



# Artificial intelligence reconstructs missing climate information

Christopher Kadow<sup>1,2</sup>✉, David Matthew Hall<sup>3</sup> and Uwe Ulbrich<sup>1,2</sup>

**Historical temperature measurements are the basis of global climate datasets like HadCRUT4. This dataset contains many missing values, particularly for periods before the mid-twentieth century, although recent years are also incomplete. Here we demonstrate that artificial intelligence can skilfully fill these observational gaps when combined with numerical climate model data. We show that recently developed image inpainting techniques perform accurate monthly reconstructions via transfer learning using either 20CR (Twentieth-Century Reanalysis) or the CMIP5 (Coupled Model Intercomparison Project Phase 5) experiments. The resulting global annual mean temperature time series exhibit high Pearson correlation coefficients ( $\geq 0.9941$ ) and low root mean squared errors ( $\leq 0.0547$  °C) as compared with the original data. These techniques also provide advantages relative to state-of-the-art kriging interpolation and principal component analysis-based infilling. When applied to HadCRUT4, our method restores a missing spatial pattern of the documented El Niño from July 1877. With respect to the global mean temperature time series, a HadCRUT4 reconstruction by our method points to a cooler nineteenth century, a less apparent hiatus in the twenty-first century, an even warmer 2016 being the warmest year on record and a stronger global trend between 1850 and 2018 relative to previous estimates. We propose image inpainting as an approach to reconstruct missing climate information and thereby reduce uncertainties and biases in climate records.**

The deeper a research period lies in the past, the fewer observations are available. The atmospheric variable with the longest measurement record is the near-surface air temperature over land, usually at a height of 2 m. Around the globe, individual locations have temperature records from as early as the late seventeenth century<sup>1</sup> (for example, Zurich, Prague and Berlin). These records are extremely valuable for science<sup>2</sup>, but these locations are too sparse to derive global or even regional statements. Station-combining datasets, which include measurements of sea surface temperature by ships, start in the mid-nineteenth century. The Fifth Assessment Report (AR5) of the United Nations Intergovernmental Panel on Climate Change (IPCC) Chapter 2<sup>3</sup> investigates three of those 'Global Combined Land and Sea Surface Temperature' datasets: HadCRUT4<sup>4</sup>, MLOST (Merged Land–Ocean Surface Temperature)<sup>5</sup> and GISS (Goddard Institute for Space Studies)<sup>6</sup>. The latter two use interpolation or smoothing techniques to reach a nearly global scale, starting in 1880. HadCRUT4 has more stringent conditions, using grid points on a  $5 \times 5^\circ$  grid only if measurements exist for the monthly analysis starting in 1850. No interpolation in grid space is applied. Therefore, the HadCRUT4 analysis shows many missing values in its early phase. However, even recent years have missing grid points due to incomplete global coverage of the measurement stations, buoys, ships and so on. Missing values contribute to uncertainties and structural biases when researching our climate system<sup>4,7</sup>. The climate science community has spent a great deal of effort to estimate the missing information by developing smoothing and interpolation techniques<sup>5,6,8</sup>, such as kriging, in the more recent Berkley Earth<sup>9</sup> and Cowtan and Way<sup>7</sup> temperature reconstructions. Additionally, principle component analysis (PCA) has had a huge impact on infilling missing values and detecting signals in weather and climate datasets over the past few decades<sup>10–13</sup>. The ability of this technique to estimate real values from empirical orthogonal functions has led to successful reconstructions. In this context, even

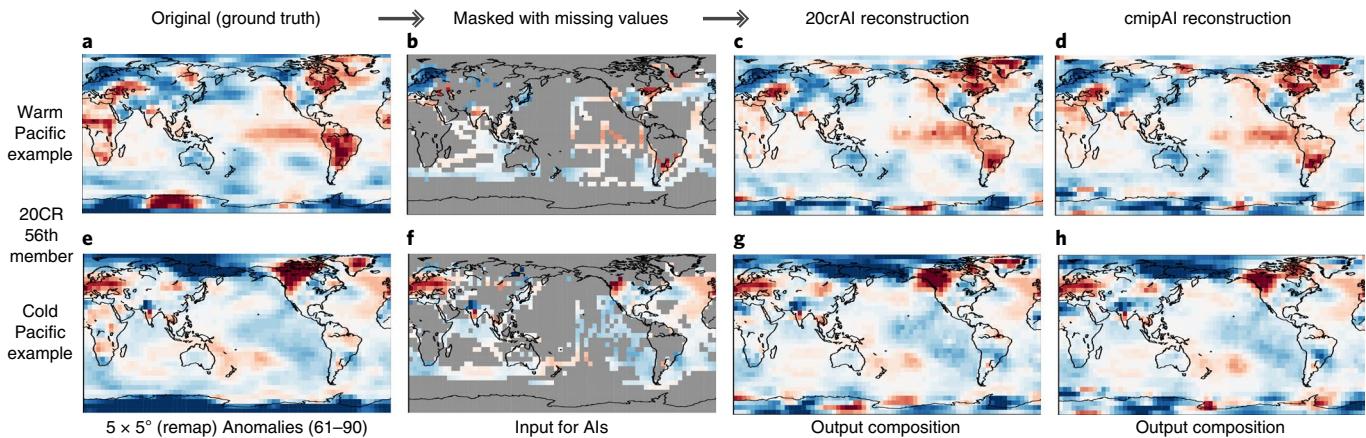
training based on the output of numerical models (NMs) has been performed on, for example, regional sea surface temperature time series in the Baltic Sea<sup>14</sup>.

The rapid progress in artificial intelligence (AI) research has substantially impacted many scientific fields, which includes climate science<sup>15,16</sup>. Examples include deep learning for the recognition of forced climate patterns<sup>17</sup> or extreme events<sup>18</sup>, ensemble learning using bootstrap aggregation to improve decadal climate predictions<sup>19</sup> and estimating ocean heat content from tidal magnetic satellite observations using neural networks<sup>20</sup>, to name a few. Recently, major progress has been made in image inpainting, the process of reconstructing missing parts of images, which has been applied primarily to photographs and paintings<sup>21</sup>, but also to, for example, satellite images of sea surface temperatures<sup>22,23</sup>. In particular, inpainting for irregular holes using stacked partial convolutions and an automatic mask updating mechanism has improved the accuracy of image reconstructions drastically<sup>24</sup>. This partial convolution method outperformed several image inpainting technologies<sup>25–28</sup> in all three major categories<sup>29</sup> of sequential-based, convolutional neural network (CNN)-based and generative adversarial network (GAN)-based methods. In addition to producing results of superior visual quality, this technique was the first to achieve high quality result when applied to images with irregularly shaped holes, which is essential for reconstructing missing values in climate datasets.

## Combining AI with Earth system modelling and observations

We propose to develop this state-of-the-art deep learning based inpainting technology<sup>24</sup> further to reconstruct missing climate information of global temperature patterns, especially for early periods with only very sparse available data records. The definition of an irregular hole in image inpainting is equivalent to the set of missing values found at various locations in a climate dataset.

<sup>1</sup>German Climate Computing Center (DKRZ), Hamburg, Germany. <sup>2</sup>Institute of Meteorology, Freie Universität Berlin, Berlin, Germany. <sup>3</sup>NVIDIA, Santa Clara, CA, USA. ✉e-mail: [kadow@dkrz.de](mailto:kadow@dkrz.de)



**Fig. 1 | AI models reconstruct two exemplary monthly show cases with many missing values.** **a–h,** Warm (September 1877) (**a–d**) and cold (August 1893) (**e–h**) eastern Pacific examples explain the reconstruction pathway of the held-out 56th member from 20CR. Shown are temperature anomalies in degrees centigrade with respect to the 1961–1990 climatology. The ground-truth original data (column 1), masked datasets with missing values (grey) of HadCRUT4 (column 2), output by 20crAI (column 3) and output by cmipAI (column 4). Missing value masks are taken from the matching time step in HadCRUT4. More examples in the Supplementary Information.

Therefore, we can use the missing value mask provided by HadCRUT4. This monthly analysis spans the years 1850–2018 and contains 2,028 masks, which may be used to train the deep neural network (DNN). In image inpainting, a hole is typically replaced with a pattern matching a particular missing object (for example, a nose or car headlight) and by a linear interpolation of the surrounding colour pattern (for example, water or sand) learned from examples to make the image as realistic as possible<sup>24</sup>. In a climate dataset, the relationship between neighbouring grid points is dictated by the laws of physics, which include conservation of mass, momentum and energy. This leads to specific variability patterns, which can help to estimate a realistic full pattern based on sparse available data.

Hence, we trained the DNN using a large set of example atmospheric states obtained from reanalysis and climate experiments (Extended Data Fig. 1) by NMs, which can be atmosphere only, coupled atmosphere–ocean climate models or Earth system models. NMs are derived by solving the primitive equations of motion. A DNN learns to reproduce the statistical relationships present in the NM data by iteratively adjusting its parameters until it can accurately predict the training examples.

We used two distinct datasets for training. One includes the Twentieth-Century Reanalysis<sup>30</sup> (20CR) of the National Oceanic and Atmospheric Administration, used to train the 20crAI model. The other is the historical experiment of the Coupled Model Intercomparison Project Phase 5 (CMIP5<sup>31</sup>), used to train the cmipAI model. 20CR is based on the available surface observations of synoptic pressure, monthly sea surface temperature and sea ice distribution, which have been assimilated into the model. It uses only the atmospheric component of an Earth system model. In more detail, 20CR consists of 56 ensemble members, differing between each other, as the assimilated data do not fully prescribe the global atmosphere. The CMIP5 historical experiment, in contrast, consists of a number of model runs carried out with different numerical codes. In addition, most of these models consist of different number of ensemble realizations, which lead to 145 experiment NM runs (Extended Data Fig. 10). These runs are driven by boundary conditions, such as volcanic eruptions, greenhouse gas concentrations and so on, rather than assimilated data, but their atmosphere and ocean submodels are fully coupled.

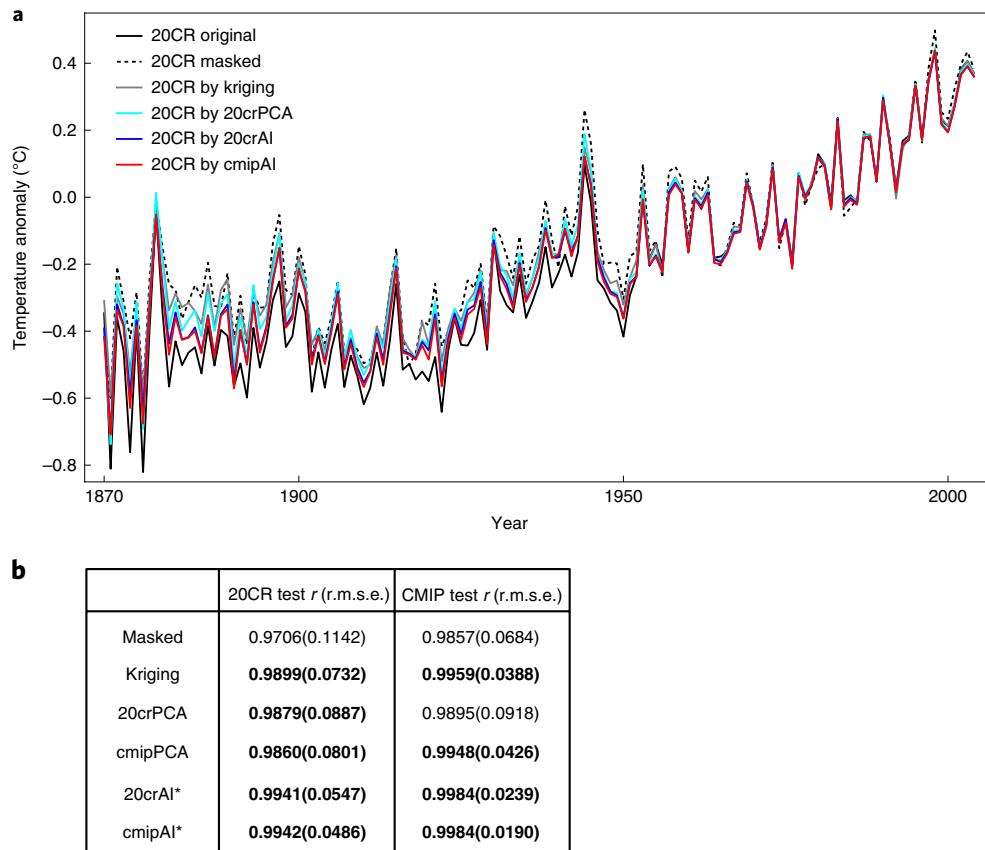
The 20crAI (cmipAI) model was trained using 1,656 (1,860) months of data from 55 (144) out of 56 (145) ensemble members, a total of 92,400 (269,568) training examples. Additional, randomly selected members, that is, the 56th (145th) members, were

completely excluded from training the statistical approach and used only for evaluation and to test reconstructions (Methods). Additionally, we performed cross-validation of the two AI models by attempting to predict the held-out ensemble members from the other dataset. This enabled us to estimate the ability of each model to generalize to data beyond its own dataset. To ensure comparability, the evaluation and test reconstructions were carried out on the period (1870–2005) of time common to both the 20CR and the CMIP experiments. In the following, we use the term ‘masked’ to indicate that the investigated climate data set has the same missing values as HadCRUT4 at each relevant time step.

We investigated if the restoration of the masked data using the AI technology is better than two alternative approaches and benchmark technologies previously applied to obtain global gridded temperature fields. These are a state-of-the-art kriging method<sup>7</sup> and an established PCA-based infilling<sup>10</sup>. After training and detailed validation of the independent test datasets, we applied both DNNs to reconstruct the missing climate values in HadCRUT4, to build and investigate new monthly climate datasets.

#### Proof-of-concept on model data with HadCRUT4 missing values

The exemplary experimental setup (Fig. 1) shows that each DNN can be trained to accurately predict the missing values from its held-out ensemble member in two example months. In each case, the neural network was able reconstruct a temperature anomaly field that is in qualitatively good agreement with the true values, when applied to data from its held-out ensemble member. This was true for both the warm and cold eastern Pacific case. To quantify the accuracy of the DNNs, we reconstructed the full time series for all the grid points and evaluated the annual global mean in comparison with the original and masked time series of 20CR (Fig. 2). Particularly in the early period, the masked time series of 20CR as the baseline for the evaluation shows a warm bias most probably imposed by the missing values<sup>32</sup>. With far fewer missing values present in the later period, the masked time series is closer to the original dataset. The 20crAI and cmipAI models achieve high temporal correlation ( $r \geq 0.9941$ ) and a low error (root mean square error (r.m.s.e.)  $\leq 0.0547$ ) when evaluated on their own held-out member and in cross-validation (Extended Data Fig. 4). As each DNN was trained on an independent dataset, it appears that both models have successfully learned to respect the underlying physical patterns and connections within the data. Both deep learning models are closer to and predict the



**Fig. 2 | Evaluation of AI models with different methods for annual global mean temperature reconstructions.** **a,b,** Temperature anomaly with respect to the 1961–1990 climatology (**a**) and metrics (**b**) of the held-out 56th 20CR member using the HadCRUT4 missing values (1870–2005). **b,** The correlation and error of the reconstruction methods compared with the original (20CR/CMIP). Bold font shows that the  $r$  and r.m.s.e. of the kriged, PCA or AI model are significantly better than that of the masked original. \*Significantly better than kriging and PCA. See Methods for the statistical significance calculations using bootstraps and Extended Data Fig. 4 for the CMIP results.

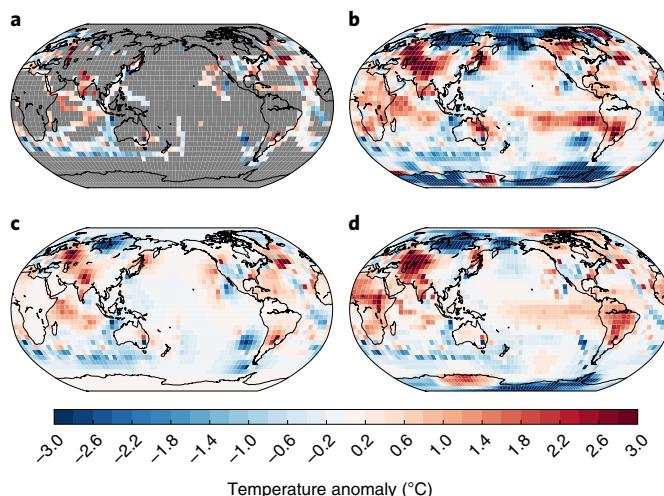
global mean temperature variability substantially better than the masked data (Fig. 2b). Especially in the early phase, the warm bias is largely reduced. An analysis using only reconstructed grid points shows a strong resemblance (Extended Data Figs. 2 and 3), especially towards the end of the time series, where the reconstructed grid points are sparse (Extended Data Figs. 5 and 6).

We also applied the kriging<sup>7</sup> interpolation scheme and the PCA-based infilling technique<sup>10</sup> used in recent temperature reconstructions on the 20CR/CMIP held-out member to be compared with the masked and DNN reconstructed data. The kriging and the PCA-based techniques improve the reconstruction relative to the masked dataset. Kriging shows its strength for the global mean time series by revealing higher correlation values and lower errors than those of the PCA method (Fig. 2b). Particularly, when applying cross-validation, the PCA method tends to introduce additional biases, which is undesirable for transfer learning<sup>7</sup>. The CMIP test case reconstructed by 20crPCA exhibits a larger error than that of the masked run (Fig. 2b and Supplementary Information). As compared with kriging and PCA, both trained DNNs (20crAI and cmipAI) provide further improvements, particularly in the early years with significantly higher correlation values and lower errors for the entire period (1870–2005).

#### Spatial and temporal reconstruction of HadCRUT4

Having established that both trained 20crAI and cmipAI are able to correctly reproduce unknown and independent climate test datasets of 20CR and CMIP, also in a cross-validation manner, now we

applied these DNNs to reconstruct HadCRUT4. As control data for the missing values do not exist, we explored reconstructions of a spatial pattern of a well-known historical event to provide qualitative validation (Fig. 3). We considered the strong El Niño phase that lasted from late 1876 to early 1878, which had a dramatic impact on both humans and nature<sup>33</sup>. In July 1877, El Niño related catastrophic rainfalls and floods were detected in western South America, especially Chile<sup>34</sup>. In most of the months in the year 1877, there is at least one north–south ship track of grid points over the eastern tropical Pacific available. However, in that region in July 1877, a large section of missing data is present in HadCRUT4 (Fig. 3a). The spatial reconstructions in that region produced by the trained DNN are remarkable, given so little information. A clear El Niño pattern can be detected in the El Niño–Southern Oscillation (ENSO)-related Pacific region, but also close to western South America and Chile (Fig. 3b). Here, the strength of the inpainting technique can be seen relative to statistical interpolation methods. Interpolation via kriging<sup>7</sup> does a good job in reconstructing grid points in the vicinity of known values (Fig. 3c). However, larger gaps are blended towards zero. Therefore, no signs of an El Niño can be seen for July 1877. Thus, although PCA does not perform as well as kriging in terms of the overall picture presented by the global mean time series (Fig. 2), in this one-month reconstruction PCA reveals more information than kriging (Fig. 3d). PCA is known to be effective for detecting ENSO-like features and was used in several ENSO-related studies<sup>12</sup>. In our case, PCA reconstructions show a present but weak warm anomaly over the central Pacific.



**Fig. 3 | AI model spatial reconstruction of an observed El Niño with many missing values in HadCRUT4.** **a–d**, Map of July 1877 by HadCRUT4 original (**a**), the reconstruction of HadCRUT4 by 20crAI (**b**), the reconstruction of HadCRUT4 by kriging from Cowtan and Way (**c**) and the reconstruction of HadCRUT4 by 20crPCA (**d**). The colours show the temperature anomalies (°C) with respect to the 1961–1990 climatology. Grid points that exist in **a** are exactly the same as those in **b–d**, because these are output compositions. More comparisons can be found in Extended Data Fig. 9.

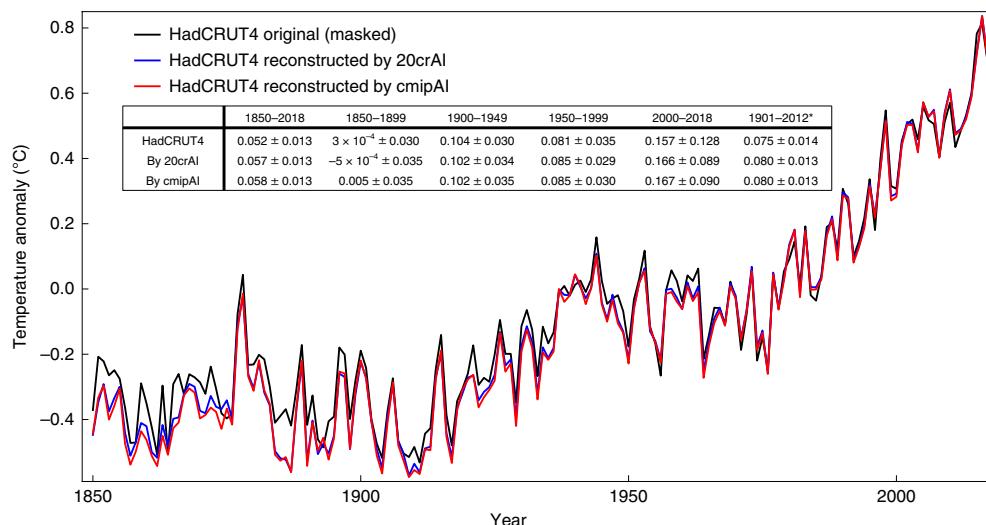
Overall, the reconstruction of the DNN shows the most meaningful signal to be interpreted. Nevertheless, the results of the two DNNs differ slightly (Extended Data Fig. 9). Both signals seem to lie further south of the equator than those produced by other interpolation techniques that also filter in time (for example, Berkley Earth<sup>9</sup> or HADISST<sup>8</sup>), and there is no way to quantitatively compare the predicted El Niño patterns with real observations. Additional analyses that support the presented DNN patterns can be found in Extended Data Fig. 9b.

Finally, we restored the missing values of HadCRUT4 by applying the two DNNs to data that spanned 1850–2018 (Fig. 4). Comparison

of HadCRUT4 with the neural network output confirms the structural behaviour, which shows a larger correction in the early years than in the later years. The early period shows a similar temperature bias in both models, as demonstrated in Fig. 2. This points to a cooler global mean temperature from the mid-nineteenth to the early twentieth century and therefore to an underestimation of the global mean temperature trend between 1850 and 2018 provided by the sparse observations (Extended Data Fig. 8). Comparing HadCRUT4 with other gridded global mean reconstructions (for example, MLOST<sup>5</sup> or GISS<sup>6</sup>) typically shows a warmer period between 1920 and 1980<sup>3,32,35</sup>. The results of the AI model reconstructions support these studies by also exhibiting a cooler period in the mid-twentieth century. The other interesting period is the early twenty-first century. Compared with HadCRUT4, both DNNs agree on a less obvious hiatus<sup>36–38</sup> due to a stronger trend<sup>7</sup> during that phase, and a general stronger trend, which includes higher values for 2016, the warmest year on record. Additional analyses that support the presented DNN time series can be found in Extended Data Fig. 7.

We anticipate our assay to be another milestone for sophisticated climate reconstructions and an alternative approach to interpolation and PCA-based infilling schemes. Temperature reconstructions from tree rings, coral, ice cores and so on, or even missing values in observational precipitation, could be investigated. Besides completing information from the past, modern reanalysis as well as seasonal to decadal prediction experiments could be initialized with retrofitted observational datasets with no missing values. The ability to reconstruct spatial maps could also be valuable for measurement campaigns, which usually provide sparse snapshots of a given climatic variable. If that same variable is present in a NM, inpainting with transfer learning could infill the missing information.

Image inpainting is a rapidly evolving field with a variety of research directions and applications that span sequence-based, GAN-based and CNN-based methods<sup>29</sup>. With no training data requirements, sequence-based methods (interpolation, diffusion and patch) are sufficient for the reconstruction of small missing parts of an image<sup>23,29,39</sup>. However, the problem of capturing the global structure remains<sup>29</sup>. GAN-based technologies can capture the global structure, learned from a rather small training set. Based on statistical distributions, GAN-based techniques mimic the data (images) in their training set. The results appear to be realistic, but are not



**Fig. 4 | AI model reconstruction of HadCRUT4 for the full time series between 1850 and 2018.** Shown is the annual global mean anomaly temperature with respect to the 1961–1990 climatology. Inset: respective linear decadal trends (°C). \*Comparison with Table 2.1 in IPCC AR5 Chapter 2 (Hartmann et al.)<sup>3</sup>. The respective trend maps to be compared with Fig. 2.21 in IPCC AR5 Chapter 2<sup>3</sup> can be found in Extended Data Fig. 8.

necessarily identical to the ground truth. This can be a disadvantage for a classification task like the one in this study. Therefore, the chosen CNN-based inpainting technology, which includes partial convolutions and the automatic mask updating mechanism, is the preferred choice for our situation in which a large training dataset is available. Transfer learning from NM data developed within this study further enhances the effectiveness of this technique. In addition, our approach requires fewer computational resources than, for example, applied kriging, and also avoids the time-intensive hyper-parameter tuning phase typically required to train GAN-based technologies. More comparisons can be found within the original study<sup>24</sup>.

These results are promising enough to warrant further research into the reconstruction of missing climate information via image inpainting using partial convolutions in a DNN: (1) HadCRUT4 consists of a 100-member dataset and usually the global mean time series is estimated by the median of the global mean (Extended Data Fig. 7), which should be further addressed in future activities. (2) Observational data reconstruction is an active research process. During the set-up of this study, an updated ocean component of the HadCRUT product was released. HadSST4<sup>40</sup> shows substantial changes to the temperature record. Additional analyses in the Extended Data show that DNN reconstructions of HadCRUT4 often point in the same direction as the blend of HadCRUT4 and HadSST4. (3) The DNN model we applied reconstructs two-dimensional images in space, but temperature reconstructions exhibit patterns in both space and time. Techniques for video-frame reconstruction based on deep learning could be investigated. (4) The observation that the cmipAI model is slightly more accurate than then 20crAI model could simply be due to the fact that the cmipAI model had almost three times as many examples available for training. Experiments that combine both training sets or include other datasets, such as the upcoming CMIP6<sup>41</sup>, could further improve the results shown here. (5) Edges of images are often problematic with image inpainting. Even though the chosen technology appears to be more effective than other solutions<sup>24</sup>, we see some artefacts and biases, especially at high latitudes (Extended Data Figs. 2 and 3). An image inpainting model that uses spherical convolutions connecting nearby points on the globe could be a beneficial research direction.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41561-020-0582-5>.

Received: 18 September 2019; Accepted: 28 April 2020;

Published online: 1 June 2020

## References

- Brázdil, R. et al. European climate of the past 500 years: new challenges for historical climatology. *Clim. Change* **101**, 7–40 (2010).
- Cubasch, U. & Kadow, C. Global climate change and aspects of regional climate change in the Berlin–Brandenburg Region. *Erde* **142**, 3–20 (2011).
- Hartmann, D. L. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T.F. et al.) Ch. 2 (IPCC, Cambridge Univ. Press, 2013).
- Morice, C. P., Kennedy, J. J., Rayner, N. A. & Jones, P. D. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 dataset. *J. Geophys. Res.* **117**, D08101 (2012).
- Vose, R. S. et al. NOAA's merged land-ocean surface temperature analysis. *Bull. Am. Meteorol. Soc.* **93**, 1677–1685 (2012).
- Lenssen, N. et al. Improvements in the GISTEMP uncertainty model. *J. Geophys. Res. Atmos.* **124**, 6307–6326 (2019).
- Cowtan, K. & Way, R. G. Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Q. J. R. Meteorol. Soc.* **133**, 459–77 (2013).
- Rayner, N. A. et al. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, 4407 (2003).
- Rhode, R. et al. A new estimate of the average Earth surface land temperature spanning 1753 to 2011. *Geoinfor. Geostat. Overview* **1**, <https://doi.org/10.4172/2327-4581.1000101> (2013).
- Beckers, J. & Rixen, M. EOF calculations and data filling from incomplete oceanographic data sets. *J. Atmos. Oceanic Technol.* **20**, 1839–1856 (2003).
- Wang, K. & Clow, G. D. Reconstructed global monthly land air temperature dataset (1880–2017). *Geosci. Data J.* <https://doi.org/10.1002/gdj3.84> (2019).
- Smith, T. M., Reynolds, R. W., Livezey, R. E. & Stokes, D. C. Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Clim.* **9**, 1403–1420 (1996).
- Kaplan, A., Kushnir, Y., Cane, M. A. & Blumenthal, M. B. Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. *J. Geophys. Res. Oceans* **102**, 27835–27860 (1997).
- Elken, J., Zujev, M., She, J. & Lagemaa, P. Reconstruction of large-scale sea surface temperature and salinity fields using sub-regional EOF patterns from models. *Front. Earth Sci.* **7**, 232 (2019).
- Reichstein, M. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).
- Monteleoni, C., Schmidt, G. A. & McQuade, S. Climate informatics: accelerating discovering in climate science with machine learning. *Comput. Sci. Eng.* **15**, 32–40 (2013).
- Barnes, E. A., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C. & Anderson, D. Viewing forced climate patterns through an AI lens. *Geophys. Res. Lett.* **46**, 13389–13398 (2019).
- Racah, E. et al. ExtremeWeather: a large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. *Adv. Neural Inform. Process. Syst.* **30**, 3405–3416 (2017).
- Kadow, C., Illing, S., Kröner, I., Ulbrich, U. & Cubasch, U. Decadal climate predictions improved by ocean ensemble dispersion filtering. *J. Adv. Modeling Earth Syst.* **9**, 1138–1149 (2017).
- Irrgang, C., Saynisch, J. & Thomas, M. Estimating ocean heat content from tidal magnetic satellite observations. *Sci. Rep.* **9**, 7893 (2019).
- Bertalmio, M., Sapiro, G., Caselles, V. & Ballester, C. Image inpainting. In *Proc. ACM Conf. Comp. Graphics (SIGGRAPH)* (eds Brown, J. R. & Akeley, K.) 417–424 (ACM/Addison-Wesley, 2000).
- Shibata, S., Iiyama, M., Hashimoto, A. & Minoh, M. Restoration of sea surface temperature satellite images using a partially occluded training set. In *24th International Conference on Pattern Recognition (ICPR), Beijing* (IEEE Computer Society) 2771–2776 (IEEE, 2018).
- Dong, J. et al. Inpainting of remote sensing SST images with deep convolutional generative adversarial network. *IEEE Geosci. Remote Sens. Lett.* **16**, 173–177 (2019).
- Liu, G. et al. in *Computer Vision—ECCV 2018 Lecture Notes in Computer Science*, Vol. 11215 (eds Ferrari, V. et al.) 19–35 (Springer, 2018).
- Barnes, C., Shechtman, E., Finkelstein, A. & Goldman, D. B. Patchmatch: a randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.* **28**, 24 (2009).
- Iizuka, S., Simo-Serra, E. & Ishikawa, H. Globally and locally consistent image completion. *ACM Trans. Graph.* **36**, 107 (2017).
- Yu, J. et al. Generative Image Inpainting with Contextual Attention. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA* 5505–5514 (IEEE/CVF, 2018).
- Perez, P., Gangnet, M. & Blake, A. Poisson image editing. *ACM Trans. Graph.* **22**, 313–318 (2003).
- Elharrouss, O., AlMaadeed, N., Al-Maadeed, S. & Akbari, Y. Image inpainting: a review. *Neural Process. Lett.* **51**, 2007–2028 (2019).
- Compo, G. P. et al. The Twentieth Century Reanalysis project. *Q. J. R. Meteorol. Soc.* **137**, 1–28 (2011).
- Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
- Folland, C. K., Boucher, O., Colman, A. & Parker, D. E. Causes of irregularities in trends of global mean surface temperature since the late 19th century. *Sci. Adv.* **eaao5297** (2018).
- Kiladis, G. N. & Diaz, H. F. An analysis of the 1877–78 ENSO episode and comparison with 1982–83. *Mon. Weather Rev.* **114**, 1035–1047 (1986).
- Aceituno, P. et al. The 1877–1878 El Niño episode: associated impacts in South America. *Clim. Change* **92**, 389–416 (2009).
- Knutson, T. R., Zhang, R. & Horowitz, L. W. Prospects for a prolonged slowdown in global warming in the early 21st century. *Nat. Commun.* **7**, 13676 (2016).
- Kosaka, Y. & Xie, S. P. Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature* **501**, 403–407 (2013).
- Saffioti, C., Fischer, E. M. & Knutti, R. Contributions of atmospheric circulation variability and data coverage bias to the warming hiatus. *Geophys. Res. Lett.* **42**, 2385–2391 (2015).

38. Marotzke, J. & Forster, P. M. Forcing, feedback and internal variability in global temperature trends. *Nature* **517**, 565–570 (2015).
39. Yan, Z. X., Li, M., Zuo, W. & Shan, S. in *Computer Vision—ECCV 2018* Lecture Notes in Computer Science, Vol. 11215 (eds Ferrari, V. et al.) 3–19 (Springer, 2018).
40. Kennedy, J. J., Rayner, N. A., Atkinson, C. P. & Killick, R. E. An ensemble data set of sea-surface temperature change from 1850: the Met Office Hadley Centre HadSST.4.0.0.0 data set. *J. Geophys. Res. Atmos.* **124**, 7719–7763 (2019).
41. Eyring, V. et al. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

## Methods

**Data.** The monthly gridded HadCRUT4-median dataset version 4.6.0.0 was downloaded from <https://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/download.html>. A total of 2,028 missing value masks (monthly values between 1850 and 2018) were prepared for training and validation using ‘1’ for existing and ‘0’ for missing values.

Monthly gridded near-surface air temperature datasets from 20CR-v2/CMIP5 were downloaded from [https://portal.nersc.gov/project/20C\\_Reanalysis](https://portal.nersc.gov/project/20C_Reanalysis) and <https://esgf-data.dkrz.de/projects/esgf-dkrz/>.

For 20CR, 55 members were used that spanned 1870–2009, which resulted in 92,400 monthly samples. The randomly selected 56th extra member was used for testing only. For the historical experiment of CMIP5 (36 models), 144 members were used that spanned 1850–2005, which resulted in 269,568 monthly samples (see Extended Data Fig. 10 for details). The randomly selected 145th extra member (ipsl-cm5b-lr<sup>12</sup>, r1i1p1) was used for testing only. Both data sets were preprocessed to provide comparability with HadCRUT4. Monthly anomalies were built using climatology values spanning 1961 and 1990. The datasets were conservatively remapped onto a 5° global grid (72×36) matching the HadCRUT4 dataset.

The kriging dataset for Fig. 3 and the software for kriging of the masked 20CR 56th member in Fig. 2 were downloaded from <https://www-users.york.ac.uk/~kdc3/papers/coverage2013/series.html>.

**Preparation for the AI models.** Applying the climate data operators, all datasets were conservatively remapped to a 5°/2.5° (72×72) equally shaped grid and converted into a NetCDF4 (Network Common Data Form) format to be compliant with HDF5 (Hierarchical Data Format version 5). One member each was excluded from the training set to be independently tested. Therefore, the training sets consisted of 55 and 144 members for the 20crAI and cmipAI models, respectively. The training sets of the 20crAI and cmipAI models were split into sets that contained 82,133 and 239,616 training samples and 10,267 and 29,952 validation samples, respectively, using every ninth month for validation (see the scheme in Extended Data Fig. 1).

**Deep-learning method and software.** Liu et al.<sup>24</sup> introduced the neural network architecture applied in this article. The goal was to propose a model for image inpainting that performs robustly on irregular holes and produces semantically meaningful predictions that match smoothly with the rest of the data without the need for additional postprocessing or blending. Their technique outperformed previous state-of-the-art techniques, which include PatchMatch<sup>25</sup> (iterative search for best-fitting patched), Iizuka et al.<sup>26</sup> (fast marching and Poisson image blending) and Yu et al.<sup>27</sup> (following-up refinement network). A main advantage of the Liu et al.<sup>24</sup> technique is that it works with irregular-shaped holes, whereas the other methods are limited to rectangularly shaped holes. Our specific implementation was based on the original NVIDIA implementation downloaded from <https://github.com/NVIDIA/partialconv>. However, we applied and adapted the ‘ready2start’ modification described in <https://github.com/naoto0804/pytorch-inpainting-with-partial-conv>.

The required software are Python3.6, PyTorch and CUDA10.

The software was adapted as follows:

- Read HDF5 and/or NetCDF input data into tensors using Numpy.
- Reduce to a 72×72 pixel ‘image’ or HDF5 set.
- Read random (train) and specific (evaluate) missing-value masks.
- Reduce number of partial convolutional layers to three instead of seven.
- Replicate the temperature data to fill the three RBG image channels, but output a single channel.
- Disable image normalization.
- Convert the HDF5 output data back to the NetCDF4 format using NetCDF operators.

Both AI models were trained using 500,000 iterations with an additional 500,000 iterations for fine-tuning by applying a batch size of 18 on a NVIDIA GeForce 1080Ti at approximately 17 bits s<sup>-1</sup>.

The trained DNNs were used to test the extra members of 20CR/CMIP5 and finally to reconstruct the HadCRUT4 median dataset and others.

**Post-processing AI model output.** The NetCDF climate data operators were used to remap the network output back to the original 72×32 grid for comparison and evaluation, which included temporal and spatial field correlation, r.m.s.e., trends and so on. Climate data operators were also used to calculate the global annual mean time series of the original, masked, AI, kriged and PCA-based data following Intergovernmental Panel on Climate Change<sup>3</sup> rules—field mean before annual mean. Statistical significances for the annual global mean time series in correlation and the r.m.s.e. in terms of its skill score were calculated using the MurCSS<sup>43</sup> tool for climate evaluation. The statistical significance was calculated with values significantly different from zero, exceeding a 5% level, by applying 1,000 block bootstraps<sup>43</sup>.

**Inpainting via PCA-learning method.** We followed the instructions provided by Beckers and Rixen<sup>10</sup> to reconstruct the test datasets of 20CR and CMIP5 with

the principal component analysis (PCA). We replaced missing values with zeros, performed the PCA, performed a back transformation, replaced the grid points with their original values where possible, calculated the PCA again and so on, iteratively. Using this approach, 50 iterations were performed, at which point no substantial differences could be detected in the known values. To find the optimum number of principal components (PCs) to be calculated, the r.m.s.e. test was applied to held-out grid points as suggested by Beckers and Rixen<sup>10</sup>. Three additional numerical experiments were conducted with the PCA reconstruction to test the transfer-learning approach. PCA reconstruction of each 20CR/CMIP5 test case was performed with: (1) the set of test data (full time series) itself, which is the basic method and is comparable with several other studies, (2) the test data (full time series) together with the 20CR/CMIP5 training data, which introduces transfer learning but retains the test data for potential additional accuracy, and (3) each timestep of the test run plus the 20CR/CMIP5 training set, which is comparable to the AI reconstruction, which is also done on each time step. A detailed comparison can be found in the Supplementary Information. Method (3) was applicable and performed better than method (1).

## Data availability

A software snapshot, trained AI models (checkpoints), missing value masks and the HadCRUT4 reconstructions by the AI models can be downloaded at <https://doi.org/10.5281/zenodo.3766741>. Training data from 20CR and CMIP5 cannot be hosted due to copyrights, but are available at National Oceanic and Atmospheric Administration and ESGF (Methods). Contact [kadow@dkrz.de](mailto:kadow@dkrz.de) for further information. Source Data are provided with this paper.

## Code availability

All the code utilized in this project can be downloaded here or cloned here at <https://github.com/FREVA-CLINT/climaterereconstructionAI>. This code will be updated and changed over time.

## References

42. Dufresne, J. L. et al. Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Clim. Dyn.* **40**, 2123–2165 (2013).
43. Illing, S., Kadow, C., Oliver, K. & Cubasch, U. MurCSS: a tool for standardized evaluation of decadal hindcast systems. *J. Open Res. Softw.* **2**, e24 (2014).

## Acknowledgements

We thank the HPC-Service of ZEDAT, Freie Universität Berlin and the German Climate Computing Center (DKRZ) for the computation resources; the Climatic Research Unit (CRU) of the University East Anglia (UEA) and the MetOffice UK for providing the HadCRUT4 and HadSST4 datasets; the Earth System Grid Federation (ESGF) for providing the CMIP5 experiments; J. Marotzke (MPI-M), M. Schuster (FUB), E. Barnes (CSU), K. Buscher (UKM) for discussions; N. Inoue (University of Tokyo) for providing the applicable code for image inpainting; A. Richling (FUB) for reproducing the Intergovernmental Panel on Climate Change trend, uncertainty and confidence values; K. Cowtan, R. Way, and the University of York for not just providing the reconstructed HadCRUT4 data (used in Fig. 3b), but also software to apply the kriging scheme (used in Fig. 2). Support for the 20CR Project dataset is provided by the US Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) programme, by the Office of Biological and Environmental Research (BER) and by the National Oceanic and Atmospheric Administration Climate Program Office.

## Author contributions

C.K. initiated the study design, coded the AI technology for climate research, performed the analysis and drafted the paper. D.M.H. supervised the NVIDIA AI technology and U.U. supervised the climate research results. All the authors discussed the results and edited the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

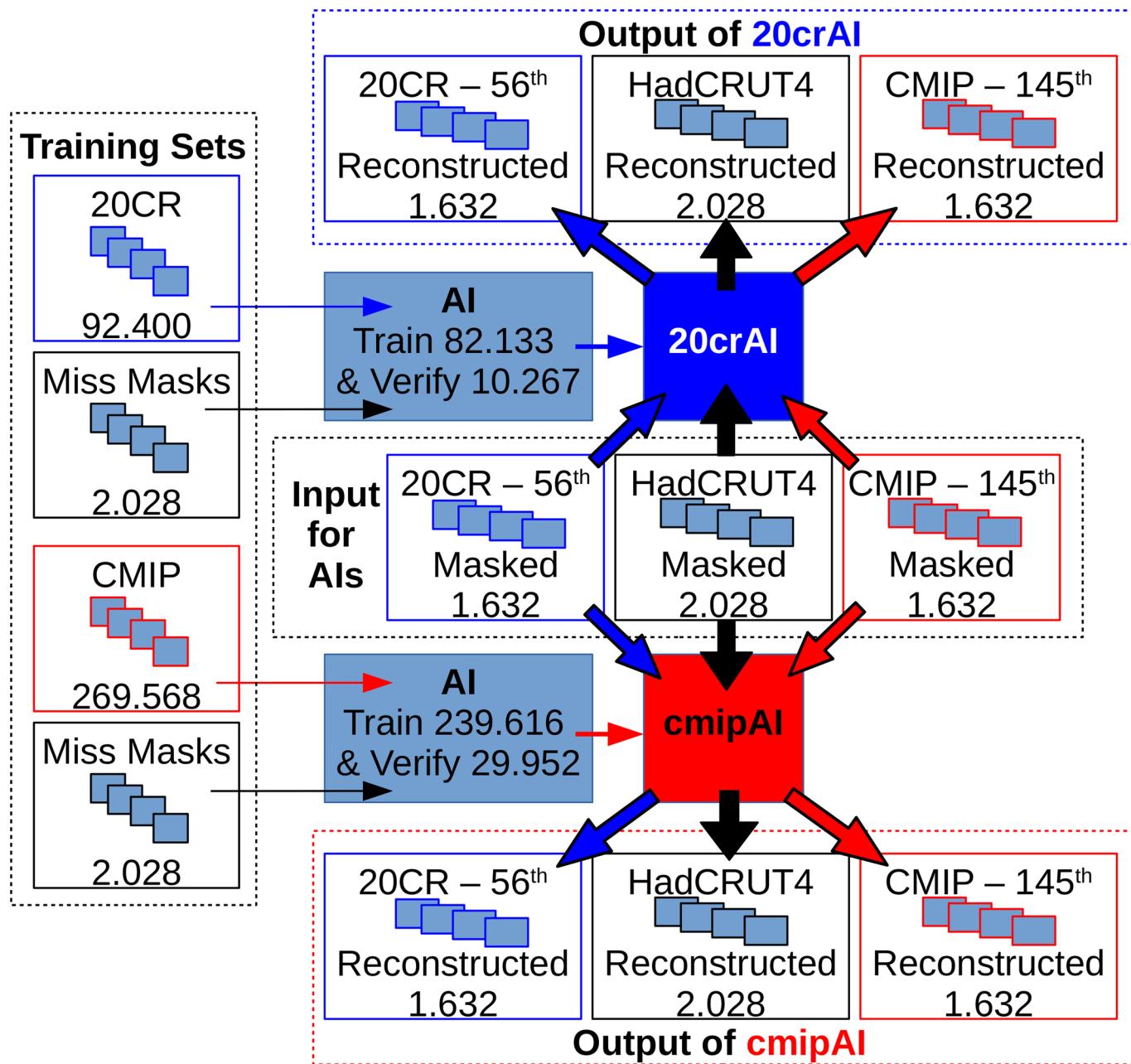
**Extended data** is available for this paper at <https://doi.org/10.1038/s41561-020-0582-5>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41561-020-0582-5>.

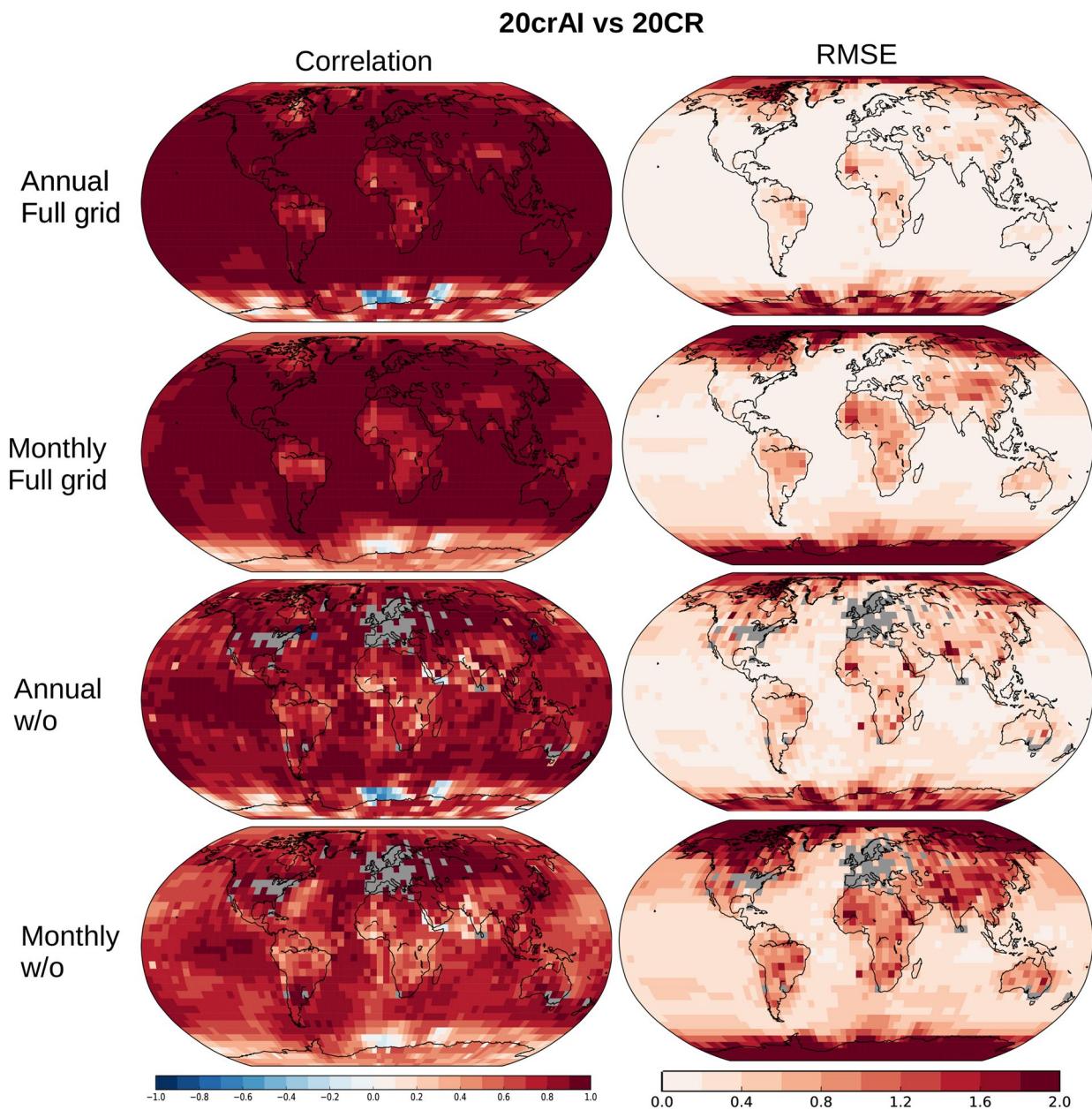
**Correspondence and requests for materials** should be addressed to C.K.

**Peer review information** Primary Handling Editors: Stefan Lachowycz; Heike Langenberg.

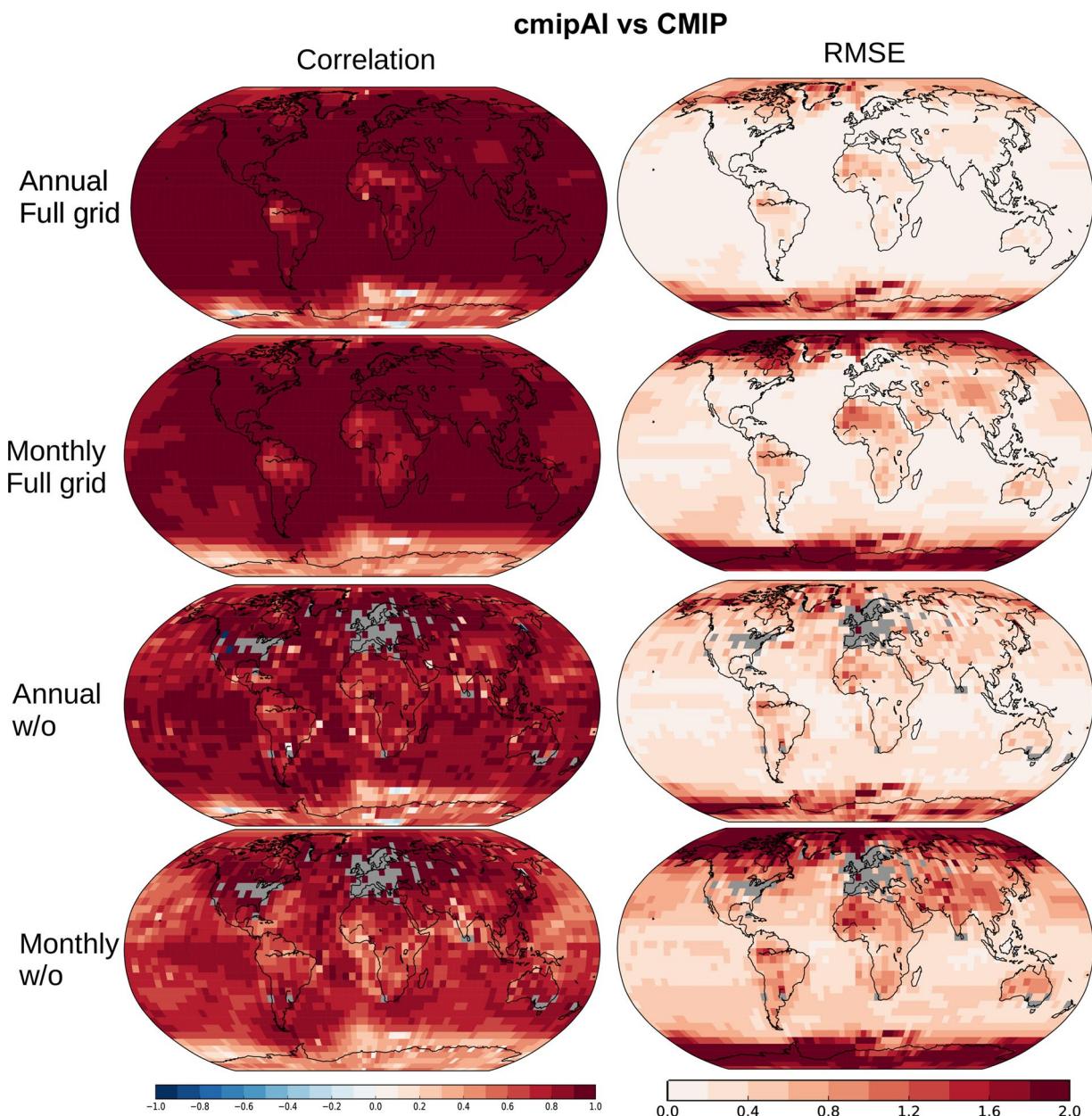
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



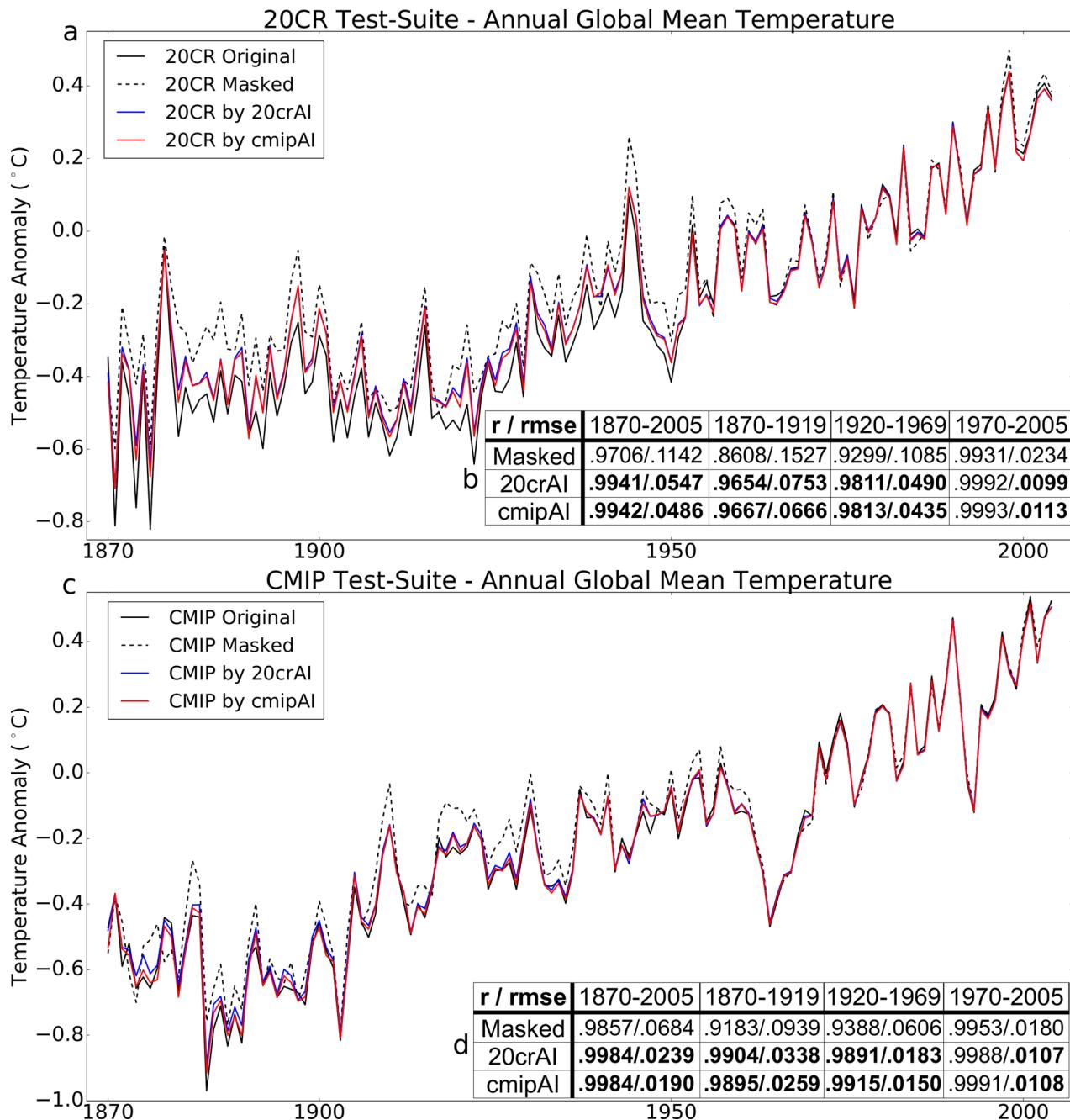
**Extended Data Fig. 1 | Scheme for the study setup including training set.** Input for the AI models, training of the models, and their output. HadCRUT4 data in black, CMIP data or AI in red, 20CR data or AI in blue. Numbers on the bottom of the boxes represent the number of ‘images’ / months / time steps, which are used as input or result as output (see Method section).



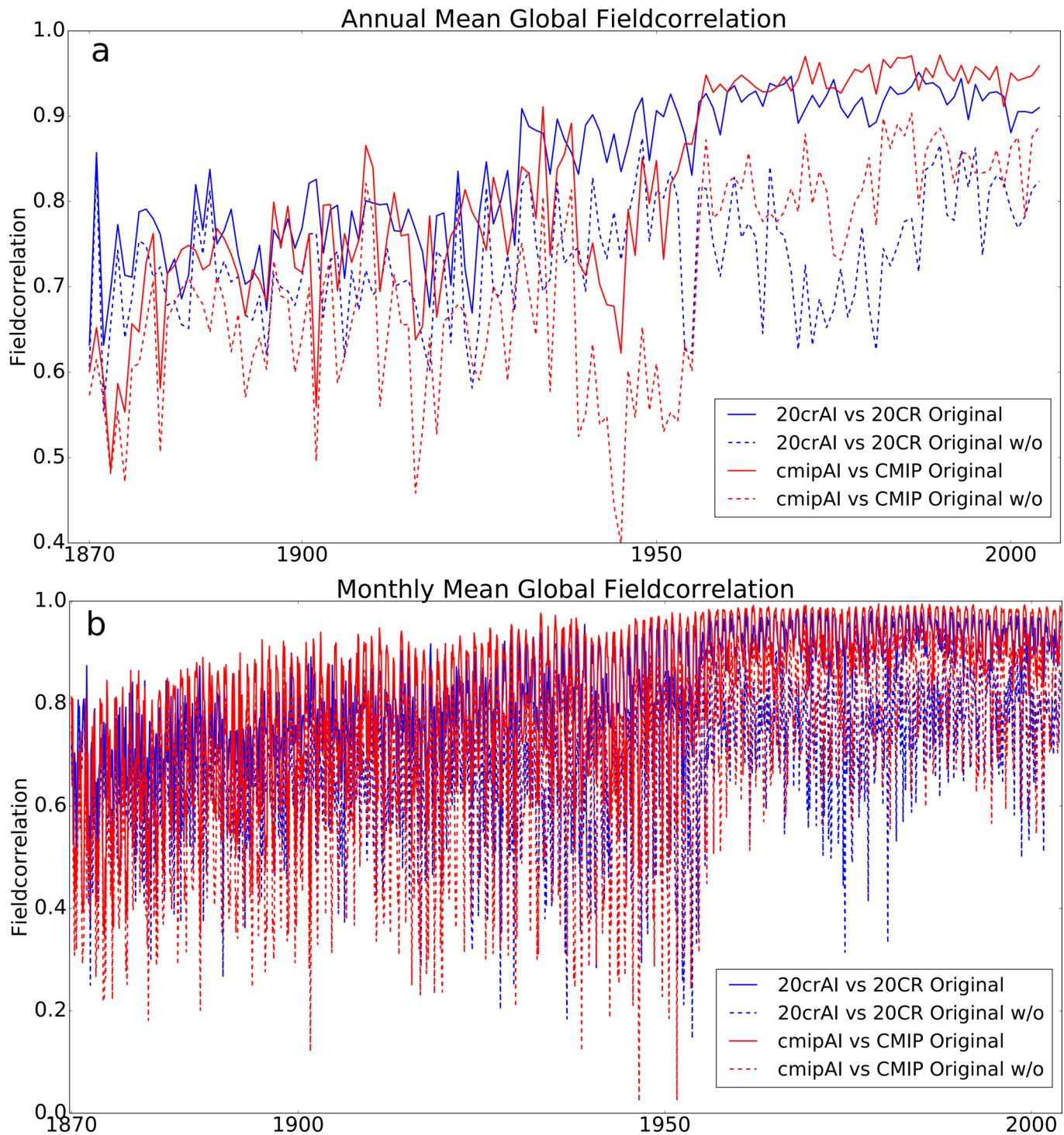
**Extended Data Fig. 2 | Detailed grid space evaluation of 20CR reconstruction.** Correlation (left) and root mean squared error in centigrade (right) comparing the reconstructed 20CR 56th member by the 20crAI model with the original 20CR 56th member. Comparison of all grid points in an annual (row 1) and monthly (row 2) analysis. The respective analysis for the reconstructed grid points only, without (w/o) grid points which were evident during reconstruction below (row 3/4). Grey grid points indicate points that exist for the whole time series.



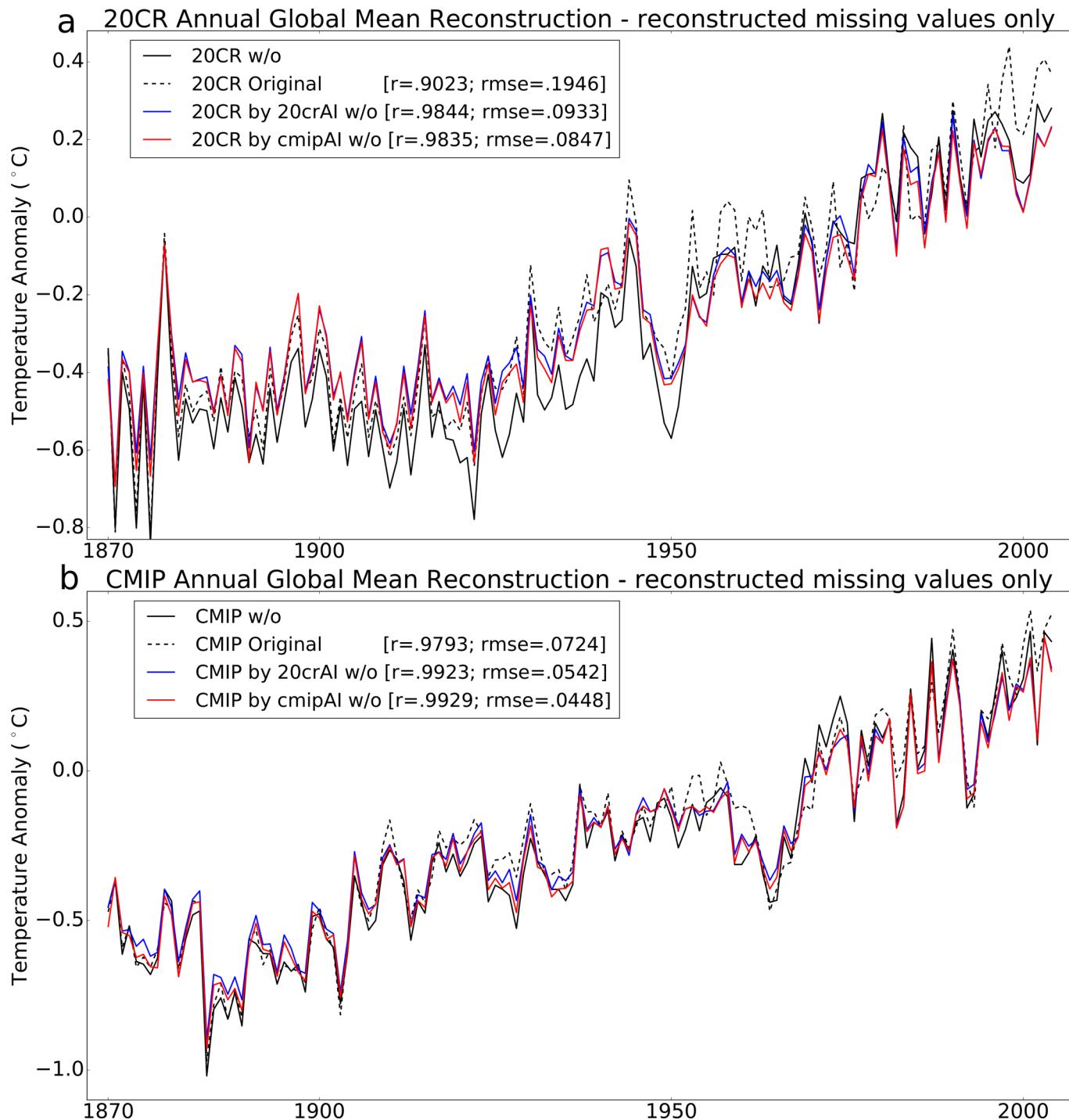
**Extended Data Fig. 3 | Detailed grid space evaluation of CMIP reconstruction.** Correlation (left) and root mean squared error in centigrade (right) comparing the reconstructed CMIP 145<sup>th</sup> member by the cmipAI model with the original CMIP 145<sup>th</sup> member. Comparison of all grid points in an annual (row 1) and monthly (row 2) analysis. The respective analysis for the reconstructed grid points only, without (w/o) grid points which were evident during reconstruction below (row 3/4). Grey grid points indicate points that exist for the whole time series.



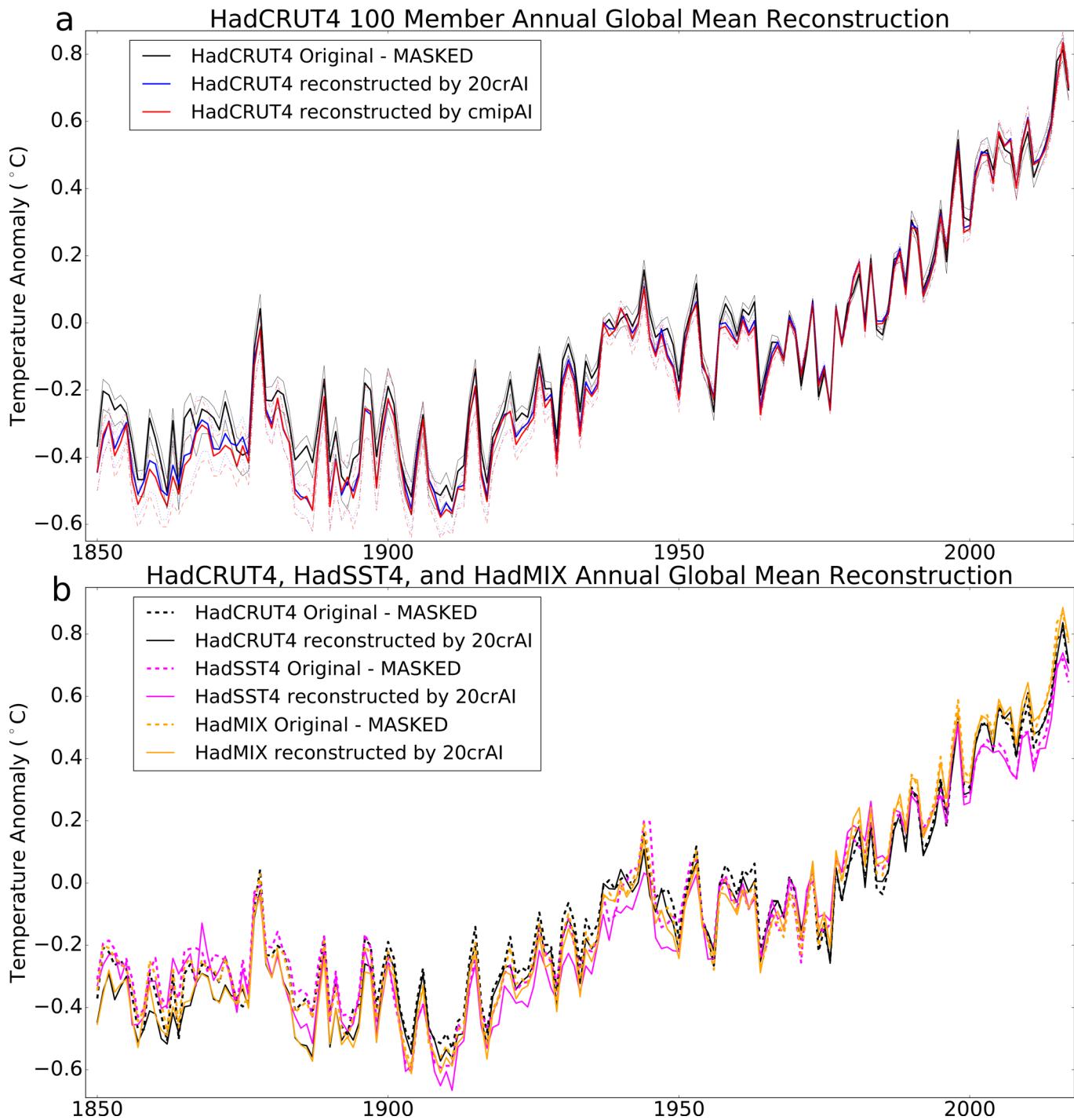
**Extended Data Fig. 4 | Time-series analysis and evaluation of AI model reconstruction.** As Fig. 2, but the annual global mean anomaly temperature reconstructions in centigrade of 20CR (a, b) / CMIP (c, d) test-suite of monthly grid reconstructions of the held-out 56th / 145th member using the HadCRUT4 missing value mask (1870-2005). In black the original held-out member, in black-dashed the original but masked held-out member to see the effect of the missing values. In blue/red the reconstructed grid time-series of the 20crAI/cmipAI. Tables show anomaly correlation ( $r$ ) and root mean squared error (rmse) compared to the original dataset on four selected time ranges. (see also Fig. 2).



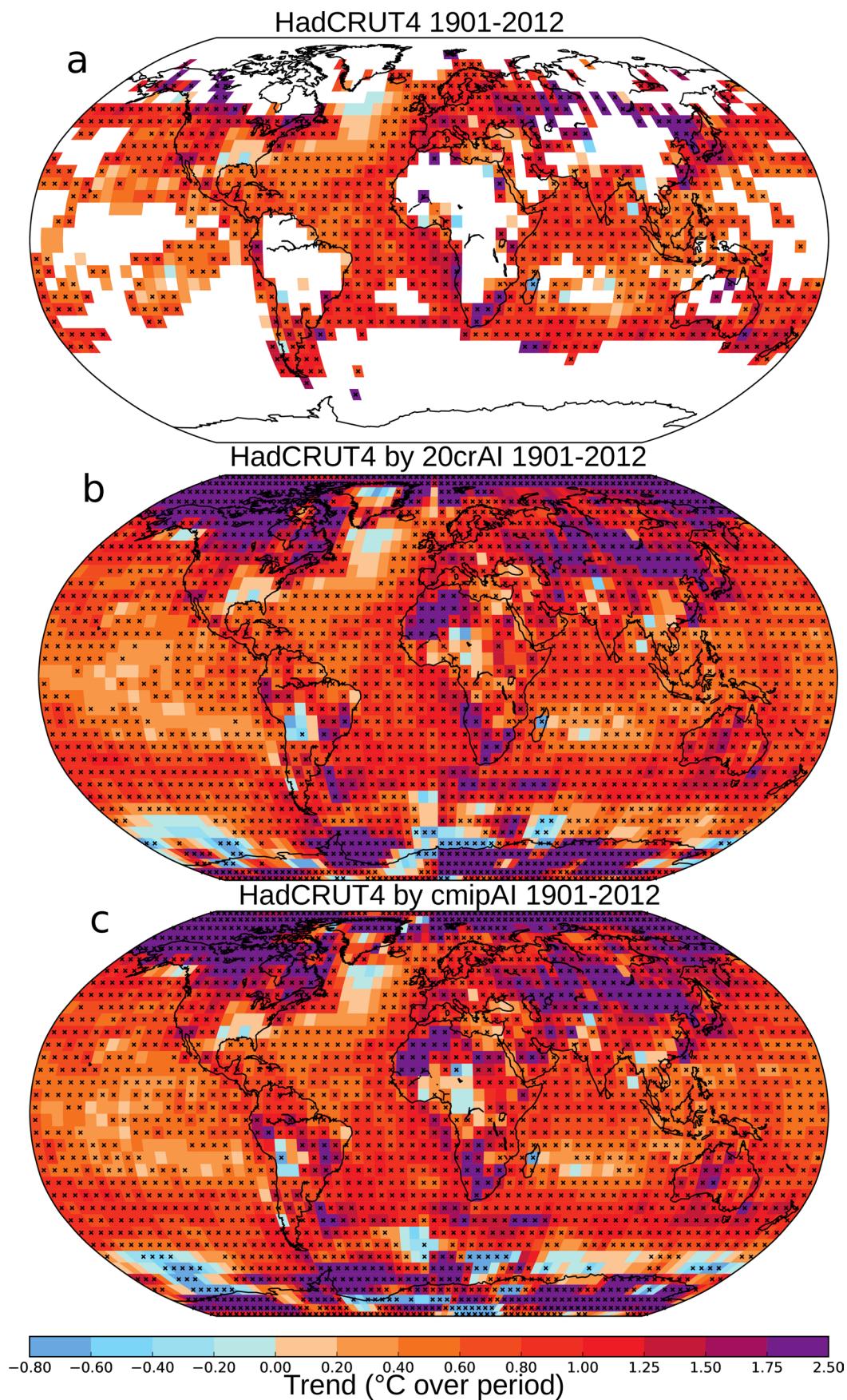
**Extended Data Fig. 5 | Spatial evaluation of AI models over time.** Fieldcorrelation of the annual (a) and monthly (b) mean reconstruction of the 20CR 56th / CMIP 145th member by the 20crAI / cmipAI models with the original 20CR 56th / CMIP 145th member in blue / red. Solid line compares the full grid space, while the dashed line respective analysis for the reconstructed grid points only, without (w/o) grid points which were evident during reconstruction.



**Extended Data Fig. 6 | Evaluation on reconstructed grid points only.** Annual global mean anomaly temperature reconstruction in centigrade of 20CR (a) and CMIP (b) of monthly grid reconstructions applying only reconstructed missing values the extra 56th / 145th member using the HadCRUT4 missing value mask between 1870 and 2005. In black the extra member without (w/o) existing grid points, in black-dashed the original full left-out member to see the effect of the missing values. In blue/red the reconstructed grid time-series of the 20crAI/cmipAI models without (w/o) existing grid points.

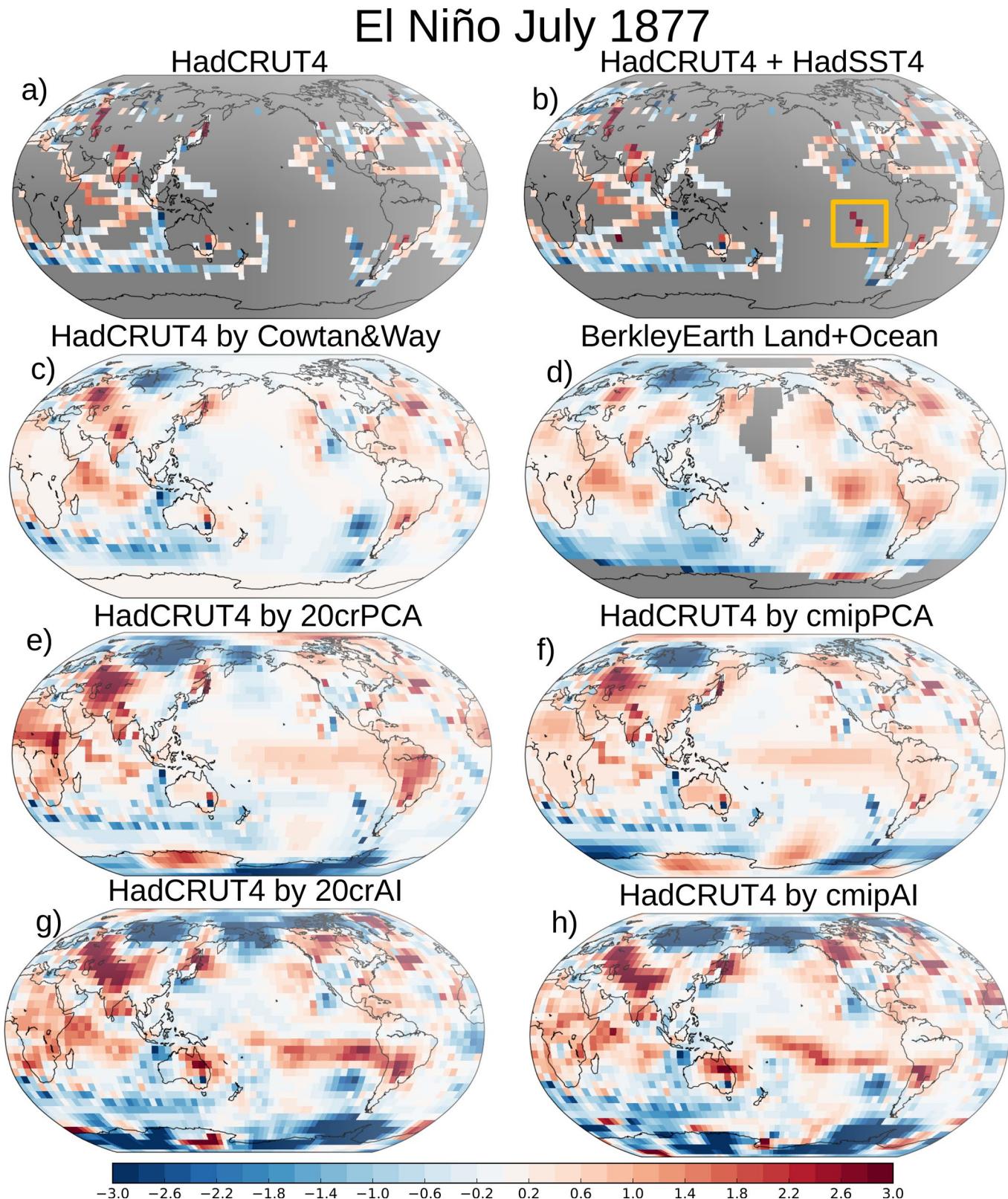


**Extended Data Fig. 7 | Reconstruction analysis of additional Hadley Centre products.** Annual global mean anomaly temperature time series between 1850 and 2018. (a) HadCRUT4 original (masked) 100 member data in black (median, 95th, 5th percentile). The HadCRUT4 reconstruction of the 20crAI/cmipAI models in blue/red (median, 95th, 5th percentile). (b) HadCRUT4 original (masked) data in black, HadSST4 original (masked) data in pink, HadMIX original (masked) data in orange. The originals are dashed, the reconstructions have straight lines. HadMIX has all grid points available of HadSST4, if not available (usually over land) HadCRUT4 grid points are used.



Extended Data Fig. 8 | See next page for caption.

**Extended Data Fig. 8 | HadCRUT4 trends of AI models in grid space.** Trends in surface temperature from Fig. 4 for 1901–2012. White areas indicate incomplete or missing data. Trends have been calculated only for those grid boxes with greater than 70% complete records and more than 20% data availability in first and last decile of the period. Black plus signs (+) indicate grid boxes where trends are significant (i.e., a trend of zero lies outside the 90% confidence interval). Graphics are constructed, to be compared with IPCC AR5 Chapter 2 Figure 2.21. Here HadCRUT4 Version 4.6.0.0 is used, IPCC report used Version 4.1.1.



**Extended Data Fig. 9 | Spatial reconstruction of an observed El Niño.** As Fig. 3 but with additional datasets. Recently, the HadSST4 (b) data set was released as an update to HadSST3 (ocean component of HadCRUT4 (a)). Kriging analysis of Cowtan&Way (c) is set next to Berkley Earth (d). In July 1877 HadSST4 has three new grid points, which show very high (warm) temperature anomalies in a region (further south than usual) where the PCA reconstruction of 20crPCA (e) and cmipPCA (f) show some weak signal. Neural network reconstructions of 20crAI (g) and cmipAI (h) show some strong signal of an El Niño like temperature pattern.

Model	N°	Ensemble	Model	N°	Ensemble	Model	N°	Ensemble
access1-0 <sup>44</sup>	3	r1i1p1 - r3i1p1	cmcc-cm <sup>50</sup>	1	r1i1p1	inmcm4 <sup>57</sup>	1	r1i1p1
access1-3 <sup>45</sup>	3	r1i1p1 - r3i1p1	cmcc-cms <sup>50</sup>	1	r1i1p1	ipsl-cm5a-lr <sup>42</sup>	6	r1i1p1 - r6i1p1
bcc-csm1-1-m <sup>46</sup>	3	r1i1p1 - r3i1p1	cnrm-cm5-2 <sup>51</sup>	1	r1i1p1	ipsl-cm5a-mr <sup>42</sup>	3	r1i1p1 - r3i1p1
bcc-csm1-1 <sup>46</sup>	3	r1i1p1 - r3i1p1	cnrm-cm5 <sup>51</sup>	10	r1i1p1 - r10i1p1	miroc5 <sup>58</sup>	5	r1i1p1 - r5i1p1
bnu-esm <sup>47</sup>	1	r1i1p1	csiro-mk3-6-0 <sup>52</sup>	10	r1i1p1 - r10i1p1	miroc-esm-chem <sup>58</sup>	1	r1i1p1
canesm2 <sup>48</sup>	5	r1i1p1 - r5i1p1	ec-earth <sup>53</sup>	5	r1i1p1, r2i1p1, r8i1p1, r9i1p1, r12i1p1	miroc-esm <sup>58</sup>	3	r1i1p1 - r3i1p1
ccsm4 <sup>49</sup>	6	r1i1p1 - r6i1p1	fgoals-g2 <sup>54</sup>	4	r1i1p1, r3i1p1, r4i1p1, r5i1p1	mpi-esm-lr <sup>59</sup>	3	r1i1p1 - r3i1p1
cesm1-bgc <sup>49</sup>	1	r1i1p1	fio-esm <sup>55</sup>	3	r1i1p1 - r3i1p1	mpi-esm-mr <sup>59</sup>	3	r1i1p1 - r3i1p1
cesm1-cam5 <sup>49</sup>	3	r1i1p1 - r3i1p1	giss-e2-h-cc <sup>56</sup>	1	r1i1p1	mpi-esm-p <sup>59</sup>	2	r1i1p1, r2i1p1
cesm1-fastchem <sup>49</sup>	3	r1i1p1 - r3i1p1	giss-e2-h <sup>56</sup>	17	r1i1p1 - r6i1p1, r1i1p2 - r5i1p2, r1i1p3 - r6i1p3	mri-cgcm3 <sup>60</sup>	5	r1i1p1 - r5i1p1
cesm1-waccm <sup>49</sup>	1	r1i1p1	giss-e2-r-cc <sup>56</sup>	1	r1i1p1	noresm1-me <sup>61</sup>	2	r1i1p1, r1i1p2
cmcc-cesm <sup>50</sup>	1	r1i1p1	giss-e2-r <sup>56</sup>	18	r1i1p1 - r6i1p1, r1i1p2 - r6i1p2, r1i1p3 - r6i1p3	noresm1-m <sup>61</sup>	3	r1i1p1 - r3i1p1

**Extended Data Fig. 10 | CMIP numerical models to train the neural network.** CMIP5 Historical monthly experiments between 1850 and 2005 applied to train the cmipAI. Data from refs. <sup>42,44–61</sup>.

44. Lewis, S. C. & Karoly, D. J. Assessment of forced responses of the Australian Community Climate and Earth System Simulator (ACCESS) 1.3 in CMIP5 historical detection and attribution experiments. *Aust. Meteorol. Oceanogr. J.* **64**, 87–101 (2014).
45. Collier, M. & Uhe, P. *CMIP5 Datasets from the ACCESS1.0 and ACCESS1.3 Coupled Climate Models* CAWCR Technical Report 059 (CAWCR, 2012).
46. Xin, X., Wu, T. & Zhang, J. Introduction of CMIP5 experiments carried out with the climate system models of Beijing Climate Center. *Adv. Clim. Change Res.* **4**, 41–49 (2013).
47. Ji, D., Wang, L., Feng, J., Wu, Q. & Cheng, H. *BNU-ESM Model Output Prepared for CMIP5 rcp45 Experiment, Served by ESGF* (WDCC at DKRZ, 2015); <https://doi.org/10.1594/Wdcc/CMIP5.BUBU4>
48. Canadian Centre for Climate Modelling and Analysis (CCCma). *CanESM2 Model Output Prepared for CMIP5 Historical, Served by ESGF* (WDCC at DKRZ, 2015); <https://doi.org/10.1594/Wdcc/CMIP5.CCE2hi>
49. Scoccimarro, E. et al. Effects of tropical cyclones on ocean heat transport in a high resolution coupled general circulation model. *J. Clim.* **24**, 4368–4384 (2011).
50. Centre National de Recherches Météorologiques and Centre Européen de Recherche et Formation Avancée en Calcul Scientifique WCRP CMIP5: *The CNRM-CERFACS Team CNRM-CM5-2 Model Output for the Historical Experiment* (Centre for Environmental Data Analysis, 2017); <http://catalogue.ceda.ac.uk/uuid/6ea812758cf14de8a5577406e896c3f9>
51. Rotstayn, L. et al. Improved simulation of Australian climate and ENSO-related climate variability in a GCM with an interactive aerosol treatment. *Int. J. Climatol.* **30**, 1067–1088 (2010).
52. Hazeleger, W. et al. EC-Earth. *Bull. Am. Meteor. Soc.* **91**, 1357–1364 (2010).
53. Li, L. et al. The flexible global ocean-atmosphere-land system model, Grid-point Version 2: FGOALS-g2. *Adv. Atmos. Sci.* **30**, 543–560 (2013).
54. Qiao, F. et al. Development and evaluation of an Earth System Model with surface gravity waves. *J. Geophys. Res. Oceans* **118**, 4514–4524 (2013).
55. Miller, R. L. et al. CMIP5 historical simulations (1850–2012) with GISS ModelE2. *J. Adv. Model. Earth Syst.* **6**, 441–477 (2014).
56. Volodin, E. M., Dianskii, N. A. & Gusev, A. V. Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations. *Atmos. Ocean. Phys.* **46**, 414–431 (2010).
57. Watanabe, M. et al. Improved climate simulation by MIROC5: mean states, variability, and climate sensitivity. *J. Clim.* **23**, 6312–6335 (2010).
58. Giorgetta, M. et al. *CMIP5 Simulations of the Max Planck Institute for Meteorology (MPI-M) based on the MPI-ESM-LR Model: the rcp45 Experiment, Served ESGF* (WDCC at DKRZ, 2012); <https://doi.org/10.1594/Wdcc/CMIP5.MXELr4>
59. Meteorological Research Institute (MRI) *MRI-CGCM3 Model Output Prepared for CMIP5, Served by ESGF* (WDCC at DKRZ, 2012); <http://cera-www.dkrz.de/Wdcc/CMIP5/Compact.jsp?acronym=MRMC>
60. Iversen, T. et al. The Norwegian Earth System Model, NorESM1-M—Part 2: climate response and scenario projections. *Geosci. Model Dev.* **6**, 389–415 (2013).
61. Gent, P. R. et al. The Community Climate System Model version 4. *J. Clim.* **24**, 4973–4991 (2011).