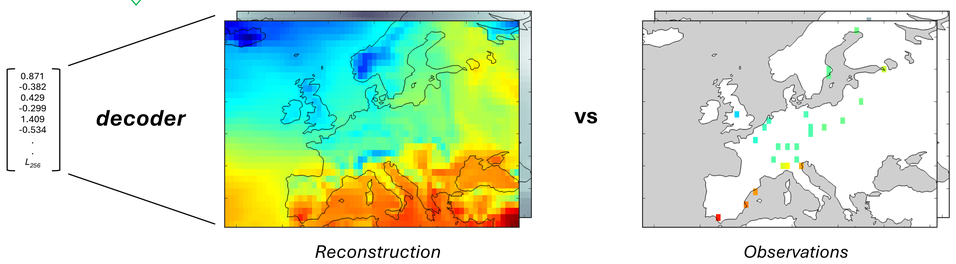
- Here, we present a deep-learning model based on a variational auto-encoder (VAE) architecture for reconstructing fields of mean daily air temperature and sea-level pressure at a 1° resolution across Europe (36N-67N, 22W-41E).
- Using this, we reconstruct the year 1807 from 25 historical temperature series and 18 historical pressure series (Fig. 1) and evaluate against a set of hold-out records (Tab. 1, 2). We also reconstruct the period 1950-1954 assuming the same data availability as 1807 and evaluate against ERA5[1] (Fig. 2-4).
- The model is also compared to the re-analysis dataset *20CRv3*[2] and to an existing deep-learning model *WeRec3D*[3] inspired by video inpainting, which uses the same input series.

3 VAE model concept



Model set-up: A pair of convolutional neural networks – an *encoder* and a *decoder* – are trained to respectively reduce the normalised ERA5 fields of temperature and pressure for a given day to a single set of 256 *latent variables* and then upsample this to re-create the input fields.

Model application: Once trained, we can insert a random set of latent variables and decode this to generate a new but plausible set of output fields. The latent variables can then be iteratively adjusted such that the output best matches any observations from a given day of interest, thus providing the reconstruction.



4 Reconstruction of 1807 vs the hold-out series

RMSEs and correlations

	VAE	*WeRec 3D*	*20CRv3*
Gdansk[4]	1.81 K (0.98)	*1.80 K (0.98)*	*2.73 K (0.96)*
Zurich	5.22 K (0.98)	*5.02 K (0.98)*	*7.05 K (0.94)*
Aarau	5.41 K (0.97)	*6.28 K (0.98)*	*7.42 K (0.94)*
Vevey	6.75 K (0.96)	*6.07 K (0.96)*	*10.24 K (0.91)*
Marschlins	3.52 K (0.96)	*5.13 K (0.95)*	*5.54 K (0.93)*
Delemont	3.32 K (0.95)	*3.31 K (0.97)*	*5.21 K (0.90)*
Bern (Studer)	3.30 K (0.98)	*3.42 K (0.99)*	*5.34 K (0.92)*
Delfft	2.19 K (0.98)	*3.04 K (0.98)*	*2.36 K (0.94)*
Zwanenburg	0.95 K (0.99)	*2.31 K (>0.99)*	*2.41 K (0.95)*
mean	**3.61 K (0.97)**	***4.04 K (0.98)***	***5.37 K (0.93)***

Table 1: Root mean squared errors and correlation coefficients (in parentheses) of each model with respect to the hold-out **temperature** series

	VAE	*WeRec 3D*	*20CRv3*
Gdansk[4]	4.31 hPa (0.88)	*4.16 hPa (0.88)*	*4.17 hPa (0.89)*
Zurich	5.11 hPa (0.95)	*5.11 hPa (0.92)*	*4.64 hPa (0.93)*
Aarau	3.66 hPa (0.92)	*4.04 hPa (0.91)*	*4.36 hPa (0.91)*
Vevey	3.70 hPa (0.94)	*4.06 hPa (0.93)*	*4.98 hPa (0.92)*
Marschlins	3.78 hPa (0.95)	*4.77 hPa (0.91)*	*4.33 hPa (0.88)*
Delemont	11.05 hPa (0.89)	*11.75 hPa (0.88)*	*11.81 hPa (0.87)*
Bern (Studer)	4.31 hPa (0.97)	*3.95 hPa (0.96)*	*4.73 hPa (0.93)*
Bern (Fueter)	4.99 hPa (0.85)	*5.73 hPa (0.83)*	*5.55 hPa (0.82)*
mean	**5.11 hPa (0.92)**	***5.45 hPa (0.90)***	***5.57 hPa (0.89)***

Table 2: Root mean squared errors and correlation coefficients (in parentheses) of each model with respect to the hold-out **pressure** series

The VAE model reconstructs the hold-out series of 1807 relatively well with RMSEs and correlations similar to or better than those of *WeRec3D* and *20CRv3*. The standard deviations of the series are also well represented (not shown here).

5 Reconstruction of 1950-1954 (assuming the data availability of 1807) vs ERA5

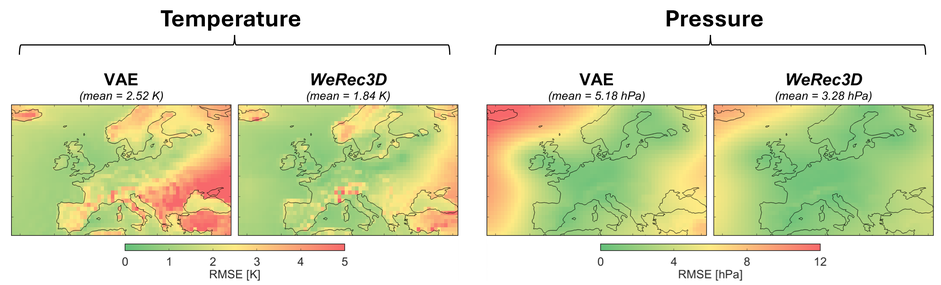


Fig. 2: **Root mean squared errors** of the (absolute) VAE and *WeRec3D* reconstructions with respect to ERA5.

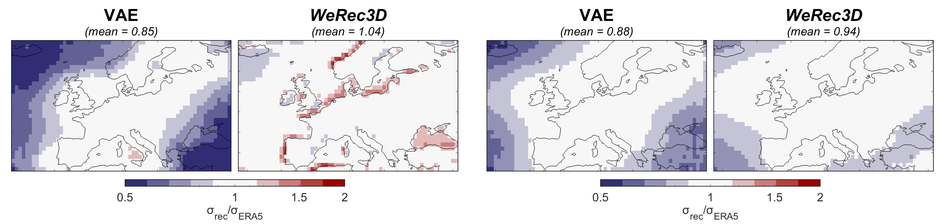


Fig. 3: **Standard deviation ratios** of the normalised VAE and *WeRec3D* reconstructions with respect to ERA5. Values <1 (blue) indicate that the standard deviation of the reconstruction is too low; values >1 (red) indicate it is too high.

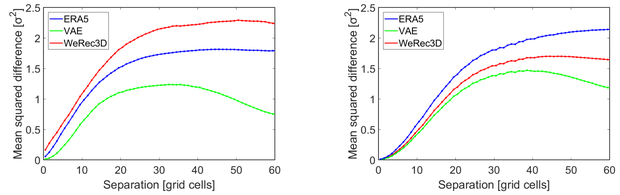


Fig. 4: **Variograms** (illustrating the mean squared difference between any two grid cells as a function of the distance between them) of the normalised VAE, *WeRec3D* and ERA5 fields. Differences smaller than those expected from ERA5 indicate that the reconstructed fields are too smooth; greater differences indicate that the reconstructed fields are too rugged.

Considering the entire study domain over the period 1950-1954, the VAE model has a relatively high reconstruction error with mean RMSEs of 2.52 K and 5.18 hPa (compared to 1.84 K and 3.28 hPa for *WeRec3D*; Fig. 2). It also under-represents the true (ERA5) amount of weather variability (given by values <1 in Fig. 3), while *WeRec3D* captures this relatively well. These issues are especially pronounced towards the northwest and southeast of the domain where observations are minimal.

Furthermore, the variograms (Fig. 4) indicate that the reconstructed VAE fields of both variables are considerably too smooth relative to ERA5. Again, *WeRec3D* is closer to reality with temperature and pressure fields that are only somewhat too rugged and smooth, respectively.

6 Conclusions

The VAE model reconstructs the hold-out series of 1807 relatively well compared to *WeRec3D* and *20CRv3*, successfully demonstrating the application of this deep-learning architecture for the task of historical weather reconstruction.

Overall, however, the VAE does not outperform *WeRec3D* when considering the entire spatial domain, with the latter providing more realistic and accurate reconstructions, particularly in the most sparely observed regions.

[1] Hersbach, H. et al. (2023) 'ERA5 hourly data on single levels from 1940 to present', *Copernicus Climate Change Service (CS3) Climate Data Store (CDS)*, doi: https://doi.org/10.24381/cds.adbb2d47 (accessed 20 December 2024)

[2] Slivinski, L.C. et al. (2019) 'Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system', *Quarterly Journal of the Royal Meteorological Society* 145 (724): 2876-908, doi: https://doi.org/10.1002/qj.3598

[3] Schmutz, Y., N. Imfeld, S. Brönnimann, and E. Graf (2024) 'Enhanced video inpainting: A deep learning approach for historical weather reconstruction', *Journal of Geophysical Research: Machine Learning and Computation* 1 (e2024JH000299), doi: https://doi.org/10.1029/2024JH000299

[4] Stucki, P. et al. (in preparation) 'Assessment of reanalyses-based and machine-learning datasets for hindcasting the European hot summer of 1807', *peter.stucki@unibe.ch*