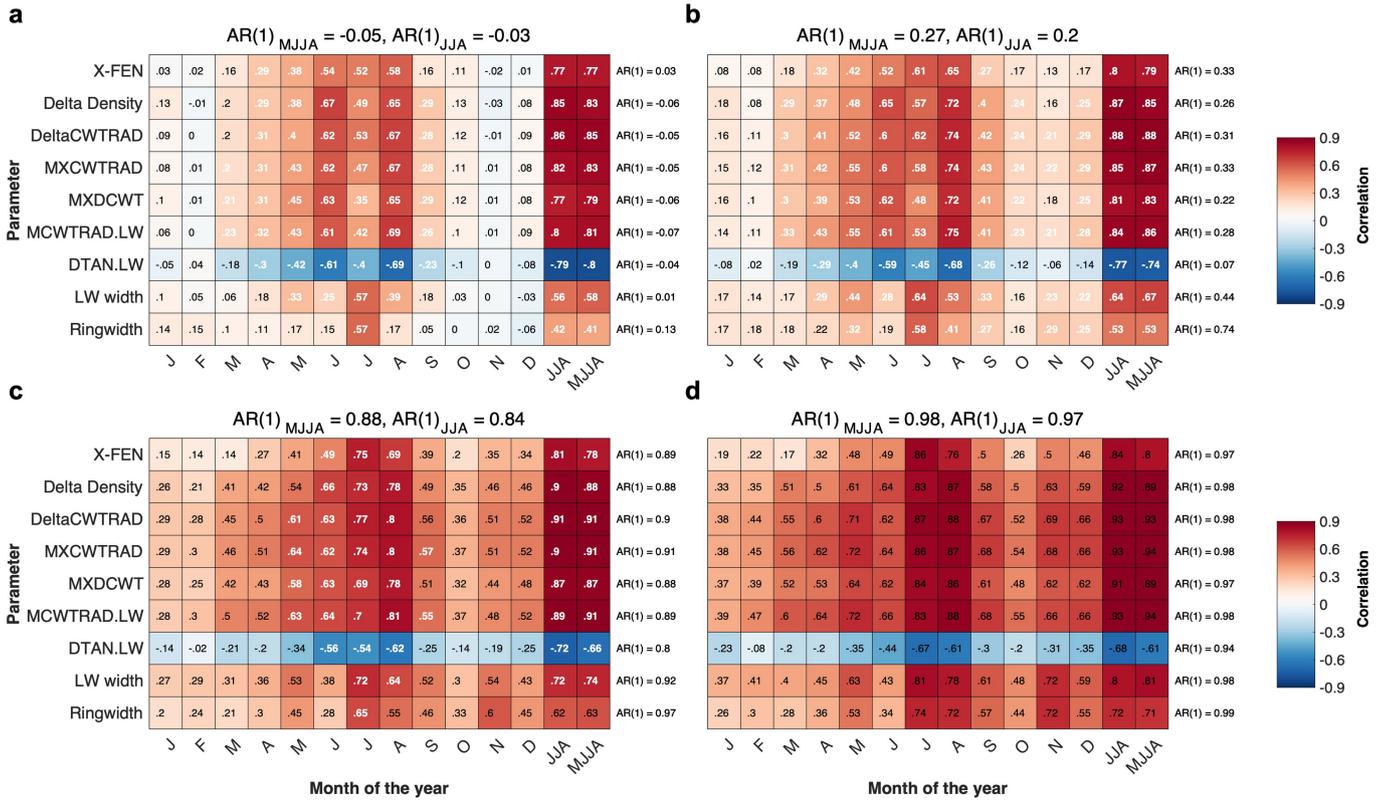


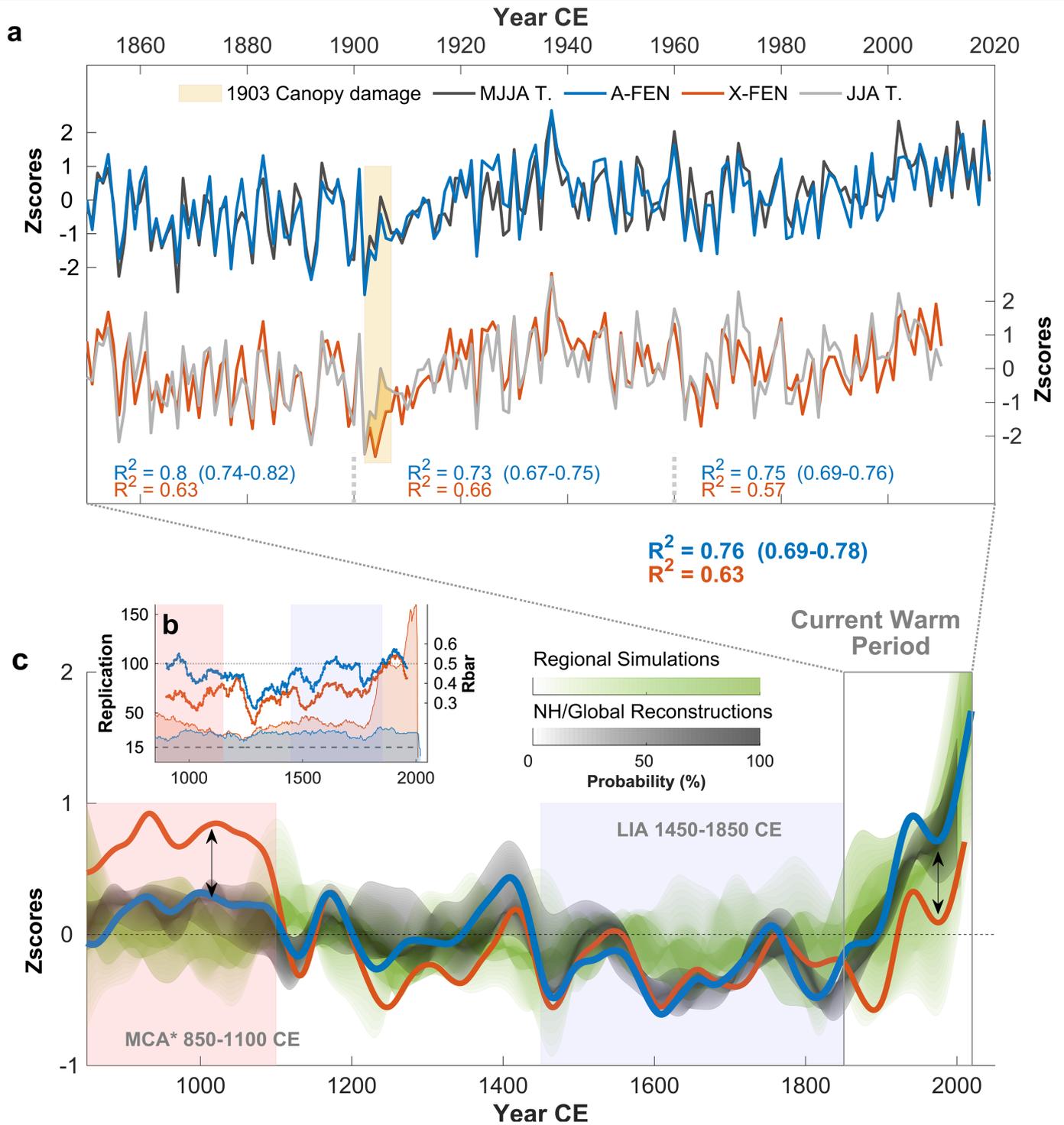
**Extended Data Fig. 1 | Evaluation of the ability of the A-FEN and X-FEN reconstructions to capture extremes in climate targets.** The instrumental temperature data was sorted from coldest to warmest and plotted together with the reconstruction values of the corresponding years. The grey boxes are bound by the 10<sup>th</sup> coldest and warmest years, and the 10<sup>th</sup> and 90<sup>th</sup> percentile of the zscore temperatures, respectively. If an extreme reconstruction value is found within the grey box, the extreme is defined as “captured”. The sum of the captured values divided by the potential sum of values, is calculated and presented as a percentage of extreme value capturing (EVC). In McCarroll, et al.<sup>21</sup>, a significance testing was implemented, and for 160–170 years of climate data,  $p < 0.001$  is achieved if more than 40% of values are captured. **a)** A-FEN’s ability

to capture JJA temperature extremes. **b)** X-FEN’s ability to capture JJA temperature extremes. **c)** and **d)** show same analysis as **a)** and **b)** but using the target MJJA. Both datasets thus display significant amounts of extremes captured, but the A-FEN captures significantly more than the X-FEN for the MJJA target season. The X-FEN captures a higher percentage of cold extremes if the MJJA target season is used but the same percentage of warm extremes regardless of target season. The rationale for using MJJA as the target season for X-FEN is thus less clearcut than for the A-FEN. JJA is the target season used in the publications originally presenting the MXD data<sup>2,37</sup> and is thus used in the main text for the other comparisons.



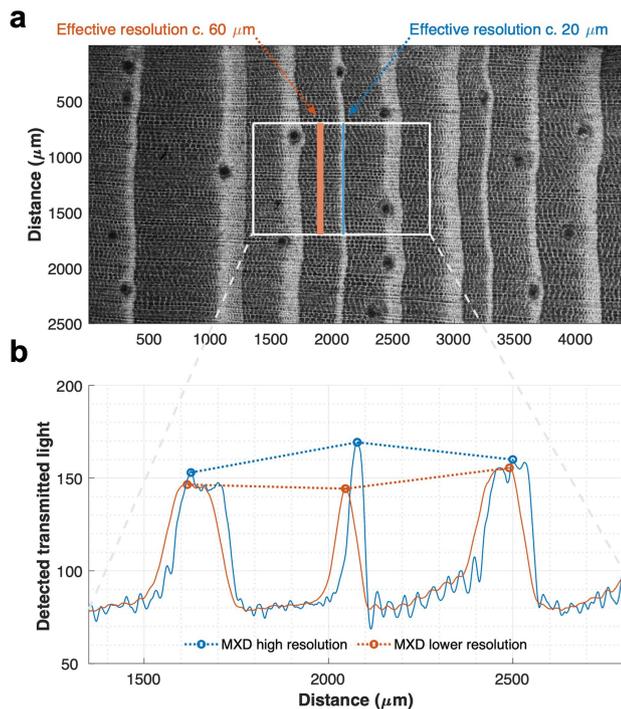
**Extended Data Fig. 2 | Monthly climate correlations across a range of frequency domains.** **a**) high-pass filtered data (cubic smoothing splines with 50% frequency response cut-off at 40 years (HP40yrs)) correlated with identically treated temperature data. **b**) RCS detrended data correlated with untreated temperature data. **c**) low-pass filtered data (LP5yrs), and **d**) (LP10yrs), correlated with identically treated temperature data, respectively. The monthly temperature data were retrieved from HadCRUT5<sup>68</sup> (5° gridded monthly dataset, Lat. 65–70° N, Lon. 15–30° E). Correlation coefficients in white are significant at  $p < 0.01$ , and black coefficients are insignificant. When 10-year low-pass filtered data are used, the autocorrelation is so high that it is impossible to detect significance after adjusting for loss of degrees of

freedom<sup>79</sup>, why it is meaningless to continue the analysis over even lower frequencies. The parameters or reconstructions reside on the y-axis, and each monthly temperature or monthly target season on the x-axis. First order autocorrelation,  $AR(1)$ , of the JJA and MJJA temperatures are given on top of each panel as a reference, and the tree-ring parameter  $AR(1)$  can be found in the right margin of each panel. The period of analysis covers the full length-overlap between all datasets (1850–2019 for anatomical parameters and 1850–2010 for the X-FEN). The results are very similar if the 1850–2010 period is used for the QWA data. The delta radial cell wall thickness (DeltaCWTRAD) parameter was established as predictor for the A-FEN reconstruction due to overall performance in the analysis.

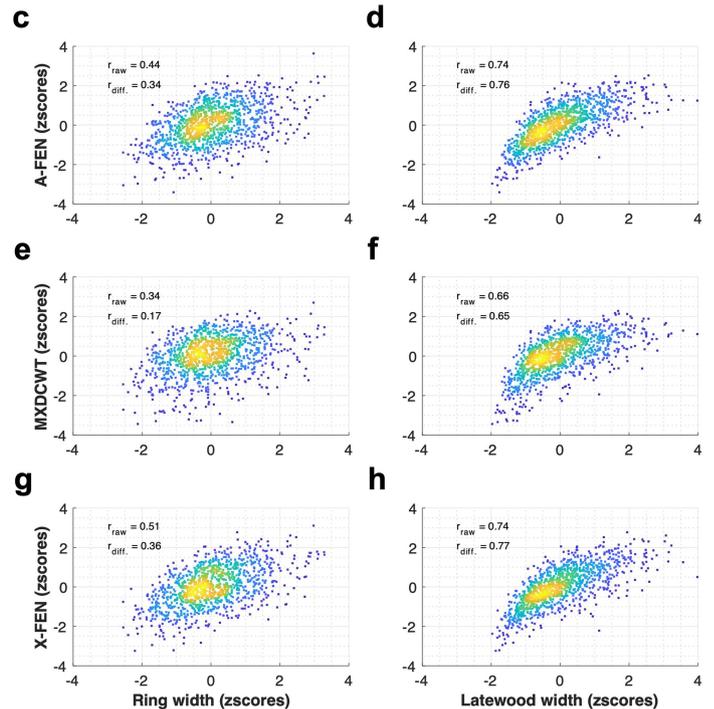


**Extended Data Fig. 3 | Warm-season temperature reconstruction skill of A-FEN and X-FEN, as well as comparisons with existing large-scale reconstructions and regional climate simulations but using QWA data that has not been detrended using the RCS approach.** The non-QWA datatypes are identical to Fig. 2 of the Main manuscript and the vertical arrows have the exact positions and dimensions as in Fig. 2 for reference. **a)** A-FEN (produced in this study) calibrated using regional mean air MJJA temperatures<sup>68</sup> ( $R^2$  ensemble range within brackets ( $\alpha = 0.05$ )), and results for the X-FEN (from Wilson, et al.<sup>9</sup>) using corresponding JJA temperatures. The irregular winter/spring of 1902/1903, led to a massive dieback of yearly branch-shoots in the region<sup>55</sup>, highlighted by

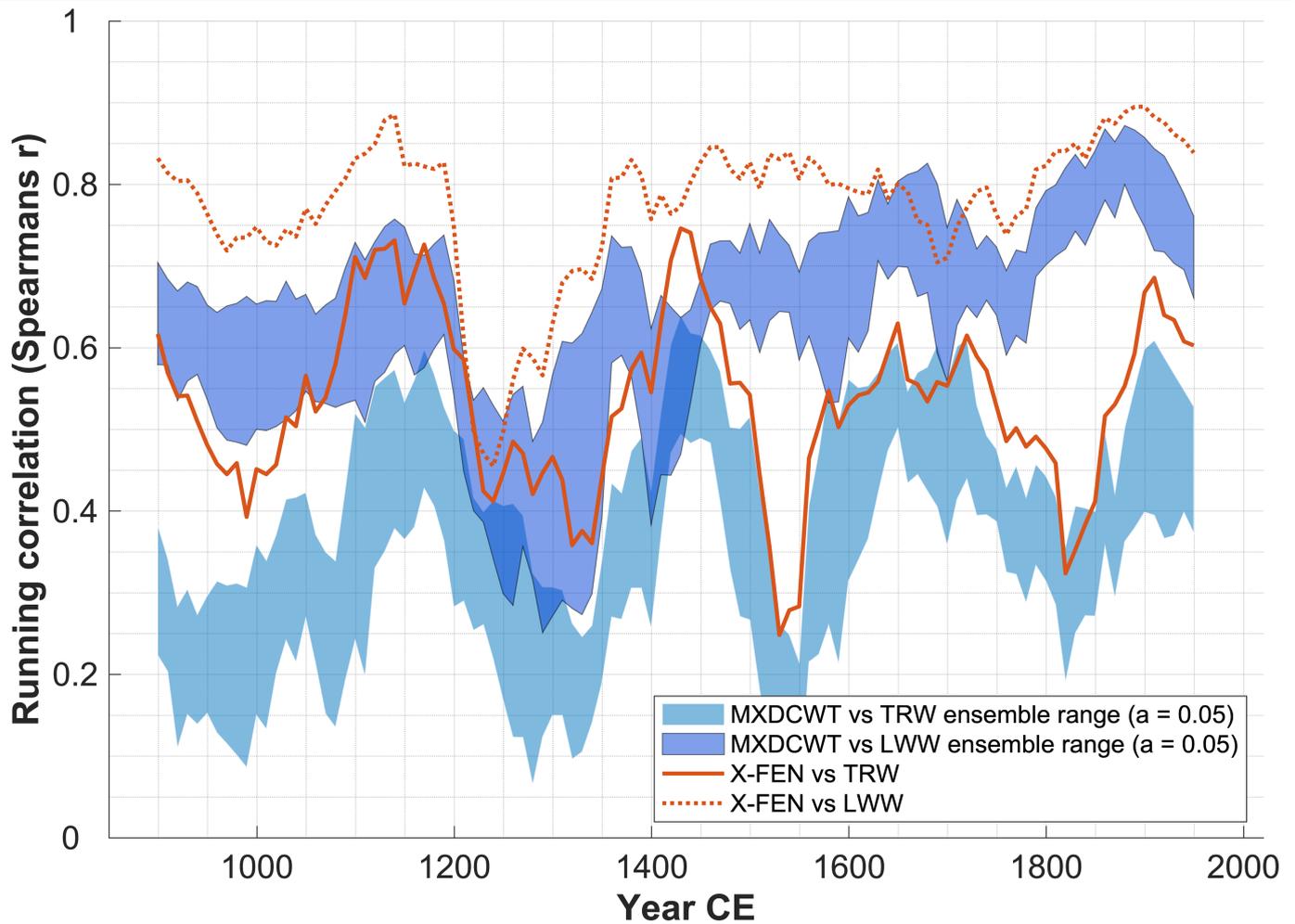
the yellow area. In these years with extremely narrow rings, the X-ray technique struggles to measure high MXD values due to its comparatively lower effective measurement resolution<sup>29</sup> (see Extended Data Fig. 4). **b)** Replication and pairwise inter-series correlation ( $\bar{R}$ ) of A-FEN in blue and the X-FEN in red. **c)** Centennial-scale variations (see Methods) compared between A-FEN, X-FEN, climate model simulations, and NH and global temperature reconstructions. The five large-scale reconstructions<sup>1,9,10,38,39</sup>, as well as the eleven regionally extracted climate-model simulations<sup>40-50</sup> are represented by probabilistic percentile ranges. The vertical arrows highlight the overall discrepancies of the X-FEN compared to the other data.



**Extended Data Fig. 4 | Illustration of the issue with comparatively low measurement resolution for X-ray MXD.** **a)** X-ray image with analysis track path indicated within the solid white rectangle, and examples of the effect of different effective measurement resolutions. **b)** The photosensors in **a)** build up measurement profiles, where the blue sensor builds the blue profile corresponding to a 20-micron effective measurement resolution, and the orange sensor builds up the orange profile corresponding to a 60 micron effective measurement resolution, approximating the effective measurement resolution of the X-ray methodology<sup>29</sup>. Note how the time series of MXD reflect inverse variations if developed using high-resolution or low-resolution equipment, i.e., the middle ring exhibits the lowest or highest value depending on resolution. The explanation for this is that very narrow latewood widths are

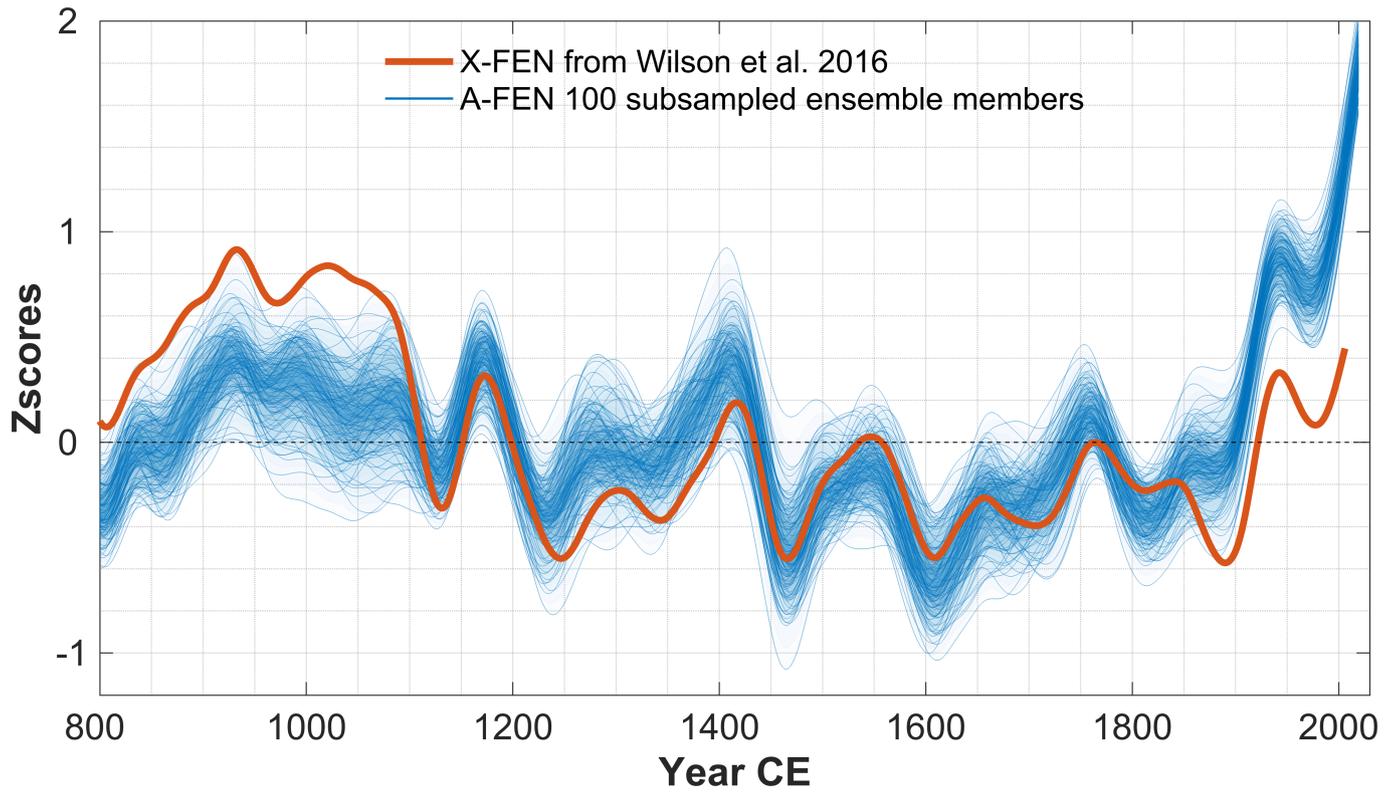


associated with comparatively lower MXD values even though the “true” MXD value may be high. **c)** Relationships between TRW and A-FEN and **d)** LWW and A-FEN. **e)** Relationships between TRW and anatomical MXD (MXDCWT) and **f)** LWW and MXDCWT. **g)** Relationships between TRW and X-FEN and **h)** LWW and X-FEN. All datasets display correlations and using datapoints covering 850–2005 CE. Note how the X-FEN always is stronger correlated with TRW and LWW than the MXDCWT. A higher correlation is expected if TRW or LWW is affecting the measurement. Spearman rank correlation coefficients were used due to the possibly non-linear relationships between width and density.  $R_{raw}$  and  $r_{diff}$  refers to untreated and first differenced data prior to correlations, respectively.



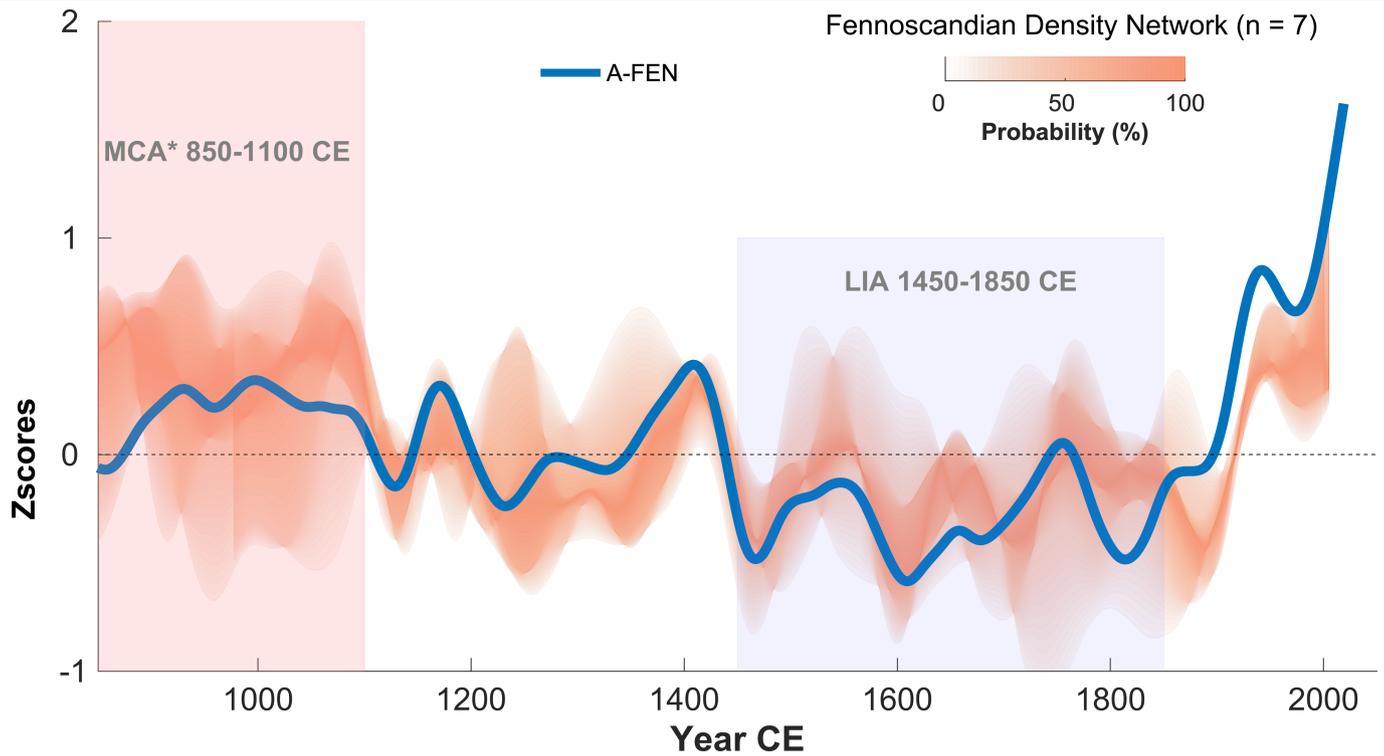
**Extended Data Fig. 5 | Moving window correlation coefficients revealing that X-ray MXD exhibits a stronger relationship with ring width and latewood width than anatomical MXD.** Ring width (TRW) versus anatomical MXD (MXDCWT) and X-FEN, as well as latewood width (LWW) versus MXDCWT and X-FEN. Spearman rank correlations were used on RCS-detrended chronologies with a 100-year base-lengths and 10-year overlaps. For the anatomical MXDCWT data, 100 sub-sampled chronologies with 15 trees/year were used to create

ensemble ranges represented in blue shades. Deviations from these blue shaded areas represent significant differences ( $p < 0.05$ ) from the TRW and LWW correlations with MXDCWT, respectively. The X-FEN correlations often reside outside the blue areas, and at higher correlations with TRW and LWW respectively, indicating occasionally stronger dependence of X-FEN on these parameters.



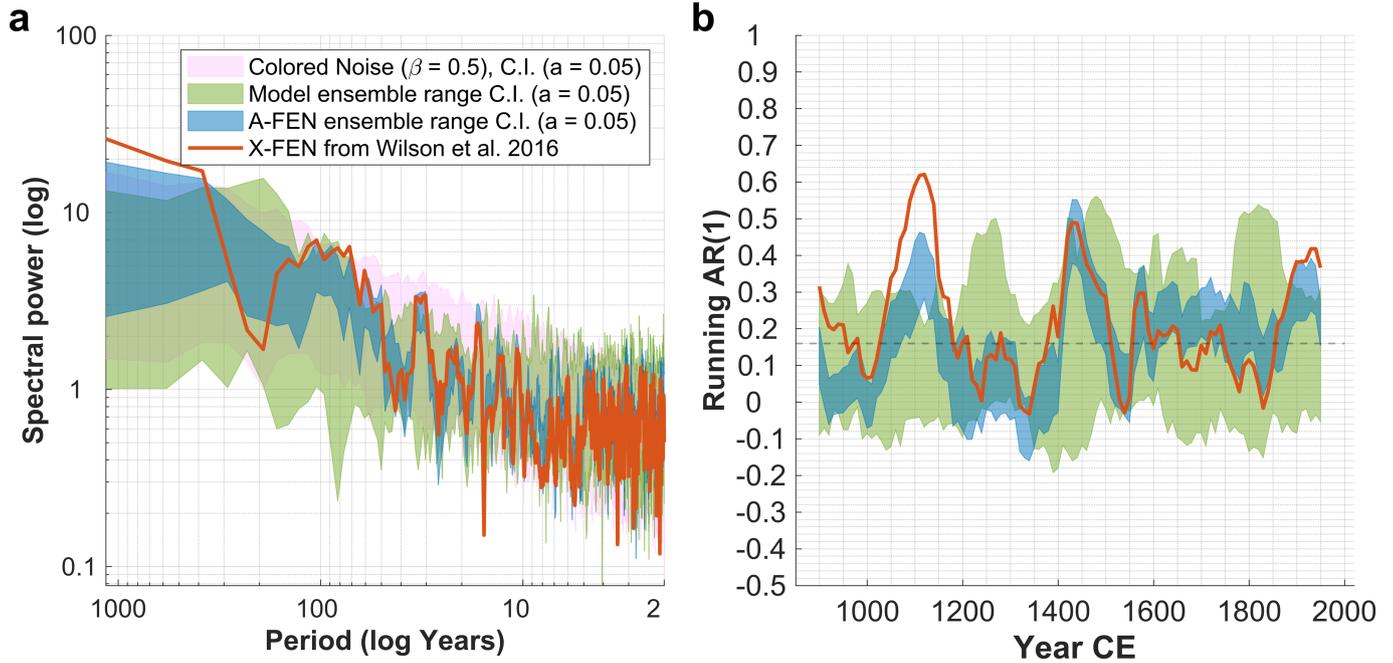
**Extended Data Fig. 6 | Comparison of A-FENs complete set of ensemble members with X-FEN.** RCS-detrended A-FEN (data from this study) versus the X-FEN (data from Wilson, et al.<sup>9</sup>), smoothed using cubic smoothing splines with

50% frequency response cut-off at 100 years. Note that no A-FEN ensemble member exhibit the protracted warmth during the MCA and the relatively low temperatures during the CWP, as does the X-FEN.



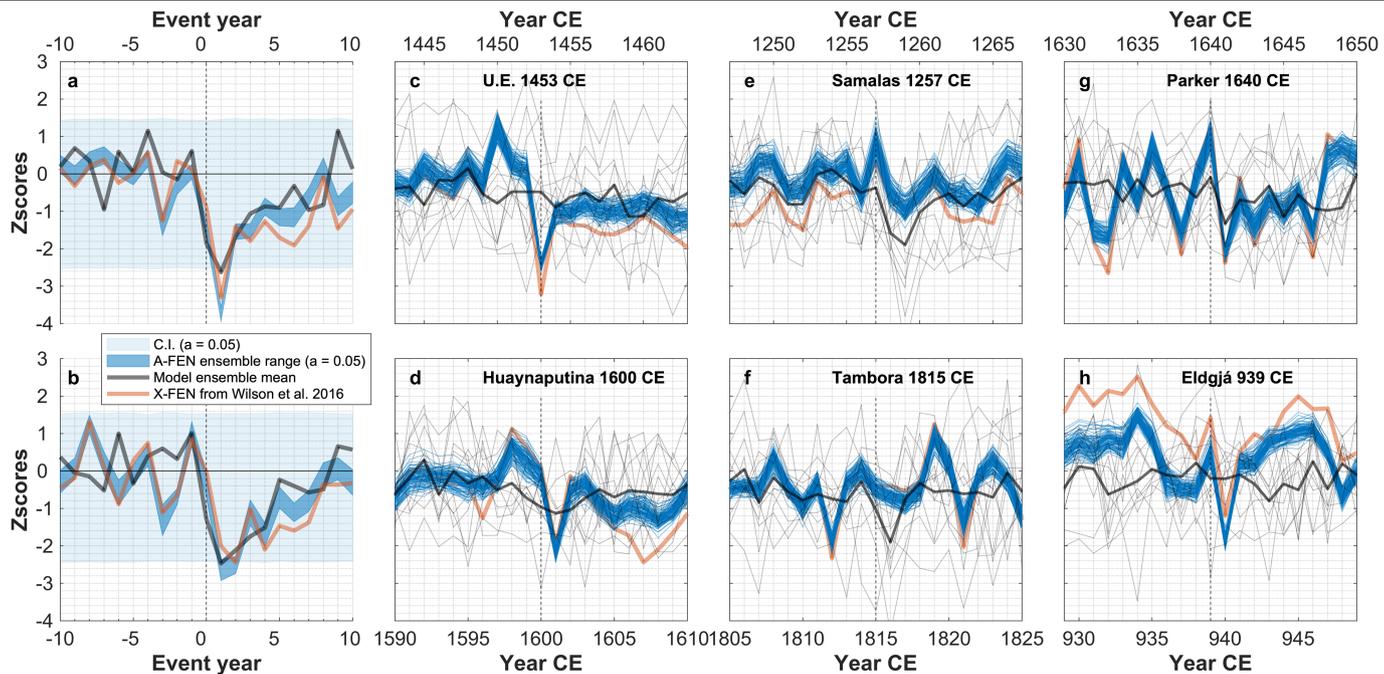
**Extended Data Fig. 7 | Comparison of A-FEN versus a network of millennium-long Fennoscandian MXD datasets showing the wide range of medieval estimates and comparably modest modern warming.** The MXD based

temperature reconstructions from Fennoscandia are retrieved from Wilson, et al.<sup>9</sup>, Schneider, et al.<sup>10</sup> and McCarroll, et al.<sup>61</sup> represented by a probabilistic percentile range.



**Extended Data Fig. 8 | Spectral properties and first order autocorrelation of the reconstructions and models compared in this study. a)** Spectral properties of the A-FEN ensemble and X-FEN on the backdrop of the model ensemble range, as well as the range of a 1000 timeseries, of equal length to the

A-FEN, of colored noise with a beta coefficient of 0.5. (Beta coefficient for White noise = 0, Pink noise = 1). **b)** Running autocorrelations AR(1) calculated for 100-year window lengths, shifted by 10-year lags.



**Extended Data Fig. 9 | Superposed epoch analysis and comparisons over individual eruption years for the A-FEN ensemble, X-FEN and the model ensemble.** SEA's using Gao, et al.<sup>86</sup> event lists of the 10 **a)** and 30 **b)** of the largest (based on sulfate aerosol injection) northern Hemisphere events. The model simulations were all extracted from the corresponding grid cells Lat 65–70° N, Lon 15–30° E. We used only Gao et al as basis for the event lists because most models in our ensemble were forced with Gao et al, but note that this list may

not be optimal for some models and the tree-ring data. We employed a model ensemble mean in the SEA, to explore the degree of volcanic cooling the models express. **c-h)** Proxy vs model response to some specific major volcanic events dated according to Toohey and Sigl<sup>87</sup>. The responses to U.E. 1453 CE, Huaynaputina and Eldgjá are pronounced in the proxy data but not in the models. The responses to Samalas and Tambora are pronounced in the models but not in the proxy data. The response to Parker is present in both models and proxy.

**Extended Data Table 1 | Reconstruction statistics for Fennoscandian tree-ring anatomy A-FEN**

<b>Calibration periods</b>	<b>1960–2019</b>	<b>1900–1959</b>	<b>1850-1899</b>	<b>1850-2019</b>
Explained variance, $R^2$ adjusted	0.74	0.72	0.81	0.77
No. of observations, (d.f.)	60 (58)	60 (58)	50 (48)	170 (168)
<b>First validation period</b>	<b>1900–1959</b>	<b>1960–2019</b>	<b>1900–1959</b>	
Reduction of error (RE)	0.68	0.77	0.72	
Coefficient of efficiency (CE)	0.72	0.73	0.71	
<b>Second validation period</b>	<b>1850-1899</b>	<b>1850-1899</b>	<b>1960–2019</b>	
Reduction of error (RE)	0.78	0.77	0.69	
Coefficient of efficiency (CE)	0.81	0.77	0.64	

Reconstruction statistics were calculated using May-August from HadCRUT5<sup>68</sup> (5° gridded monthly dataset, Lat. 65–70° N, Lon. 15–30° E). The reconstruction statistics were calculated over three split calibration/validation periods.