A 1150-year-long AMV reconstruction suggests early warning for a North Atlantic climate tipping point

Simon Michel (simon.michel@hotmail.fr)
Environnements et Paléoenvironnements Océaniques et Continentaux
Didier Swingedouw
EPOC, Universite Bordeaux https://orcid.org/0000-0002-0583-0850
Juliette Mignot
Sorbonne université (LOCEAN) https://orcid.org/0000-0002-4894-898X
Guillaume Gastineau
Sorbonne université (LOCEAN)
Pablo Ortega
Universitat Politecnica de Catalunya (BSC)
Myriam Khodri
Sorbonne université (LOCEAN)
Gerard McCarthy
Maynooth University https://orcid.org/0000-0002-2363-0561

Article

Keywords: Atlantic Multidecadal Variability (AMV), climate change, ecosystems

Posted Date: June 7th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-588180/v1

Version of Record: A version of this preprint was published at Nature Communications on September 2nd, 2022. See the published version at https://doi.org/10.1038/s41467-022-32704-3.

A 1150-year-long AMV reconstruction suggests early warning for a North Atlantic climate tipping point

Simon Michel[1,5]*, Didier Swingedouw[1], Juliette Mignot[2], Guillaume Gastineau[2], Pablo Ortega[3], Myriam Khodri[2], Gerard McCarthy[4]

[1] EPOC, Université de Bordeaux, Allée Geoffroy Saint-Hilaire, Pessac 33615, France.

[2] LOCEAN, Sorbonne université (UPMC, Univ. Paris 06)-CNRS-IRD-MNHN, 4 place Jussieu, 75005 Paris, France

[3] Barcelona Supercomputing Center (BSC-CNS), Edificio NEXUS I, Campus Nord UPC, Grand Capitán, 2-4, 08034, Barcelona, Spain.

[4] Irish Climate Analysis and Research UnitS (ICARUS), Department of Geography, Maynooth University, Maynooth, Ireland.

[5] Institute of Marine and Atmospheric research Utrecht (IMAU), Department of Physics, Utrecht University, Utrecht, Netherlands.

*s.l.l.michel@uu.nl

The Atlantic Multidecadal Variability (AMV) is a large-scale climate phenomenon with crucial impacts on human societies and ecosystems. Its periodicity and drivers are controversial due to the short temporal extent of instrumental observations and competing impacts of external forcing and internal variability. Here, we use a well-verified set of paleoclimate proxy records and compare four regression methods to perform different reconstructions of the AMV since 850 C.E., built to only reflect North Atlantic internal

variability. The best performing reconstruction, when verified both against climate model outputs and independent proxy records is obtained using the non-linear random forest method. It exhibits large multi-decadal variations in the range of 20-90 years, a broader range than the 50-70 years identified in instrumental records. The reconstruction shows that AMV autocorrelation properties have experienced significant changes in the recent decades, suggesting an early warning signal for the proximity of a tipping point in the Atlantic.

Introduction:

Since the beginning of the 20th century, the North Atlantic region has exhibited successive decades of anomalously warm and cold sea surface temperatures[1] (SST) relative to the global average, subsequently contributing to amplify or damp the global warming effects in the Atlantic sector[2,3]. The underlying variability mode, the Atlantic Multidecadal Variability (AMV), is related to diverse climatic effects in the North Atlantic neighboring regions[4,5] (Fig. 1). At the global scale, it also influences drought and rainfall in the Sahel[6,7], Northeastern Brazil[7] and Central Asia[7] (Fig. 1); Atlantic hurricane frequency and intensity[8,9]; sea ice thickness and extent over the Arctic[10]; and is linked to Pacific climate variability[11]. The processes driving the AMV and its resulting spectral properties remain a source of controversy. Disagreements partly come from the relatively short period over which the AMV is directly observed, which mainly encompasses climate changes significantly affected by human activity[12]. Internal ocean variability has been identified in climate models as a driver of the AMV pattern through variations in the Atlantic Meridional Overturning Circulation[13,14] (AMOC), a large-scale circulation that is known to be potentially unstable in the long-term projections of future climate[15,16]. These results, based on coupled ocean-atmosphere general circulation models,

48 have been further supported by a tide gauge-based reconstruction of the ocean circulation
49 intensity in the North Atlantic intergyre region that exhibits strong correlations with the AMV
50 index over the last 60 years[17]. Several long paleoclimate records also indicate the existence
51 of such multidecadal variability on longer time scales[18,19]. Nevertheless, it has also been
52 shown that many features of the observed AMV can be reproduced in slab-ocean models by
53 the integration in the ocean of the stochastic forcing of the North Atlantic Oscillation, thus
54 without any influence of ocean dynamics[20]. It has also been highlighted that periods of cold
55 AMV are dominated by both strong anthropogenic aerosol emissions from Europe and North
56 America and strong volcanic activity[21]. Notably, a recent study[22] has shown that no consistent
57 multidecadal variability in the global mean surface temperature (GMST) is found in control
58 simulations of 16 models contributing to the Climate Model Intercomparison Project phase 6
59 (CMIP6), as compared to last millennium experiments including estimations of volcanic and
60 solar forcing as boundary conditions. This study argues that the incoming solar radiation has
61 varied in a 50 to 70 years band due to volcanic aerosols during the last millennium. Their
62 conclusion is that these spectral peaks in last millennium natural forcings can solely explain
63 the simulated basin-wide multidecadal variability of North Atlantic SSTs. However,
64 multidecadal variability is believed to be inherent to the North Atlantic, while this former
65 study focused on the GMST, which can also include important contributions from interannual
66 variability modes, like El-Niño Southern Oscillation. In addition, GCMs might be overly
67 sensitive to volcanic forcing, as in proxy-based climate reconstructions of the largest
68 eruptions are generally associated with weaker temperature responses than in the models[23].
69 Moreover, a study of the AMV over the last 8000 years, investigating a set of ice core and
70 marine records, has shown that its 55-70 years timescale of variability were largely driven by
71 internal ocean-atmosphere variability and poorly influenced by natural forcings for a large

part of the Holocene[24]. The coming years of AMV observations could be very informative in this respect, as recent observations indicate the likely onset of a negative phase[25]. The study of this potential transition to a new negative phase of the AMV could be very fruitful in understanding the relative roles of natural forcings and internal variability behind its fluctuations.

In terms of spectral properties, there are notable discrepancies between the observed periodicity in AMV (50-70 years) and the periodicity found in most of the preindustrial control experiments contributing to the CMIP5 (10-30 years)[26]. Disentangling the relative contributions of the AMV driving factors to its spatial and spectral properties in observations could therefore be useful in assessing models and eventually developing emergent constraint approaches to reduce uncertainty in decadal climate predictions[27]. To reach this long-term goal, an improved knowledge of internal variability of North Atlantic Sea Surface Temperature (NASST) is crucial.

The AMV index definition is of paramount importance to tackle correctly the lingering questions concerning its drivers. Since climate is currently in a period with a dominant anthropogenically-induced warming signal, several methods are used to isolate the internal variability inherent to the North Atlantic basin. In this study, we use three yearly AMV indices in which all the externally forced signals (including anthropogenic forcing, solar variations, and volcanic eruptions) in annual NASST have been removed with different techniques, to thus explore the sensitivity to the method. We denote these indices as AMV_{TS}[8], AMV_{T}[2] and AMV_{F}[28] (see Methods and Fig. 1a), where the subscripts TS, T, and F refer to the associated references[2,8,28]. They all have the advantage of removing a large part of the externally forced variability from greenhouse gases and aerosols (see Methods) as opposed to detrending methods[9,29], which only remove the effect of anthropogenic greenhouse gases emissions[28]. A

96 regression analysis between the mean of the three AMV indices with surface temperatures
97 and precipitation observations from the instrumental dataset CRUTS4[30], shows significant and
98 widespread links of the AMV with both variables, in particular over the continents (Fig. 1b-k),
99 consistently with early studies[4-7].

100 The last millennium is well-covered in space by high-resolution proxy records
101 describing local or regional precipitation or temperature variations, making it very suitable
102 for producing annually-resolved paleoclimate reconstructions. It is of particular interest since
103 it encompasses two contrasting periods in terms of global temperature: the warm Medieval
104 Climate Anomaly and the cold Little Ice Age. The relative contribution of external forcing and
105 internal climate variability to the transition between both periods is unclear, as well as their
106 temporal evolution and spatial extent[31-33]. Taking advantage of paleoclimate data covering
107 the last millennium, there have been some attempts to reconstruct AMV indices. Previous
108 studies[18,34] usually reconstructed AMV over the last millennium using a Principal Components
109 Regression (PCR) method and proxy records from continental borders of the North Atlantic.
110 However, their statistical models were calibrated using NASST anomalies[18,34], which include a
111 strong forced signal over the historical period. Indeed, the NASST variability during the last
112 millennium might have different links with global climate, as internal variability could be more
113 dominant, which could affect the resulting reconstruction. In addition, it is worth noting that
114 building a statistical model in which the predictand has a strong trend is not recommended in
115 our case because trends in the predictors can easily match by chance the one in the
116 predictand, leading to the selection of spurious predictors that compromise the quality of the
117 reconstruction. These former reconstructions of the AMV are furthermore based on a single
118 regression method[18,34] (PCR, *cf.* Methods), and do not include any objective criterion for the
119 selection of proxy records, meaning that some of them may have potentially little correlation

with the target AMV index over the instrumental period. They also do not include any pseudo-proxy validation of the reconstruction as a means to verify that the assumed climatic relationships between the AMV and the proxies are sound. To circumvent most of the previous limitations the present study combines multiple proxies with different regression methods to reconstruct the AMV for the past 1150 years, using also two validation approaches to evaluate the robustness of the reconstruction, the first based on the comparison with independent ocean proxy records and the second on the use of pseudo-proxy experiments.

Results:

An AMV reconstruction based on optimal regression model approach

The availability of large numbers of annually-resolved proxy records sensitive to temperature and precipitation provides a key opportunity to test and develop new objective reconstruction methodologies using advanced techniques. The different reconstructions produced in this study use proxy records significantly correlated at the 95% confidence level with the target AMV indices (cf. Fig. 1). The selection of these proxies is made from a large set of Northern Hemisphere and annually-resolved proxy records that includes the PAGES 2k database[35] and 41 other records published elsewhere (Extended Data Table 1). This database (hereafter P2k+) has been constructed using different quality criteria (see Methods) and comprises a total of 457 records. An important preprocessing step before selecting the proxies is to remove in each proxy record from P2k+ (see Methods) the forced variability of NASST, using linear regression starting in 1870. This is a novel step that allows us to produce an AMV reconstruction that exclusively reflects the internal variability signals. The forced component is estimated using a signal-to-noise maximizing empirical orthogonal function

144 technique[2] (Methods, Extended Data Fig. 1) from historical simulations of 37 CMIP5 models
145 (Extended Data Table 2).

146 Reconstruction methods applied in this study require that the predictors (*i.e.,* the
147 proxy records) have no temporal gaps. To use the maximum number of proxies available to
148 reconstruct each temporal window we follow a nested reconstruction approach[34] with
149 moving windows of one year. We set the reconstructions to cover the period 850 C.E. to
150 present. This will allow us to estimate the accuracy and reliability of the method through the
151 use of pseudo-proxy experiments[37] (PPE).

152 The specific methodology to be used for the final nested reconstruction will be
153 determined in a preliminary phase in which a total of 312 reconstructions covering the whole
154 temporal extension (back to 850 C.E.) will be produced, by combining the three AMV
155 definitions, 26 temporal windows (only differing in the last year covered by the proxies) and
156 four regression methods[37]: PCR, partial least squares, elastic-net and random forest (detailed
157 in Supplementary Information). We use the ClimIndRec reconstruction toolbox[38] to generate
158 these reconstructions (84 for each AMV index, *cf.* Methods). This tool is dedicated to quickly
159 produce reconstructions with up to four regression techniques, where each of their specific
160 control parameters are automatically optimized with cross-validation, and the reliability of
161 the final reconstruction is directly estimated using the Coefficient of Efficiency (CE) metric[39]
162 (Methods) over training and testing samples. The CE metric is defined between $-\infty$ and 1. As
163 the CE is positive for a given testing sample, it indicates that the statistical model gives better
164 estimations than the empirical average of the corresponding training sample, and its use for
165 reconstruction purposes can be considered[39]. In this study we set that the 312 statistical
166 models are preliminarily evaluated with the CE metric for 30 pairs of training and testing
167 samples.

CE scores of the 312 reconstructions for each level of inputs (Methods, Extended Data Fig. 2) show that the random forest[40] (RF), despite not necessarily being the best method for all the reconstructions produced, does provide the highest averaged score of all reconstructions. The latter is obtained for the AMV_F index, for the reconstruction period 850-1987. For this specific reconstruction, the average of CE scores over the 30 training/testing splits is positive at the 99% confidence level, which validates its use for reconstruction purposes ($CE \in [-0.07, 0.52], med(CE) = 0.25, avg(CE) = 0.23$, *cf.* Methods).

Using the previous setup yielding the best CE, we perform 1020 nested reconstructions of the AMV_F index based on RF models, applying one-year increments to the initial year of the reconstruction, thus going from the longest time window (*i.e.*, 850-1987) to the shortest (*i.e.*, 1869-1987). For each of them, only proxy records covering the whole window and significantly correlated with the observed index at the 95% confidence level are used. The final nested reconstruction is thus obtained by averaging the 1020 interlocking reconstructions over their common timesteps (Fig. 2a). Evaluating each of these reconstructions with CE scores for 30 training/testing splits, we find that the reconstruction has varying validation scores over time, although they overall slightly increase with generally higher scores for the shortest (most recent) reconstructions which are based on higher numbers of proxy records (Fig. 2b).

The nested reconstruction uses a total of 55 Northern Hemisphere proxy records. Their *a posteriori* weights, given by random forest importance (*cf*. Methods), and their temporal availability, are presented in Fig. 2c-d. We identify three main clusters of records with fairly distributed weights: central Asia, Europe and western North America. It is worth noting that proxies from Asia and western North America are highly represented in the PAGES 2k database[35], which basically explains their relatively large number used for the

192 reconstruction. Interestingly, there are a number of highly weighted proxy records of annual
193 and boreal summer (June-July-August) temperatures over the Eastern Pakistan/Tibetan
194 Plateau. This link has been supported in a couple of recent studies which have highlighted the
195 role of AMV variations for spring and summer temperatures in this region[41,42], notably by
196 affecting large scale pressure gradients in the Eurasian sector. We also use a large number of
197 proxy records from eastern Asia/northern China, for which climate conditions have also been
198 shown to be significantly affected by the AMV variations through atmospheric Rossby wave
199 propagation and altered heat advection in the western Pacific[43]. Given the large-scale AMV
200 teleconnections shown by these studies[41-43] and the large weights we find for Asian proxy
201 records, we justify their inclusion in our new AMV reconstruction, the first one to date to
202 include them[18,34]. Western North American proxy records are mostly sensitive to summer and
203 yearly variations of temperature and precipitation (Fig. 2c-d, Extended Data Table 3). It
204 appears that Fig. 1 does not show a consistent relationship between the AMV and
205 instrumental summer and annual temperatures in this region, although it is observed for
206 some precipitation time series over the historical period (Fig. 1). Finally, the fact that only five
207 proxy records from Europe are used is mostly due to their relatively reduced presence in the
208 proxy record database we use (<10%). However, we find that one of them has a relatively
209 large weight for the reconstruction (>7%) and covers the entire reconstruction period. It
210 corresponds to a time series of tree rings growth measurements from European Alps[44], which
211 is strongly correlated with summer temperature over the historical period ($r=0.7$, $p<0.01$,
212 Extended Data Table 3). The four other European proxy records are related to either summer
213 or annual temperature and precipitation. The selection of these records for our
214 reconstruction is thus highly consistent with the well-documented fingerprint of the AMV on
215 European summer temperatures[4,5], as also shown in Fig. 1. A detailed description of each

proxy record used for our reconstruction and their correlations with instrumental data and AMV is given in Extended Data Table 3. Additionally, the same maps of the AMV fingerprints from Fig. 1 including the corresponding selected proxy records for various seasons and climate variables are shown in Extended Data Fig. 3.

Data and model-based validations

As for the previous reconstructions[18,34], most of the proxy records used in this study are terrestrial, which might be surprising given that the AMV is an oceanic mode. Our method indeed selected a poor amount of ocean proxies from P2k+ (1 out of 55), essentially because those with a resolution lower than annual have been excluded (Methods). To evaluate if our AMV reconstruction is in line with existing ocean proxy records, we consider *a posteriori* those not used in the reconstruction. Since some of these proxies have very low temporal resolution, which might lead to spurious significant correlations with the AMV, we use a significance test that takes the time series autocorrelations into account, similarly to previous studies[17,38] (see Methods). We find that 37 ocean records (23 from the North Atlantic including the Mediterranean Sea) from the Ocean 2k database[35] are significantly correlated at least at the 90% confidence level with our AMV reconstruction (Fig. 3a). As an additional validation, we compute two composite time series of these coral and sediment-based proxy data (see methods). The first one is an average-based aggregate of the 37 ocean proxies from Fig. 3a, and the second one is only based on the 23 North Atlantic proxies from Fig. 3a (Fig. 3b). Since these composites are mostly based on low resolution sediment data, they do not capture well the annual or decadal climate variations as our reconstruction does. However, in terms of low-frequency, we find strong and significant correlations between the 30-year filtered AMV reconstruction (mostly based on terrestrial records as explained above, Fig. 2a-

d), and these composite time series based on all ocean proxies from Fig. 3a ($r=0.75$, $p<0.01$, Fig. 3b) and only those from the North Atlantic ($r=0.76$ $p<0.01$, Fig. 3b).

A complementary validation of the physical consistency of the reconstruction is performed with PPE[37], using 12 members from the Last Millennium Large Ensemble of the Community Earth System Model 1 (CESM1-LME, see Methods). Here, we conduct two types of PPE experiments for each CESM1-LME member (see Methods for detailed explanations).

The first PPE experiment consists in exactly reproducing the reconstruction from the real-world experiment (RWE) in the model simulations by only training statistical models over the same period (1870-1987), with the timeseries of the nearest grid points to the real-world proxy record for each time frame of the nested reconstruction. The validation then consists in comparing the reconstructed last millennium AMV in the model using the RWE methodological setup (see previous section) and the AMV_F effectively simulated within the model (hereafter, the model AMV). For each member, we calculate CE scores, as well as the correlation between the model AMV and the AMV reconstructed from pseudo-proxies (Fig. 4a). All correlations between the reconstructed AMV and the corresponding model AMV are significant at least at the 90% confidence level for the 10-year smoothed time series ($r \in [0.41,0.57]$ for the 12 members). In terms of skill scores we find that the real-world median skill score ($med(CE)=0.29$ for the whole nested reconstruction) falls within the range of those from the PPE, which are significantly positive at the 95% confidence for all the members except number 7 ($med(CE) \in [0,0.6]$). This first PPE validation, based on an ensemble of 12 last millennium simulations, therefore provides further confidence in our reconstruction and constitutes the first model-based validation for an AMV reconstruction.

The second experiment is adding constraints from the model for the reconstruction of the AMV using real proxy records. It consists in first subselecting pseudo-proxies according to the correlation between the model AMV and the 55 pseudo-proxies. The RF models are trained within the model simulations over the longest period covered by each proxy in the RWE, and then applied to the real-world proxies (see methods). This method is similar to the PPE performed in a previous NAO reconstruction study based on the PCR method[45]. The produced AMV reconstructions, based on AMV/pseudo-proxies relationships over long time frames within simulation members, are then compared to the AMV reconstruction from the RWE. The RF models trained over LME simulations applied to the values of the real proxy records lead to very similar reconstructions than those given by the RWE with a significance of at least 95% for the 12 members ($r \in [0.53,0.88]$, Fig. 5a). CE skill scores are also significantly positive in the 12 members for this PPE ($med(CE) \in [0.2,0.35]$, $p<0.01$ for the 12 members, Fig. 5a), which indicates a satisfactory level of robustness for the RF models trained in the CESM1-LME members. The correlation of the ensemble mean of model-based AMV reconstructions with the RWE reconstruction is also highly significant ($r=0.88$, $p<0.01$). This second PPE indicates that training RF models within the millennial-long simulations of the CESM1-LME, with respect to proxies temporal availability in the RWE (Fig 2d, see Methods), reproduces a very similar reconstruction as the RWE when they are applied to the real values of proxy records.

Another important aspect to highlight, that supports the validity of the reconstruction, is the fact that the network of proxy records used in the RWE reconstruction has similar weights in both PPE experiments: large weights are not restricted to North Atlantic bordering regions, they also occur in central to eastern Asia and western North America, in agreement with the teleconnections highlighted in numerous studies[5,41-43]

286

287 **Unforced multidecadal variability and ongoing bifurcation**

288 As a complement to the AMV reconstruction, which has been used as an index definition that
289 focuses on the internal variability, we have also performed a reconstruction of NASST, which
290 does include the influence of the external forcings, as done in previous reconstruction
291 studies[18,34]. Such an additional reconstruction might help to disentangle the role of internal
292 variability from total variability recorded in the proxy records and in the reconstruction of
293 variations in the North Atlantic. For doing so, we use the same statistical model selection and
294 the same nested approach than for the AMV to reconstruct the NASST back to 850 C.E. as
295 well, using proxy records in which the externally forced signal has not been removed. The
296 best reconstruction of NASST is obtained with the PCR approach. It exhibits a significant
297 correlation with the AMV reconstruction ($r=0.64$; $p<0.01$, Fig. 6a). This suggests that more
298 than 40% of the variability of NASST over the last millennium can be explained by internal
299 variability only, the rest being related to external forcing. Strikingly, validation scores
300 obtained by this optimal regression approach are much higher than for the AMV
301 reconstruction (med(CE)=0.44 and med(CE)=0.25, respectively). A plausible explanation for
302 these discrepancies in skill scores between NASST and AMV reconstructions is that the way
303 the AMV is constructed accompanied with external forcing removal from proxy records might
304 in turn partly decorrelate NASST and proxies by removing their common responses to the
305 same forcings.

306 Using a recent reconstruction of volcanic activity[46], we further perform a superposed
307 epoch analysis[47] (Methods) on both the NASST and the AMV reconstructions, to characterize
308 the response to the 10 largest eruptions of the last millennium (Extended Data Table 4). While
309 the reconstructed NASST has a similar response a decade after the eruption than a previous

reconstruction also based on NASST[34], no significant response is found for the AMV reconstruction (Fig. 6). This is partly expected given that we have reconstructed an index of the internal climate variability. When looking at individual responses, there are eruptions where slightly significant negative AMV responses are found, although they are not the strongest of the last millennium (1171, 1601, Extended Data Fig. 4). These responses might thus not be a direct signal forced by the eruption. Furthermore, no significant cooling is observed for any of the second and third strongest eruptions, respectively in 1815 (Tambora) and 1453 (Kuwae). There is even a large annual positive peak of temperature that is found for the strongest eruption in 1257 (Samalas), while no difference is found in the AMV state before and after it (Extended Data Fig. 4).

Regarding solar forcing, neither the 10-year or 30-year filtered time series from the PMIP3 TSI reconstruction[48] is significantly correlated with our 10-year filtered AMV reconstruction, even when solar forcing leads by a few years ($r=0.23$, $p>0.2$, lag=12; $r=0.32$, $p>0.2$, lag=13; respectively, Extended Data Fig. 5). Both the 10-year and the 30-year filtered time series of the TSI reconstruction a modestly significantly correlated with the NASST reconstruction ($r=0.5$, $0.1<p<0.2$, lag=13; $r=0.52$, $0.1<p<0.2$, lag=14; respectively). This result therefore further supports the idea that our reconstruction represents only internal variability of the North Atlantic and can be thus considered as an additional independent validation of our reconstruction.

The wavelet analysis in Fig. 7 shows that the AMV reconstruction exhibits important multidecadal variations (Fig. 7), contrary to what has been recently suggested for an ensemble of control simulations from 16 CMIP6 models[22]. Our reconstructed AMV primarily varies in the 20-90 years band except for the 1400-1800 period, which is more dominated by

shorter (20-40 years) cycles (Fig. 7). Thus, the 50-70 years periodicity suggested from observations since 1850 may not be systematic, as suggested by the variability produced by climate models, in control simulation with fixed external forcings[49]. Our reconstruction of internal variability of the North Atlantic also contradicts a recent study suggesting that North Atlantic multi-decadal variability in the 50 to 70-year frequency band can be entirely explained through the pulses of volcanic activity during the last millennium[22]. These eruptions certainly contributed to the variability in the 50-70 years spectrum band but cannot explain the whole range of variations according to this AMV reconstruction.

The wavelet analysis hints at a recent increase in the overall spectral power of the AMV. The changes in spectral characteristic of a time series can be used as an early warning of regime shift in dynamical systems, as shown in numerous studies[50-52]. In particular, the AMOC is well-known as a potential tipping element of the climate system[53] and it has been shown that an AMOC regime shift in climate models might necessitate the knowledge of hundreds of years of time series of AMOC variations[50,52], which is prohibitive with direct observations of it at 26°N that only last for less than 20 years[54]. Here, the reconstruction of our AMV as an internal mode of variability can be related to the internal dynamics of the AMOC through its impact on heat transport and the AMV[13,14,55]. Thus, the change in spectral characteristics of the AMV might be seen as a potential early warning of a regime shift[50-52] in the ocean circulation. It is, however, difficult to tell if the overall increase in spectral power is robust as the values for the largest periodicities are outside of the cone of influence and therefore subject to edge effects. To test the hypothesis of the imminent occurrence of a potential tipping point, we use a similar approach as a former model-based study evaluating early warning signals for an AMOC critical slowdown[51]. Dynamically, this approach assumes that a given system is likely to be slowing down if its memory increases over time, *i.e.*, if the

system state at time t+1 gets more and more dependent on the system state at time t when approaching the bifurcation point. In terms of time series, the memory of the system can be measured using autoregressive AR1 coefficients, which are then assumed to be increasing when approaching a critical slowdown[50]. Thus, Kendall τ statistics are computed for the AR1 coefficients for different sliding window lengths[50-52] (from 200-year to 400-year with an increment of 50 years). Kendall τ quantifies the time-evolution of autocorrelations as a ranked correlation between the AMV sliding AR1 coefficients and time (see Methods). It indicates a highly significant increase in AMV memory over the recent period for the different window lengths tested ($p<0.01$ for all, Fig 7b, Methods). According to the tipping points detection theory[50], this constitutes the first observation-based estimate that the AMV may now be approaching a tipping point, after which the Atlantic current system might change its mean state.

These recent changes in autocorrelation properties of the AMV towards higher values can be related to a possible approach of a tipping point in the AMV. Such an AMV tipping might reflect changes in the AMOC, subpolar gyre or the Arctic circulation, which were previously reported to have tipping points in models[15,16,53]. In this respect, the significant critical slowdown test of the AMV reconstruction could be interpreted as a long-term relative cooling of the North Atlantic in the near-term future. This assumption based on our real-data reconstruction has previously been proposed using CMIP5 models, among which such an abrupt change happens in projections of nearly half of the best ones in representing ocean convection in the northern North Atlantic[15], even under scenarios with low anthropogenic emissions of greenhouse gases. It thus raises serious concerns, while more and more evidence seem to indicate an on-going long-term slowdown in the Atlantic current[56] and that the impacts of such a change are numerous[16,57,58]. This further highlights the need for an

appropriate account of the potential implications in terms of climate adaptation plan in case of rapid changes in the Atlantic[59].

406 **References**

407 1. Kushnir, Y. Interdecadal variations in North Atlantic Sea Surface Temperatures and
408 associated atmospheric conditions. *J. Clim*. **7**, 1-20 (1994).
409 2. Ting, M., Kushnir, Y., Saeger, R. & Cuihua, L. Forced and internal twentieth-century SST
410 trends in the North Atlantic. *J. Clim*. **22**, 1469-1481 (2009).
411 3. Mann, M. E., Steinman, B. A & Miller, S. K. On forced temperature changes, internal
412 variability, and the AMO. *Geophys. Res. Lett*. **41**, 3211-3219 (2014)
413 4. Sutton, R. T. & Dong, B. Atlantic Ocean influence on a shift in European climate in the
414 1990s. *Nat. Geosci*. **5**, 788-792 (2012).
415 5. Sutton, R. T. & Hodson, D. L. R. Atlantic Ocean forcing of North American and European
416 summer climate. *Science*. **309(5731)**, 115-118 (2005).
417 6. Defrance, D., Ramstein, G., Charbit, S., Vrac, M., Famine, A. M., Sultan, B.,
418 Swingedouw, D., Dumas, C., Gemenne, F., Alvarez-Solas, J. & Vanderlinden, J.-P.
419 Consequences of rapid ice sheet melting on the Sahelian population vulnerability.
420 *PNAS*. **114(25)**, 6533-6538 (2017).
421 7. Knight, J., Folland, C & Scaife, A. Climate impacts of the Atlantic Multidecadal
422 Oscillation. *Geophys. Res. Lett*. **33**, L17706 (2006).
423 8. Trenberth, K. & Shea, D. Atlantic Hurricanes and natural variability in 2005. *Geophys.*
424 *Res. Lett*. **33**, L12704 (2006).
425 9. Enfield, D. & Cid-Serrano, L. Secular and multidecadal warmings in the North Atlantic
426 and their relationships with major hurricane activity. *Int. J. Climatol*. **30**, 174-184
427 (2010).

10. Miles, M., Divine, D., Furevik, T., Jansen, E., Moros, M. & Ogilvie, A. E. J. A signal of persistent Atlantic Multidecadal variability in Arctic sea ice. *Geophys. Res. Lett*. **41**, 463-469 (2014).
11. Ruprich-Robert, Y., Msadek, R., Castruccio, F., Yeager, S., Delworth, T. & Danabasoglu, G. Assessing the climate impacts of the Observed Atlantic Multidecadal Variability Using the GFDL CM2.1 and NCAR CESM1 Global Coupled Models. *J. Clim*. **30(8)**, 2785-2810 (2017).
12. Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R. L., Verheggen, B., Maibach, E. W, Carlon, J. S., Lewandowsky, S., Skuce, A. G., Green, S. A., Nuccitelli, D., Jacobs, P., Richardson, M., Winkler, B., Painting, R. & Rice, K. Consensus on consensus: a synthesis of consensus estimates on human-caused global warming. *Environ. Res. Lett*. **11**, 048002 (2016).
13. Muir, L. C. & Fedorov, A. V. How the AMOC affects ocean temperatures on decadal to centennial timescales: the North Atlantic versus an interhemispheric seesaw. *Clim. Dynam*. **45(1-2)**, 151-160 (2015).
14. Yan., X., Zhang., R. & Knutson, T. Underestimated AMOC variability and implications for AMV and predictability in CMIP models. *Geophys. Res. Lett*. **45**, 4319-4328 (2018).
15. Sgubin, G., Swingedouw, D., Drijfhout, S., Mary, Y. & Bennabi, A. Abrupt cooling over the North Atlantic in modern climate models. *Nat. Commun.* **8**, 14375 (2017).
16. Collins, M., Sutherland, M., Bouwer, L., Cheong, S.-M., Frölicher, T., Jacot Des Combes, H., Koll Roxy, M., Losada, I., McInnes, K., Ratter, B., Rivera-Arriaga, E., Susanto, R. D., Swingedouw, D. & Tibig, L. [Pörtner, H.-O., Roberts, D. C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A.,

Petzold, J., Rama, B. & Weyer, N. M. (eds.)]. Extremes, Abrupt Changes and Managing Risk. *IPCC SROCC.* **Chapter 6** (2019).

17. McCarthy, G. D., Haigh, I. D., Hirshi, J. J.-M., Grist, J. P., Smeed, D. A. Ocean impact on decadal Atlantic climate variability revealed by sea-level observations. *Nature*. **521**, 508-512 (2015).
18. Gray, S., Graumlich, L., Betancourt, J. & Pederson, G. D. A tree-ring based reconstruction of the Atlantic Multidecadal Oscillation since 1567 A.D. *Geophys. Res. Lett*. **31**, L12205 (2004).
19. Mjell, T. R., Ninnemann, U. S., Eldevik, T. & Kleiven, H. K. F. Holocene multidecadal-to millennial-scale variations in Iceland-Scotland overflow and their relationship to climate. *Paleoc. and Paleoclimatol*. **30(5)**, 558-569 (2015).
20. Clement, A., Bellomo, K., Murphy, L. N., Cane, M. A., Mauritsen, T., Rädel, G. & Stevens, B. The Atlantic Multidecadal Oscillation without a role for ocean circulation. *Science*. **350(6250)**, 320-324 (2015).
21. Booth, B. B. B., Dunstone, N. J., Halloran, P. R., Andrews, T. & Bellouin, N. Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, **484**, 228-232 (2012).
22. Mann, M. M., Steinman, B. A., Brouillette, D. J. & Miller, S. K. Multidecadal climate oscillations during the past millennium driven by volcanic forcing. *Science*. **371**, 1014-1019 (2021).

Schneider, L., Smerdson, J. E., Pretis, F., Hartl-Meier & C. Esper, J. A new archive of large volcanic events over the past millennium derived from reconstructed summer temperatures. *Environ. Res. Lett.* **12**, 094005 (2017).

23. Knudsen, M. F., Seidenkrantz, M.-S., Jacobsen, B. H., Kuijpers, A. Tracking the Atlantic Multidecadal Oscillation through the last 8,000 years. *Nat. Commun.* **2**, 178 (2011).

476 24. Frajka-Williams, E., Beaulieu, C. & Duchez A. Emerging negative Atlantic Multidecadal
477 Oscillation index in spite of warm subtropics. *Sci. Rep*. **7**, 11224 (2017).
478 25. Lin, P., Yu, Z., Lü, J., Ding, M., Hu, A. & Liu, H. Two regimes of Atlantic Multidecadal
479 Oscillation: cross-basin dependent or Atlantic-intrinsic. *Sci. Bull*. **64(3)**, 198-204 (2019).
480 26. Hall, A., Cox, P. M., Huntingford, C. & Klein, S. Progressing emergent constraints on
481 future climate change. *Nat. Clim. Chang*. **9(4)**, 269-278 (2019).
482 27. Frankignoul, C., Gastineau, G. & Kwon, Y.-O. Estimation of the SST Response to
483 Anthropogenic and External Forcing and Its Impact on the Atlantic Multidecadal
484 Oscillation and the Pacific Decadal Oscillation. *J. Clim*. **30(24)**, 9871-9895 (2017).
485 28. Enfield, D. & Cid-Serrano, L. Secular and multidecadal warmings in the North Atlantic
486 and their relationships with major hurricane activity. *Int. J. Climatol*. **30**, 174-184
487 (2010).
488 29. Harris, I., Osborn, T. J., Jones, P. & Lister, D. Version 4 of the CRU TS monthly high-
489 resolution gridded multivariate climate dataset. *Sci. Data*. **7(1)**, 109 (2020).
490 30. Shurer, A., Tett, S. & Hegeril, G. Small influence of solar variability on climate over the
491 past millennium. *Nat. Geosci*. **7**, 104-108 (2014).
492 31. Schleussner, C., Divine, D, Donges, J., Miettinen, A., Donner, R. V. Indications for a
493 North Atlantic Ocean regime shift at the onset of the Little Ice Age. *Clim. Dynam*. **45**,
494 3623-3633 (2015).
495 32. Moreno-Chamarro, E., Zanchettin, D., Lohmann, K. & Jungclaus, J. H. An abrupt
496 weakening of the subpolar gyre as trigger of LIA-type episodes. *Clim. Dynam*. **48(3-4)**,
497 727-744 (2017).

498 33. Wang, J., Yang, B., Charpentier Ljungqvist, F., Luterbacher, J., Osborn, T. J., Briffa, K. R.
499 & Zorita, E. Internal and external forcing of multidecadal Atlantic variability over the
500 past 1,200 years. *Nat. Geosci*. **10**, 512-518 (2017).
501 34. PAGES 2k Consortium. A global multiproxy database for temperature reconstructions
502 of the Common Era. *Sci. Data*. **4**, 170088 (2017).
503 35. Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P.,
504 Kent, E. C. & Kaplan, A. Global analyses of sea surface temperature, sea ice, and night
505 marine air temperature since the late nineteenth. *Journ. Geophys. Res*. **108(D14)**
506 (2003).
507 36. Neukom, R., Shurer, A. P., Steiger, N. J. & Hegerl, G. C. Possible causes of data model
508 discrepancy in the temperature history of the last Millennium. *Sci. Rep*. **8**, 7572 (2018).
509 37. Michel, S., Swingedouw, D., Ortega, P., Khodri, M., Mignot, J. & Chavent, M.
510 Reconstructing climatic modes of variability from proxy records using ClimIndRec
511 version 1.0. *Geosci. Mod. Dev*. **13**, 841-858 (2020).
512 38. Nash, J. E. & Sutcliffe, J. V. River flow forecasting through conceptual models part I: A
513 discussion of principles. *J. Clim*. **10**, 282-290 (1970).
514 39. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5-32 (2001).
515 40. Shi, C., Sun, C., Wu, G., Wu, X., Chen, D., Masson-Delmotte, V., Li, J., Xue, J., Li, Z., Ji,
516 D., Zhang, J., Fan, Z., Shen, M., Shu, L. & Ciais, P. Summer temperature over the Tibetan
517 Plateau modulated by Atlantic Multidecadal Variability. *Journal of Climate*. **32(13)**,
518 4055-4067 (2019).
519 41. Li, J., Li, F., He, S., Wang, H. & Orsolini, Y. J. The Atlantic Multidecadal Variability phase
520 dependence of teleconnection between the North Atlantic Oscillation in February and
521 the Tibetan Plateau in March. *J. Clim*. **34(11)**, 4227-4242 (2021).

42. Monerie, P.-A., Robson, J., Dong, B. & Hodson, D. Role of the Atlantic Multidecadal Variability in modulating East Asian Climate. *Clim Dynam*. **56**, 381-398 (2021).

43. Büntgen, U., Frank, D. C., Nievergelt, D & Esper, J. Summer temperature variations in the European Alps, A.D. 755-2004. *J. Clim.* **19(21)**, 5606-5623 (2006).

44. Ortega, P., Lehner, F., Swingedouw, D., Masson-Delmotte, V., Raible, C. C., Casado, M., and Yiou, P. A model-testet North Atlantic Oscillation reconstruction for the past millennium. *Nature*. **523**, 71-74 (2015)

45. Sigl, M., Winstrup, M., McConnell, J. R., Welten, K. C., Plunkett, G., Ludlow, F., Büntgen, U., Caffee, M., Chellman, N., Dahl-Jensen, D., Fisher, H., Kipfstuhl, S., Kostick, C., Maselli, J., Mekhaldi, F., Mulvaney, R., Muscheler, R., Pasteris, D. R., Pilcher, J. R., Salzer, M., Schüpbach, S., Steffensen, J. P., Vinther, B. M. & Woodruff, T. E. Timing and climate forcing of volcanic eruptions for the past 2,500 years. *Nature*. **523**, 543-549 (2015).

46. Rao, M. P., Cook, E. R., Cook B. I., Anchukaitis, K. J., D'Arrigo, R. D., Krusic, P. J. & LeGrande, A. N. A double bootstrap approach to Superposed Epoch Analysis to evaluate response uncertainty. *Dendrochronologia*. **55**, 119-124 (2019).

47. Vieira, L. E. A., Solanki, S. K., Krikova, N. A. & Usoskin, I. Evolution of the solar irradiance during the Holocene. *Astronom., Astrophys*. **531**, A6 (2011).

48. Frankcombe, L. M., von der Heydt, A. & Dijkstra, H. North Atlantic multidecadal climate variability: An investigation of dominant time scales and processes. *J. Clim.* **23**, 3626-3638 (2010).

49. Lenton, T. M. Early warning of climate tipping point. *Nat. Clim. Change.* **1**, 201-208 (2011).

50. Lenton, T. M., Livina, V. N., Dakos, V., van Nes, E. H. and Scheffer, M. Early warning of climate tipping points from critical slowing down: comparing methods to improve robustness. *Phil. Trans. R. Soc. A*. **370**, 1185-1204 (2012).
51. Boulton, C. A., Allison, L. C. & Lenton, T. M. Early warning signals of Atlantic Meridional Overturning Circulation collapse in a fully coupled climate model. *Nat. Commun.* **5**, 5752 (2014).
52. Swingedouw, D., Speranza, C. I., Bartsch, A., Durand, G., Jamet, C., Beaugrand, G., & Conversi, A. Early warning from space for a few kipping points in physical, biological and social-ecological systems. *Surv. Geophys.* **41**, 1237-1284 (2020).
53. Smeed, D. A., Josey, S. A., Beaulieu, C., Johns, W. E., Moat, B. I., Frajka-Williams, E., Rayner, D., Meinen, C. S., Baringer, M. O., Bryden, H. L. & McCarthy, G. D. The North Atlantic Ocean is in a state of reduced overturning. *Geophys. Res. Lett.* **45(3)**, 1527-1533 (2018).
54. Knight, J. R., Allan, R. J., Folland, C. K. Vellinga, M. & Mann, M. E. A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophys. Res. Lett.* **32(20)**, L20708 (2005).
55. Caesar, L., McCarthy, G. D., Thornalley, D. J. R., Cahill, N. & Rahmstorf, S. Current Atlantic Meridional Overturning Circulation weakest in the last millennium. *Nat. Geosci.* **14**, 118-120 (2021).
56. Defrance, D., Carty, T., Rajaud, A., Dessay, N. & Sultan, B. Impacts of Greenland and Antarctic ice sheet melt on future Köppen climate zone changes simulated by an atmospheric and oceanic general circulation model. *Applied Geography*. **119**, 102216 (2020).

57. Valesco, J. A., Estrada, F., Calderón-Bustamante, O., Swingedouw, D., Ureta, C., Gay, C. & Defrance, D. Synergistic impacts of global warming and thermohaline circulation collapse on amphibians. *Commun. Biol*. **4**, 141 (2021).

58. Sutton, R. T. Climate science needs to take risk assesment much more seriously. *Bull. Am. Meteor. Soc.* **100(9)**, 1637-1642 (2019).

Acknowledgements: This research was partly funded by the Universite de Bordeaux. It is also funded by the LEFE-IMAGO project. This study benefited from the IPSL Prodiguer-Ciclad and Camelot facilities, supported by CNRS, UPMC Labex L-IPSL. Part of the code has also been run on the SurfSara Cartesius supercomputer (Amsterdam). Simulations for the LME experiments have been downloaded on the NCAR Climate Data gateway website (https://www.earthsystemgrid.org/). CRUTS4 data have been downloaded from the CRU data download webpage (https://crudata.uea.ac.uk/cru/data/hrg/#current). HadISST data have been downloaded from the Met Office Hadley Center website (https://www.metoffice.gov.uk/hadobs/hadisst/). For the historical experiments of CMIP5 models, data can be downloaded from on the ESGF website (https://esgf-node.llnl.gov/projects/esgf-llnl/). Codes for reconstructions, statistical analysis, statistical tests, and figures have been integrally implemented in R and bash UNIX languages. Authors were also funded by EU-H2020 Blue Action (Grant Agreement no. 727852, D.S., G.G, and J.M.), EUCP (Grant Agreement no 776613, D.S. and J.M.), ROADMAP (J.M. and G.G.) ARCHANGE (ANR-18-MPGA-0001, J.M. and G.G.) and TiPES (S. M.) research programmes.

Author contributions: S.M. have performed the different reconstructions, statistical analysis, figures, and the pseudo-proxy experiment of this study. S. M. has mainly written the

manuscript with important participation of D. S. D. S., J. M, M. K., P. O., G. G. and G. M have contributed to the results assessments, the manuscript writing, and have made suggestions to set up the manuscript's guiding thread.

Competing interests: The authors declare no competing interests.

Codes and data availability: All codes and data needed to reproduce this study are publicly available on the following Zenodo link: https://zenodo.org/record/4896670#.YLjdOS2w3dc.

Methods:

Instrumental AMV indices:

Historical AMV indices have been calculated using annually-resolved values. The AMV reconstructions compared and presented in this study are then smoothed with a ten year kernel filter.

We have computed, over the period 1870-2019, the three instrumental AMV indices[2,8,28] of this study using the Hadley Center Global Sea Ice and Sea Surface Temperatures (HadISST) dataset[36]. The three indices are based on the spatially averaged SST over the North Atlantic (between latitudes 0° and 60°N) and differ in the way the externally forced signal is removed.

The different approaches used assume that external forcings factors have different implications for temperatures and precipitation over time. The first index uses the global averaged SST anomalies as a proxy for the externally forced signal that is subtracted from NASST, resulting in the AMV_{TS} index[8] (Fig. 1a). This approach has limitations since it does not account for regional variations in external forcings, such as the distribution of anthropogenic aerosols, which was denser in North America-Europe in the middle of last century and became more prominent in Asia in recent decades[60]. To circumvent this problem, a different index, the AMV_T (Fig. 1a), is built using climate model historical simulations to isolate the forced component in NASST, which is calculated with a signal-to-noise maximizing empirical orthogonal function (EOF) that is then removed by an estimate of its 10-year smoothed effect[2] at each grid point of the North Atlantic. $AMV_T$is then obtained as the spatial average of the previously regressed time series. Finally, the AMV_F[28] is obtained as the spatial average of the North Atlantic time series regressed onto the ten years moving average of the global mean SST between 60°S and 60°N.

Estimation of NASSTs forced component using signal-to-noise maximizing EOF:

The forced component of NASST is estimated from historical simulations of 37 climate models (Extended Data Table 2). For each, the NASST anomalies are extracted and merged as columns of the same matrix. Using a Principal Component Analysis of the latter matrix, the first Principal Component is retained as the estimated forced component of NASST (Extended Data Fig. 1). For additional details, the reader is referred to ref. 2.

Proxy records database:

To select the proxy records of this study, a first large dataset is made by merging the PAGES 2k database (686) with other proxy records in neighboring continents of the North Atlantic used for previous AMV and NAO reconstructions[34,45]. Duplicates from the 3 sources are removed leading to a final database of 727 proxy records (P2k-ALL). Since the focus is on reconstructing the annual variations of the AMV, the proxy records from P2k-ALL finally used in the reconstruction have been selected to fulfill the following conditions:

1) They are annually-resolved
2) They are located in the Northern hemisphere (latitude > 0°).
3) They are significantly correlated at the 95% confidence level with at least one historical time series of either annual or seasonal precipitation or surface temperatures from the nearest grid point within the CRUTS4 dataset[30].

The first analyses when preparing this study singled out a proxy record from Asia (named "Asia.MOR1JU" in P2k-ALL), which had abnormally large RF weights (more than 5 times higher than the second) as compared to all the other proxy records. To prevent highly biasing the reconstruction towards this single proxy we have decided to remove it from the database used in this study.

The reconstruction procedure of this study is then based on the resulting database (P2k+). In the next method sections, we describe how different reconstructions are compared and how a final nested reconstruction of the AMV is obtained. These reconstructions will also use correlation tests to select proxies from P2k+ that are significantly correlated with a given AMV index, for a given learning and reconstruction period. This means that only a subset of the most relevant proxy records from the P2k+ database is finally used in each reconstruction.

Generation of the 312 initial reconstructions compared in the study:

The 312 reconstructions compared in this study, are performed for 26 time windows Γ: from 850-1975 to 850-2000, thus subsequently incrementing the superior boundary by one year. We use this approach as a way to sample the sensitivity of the reconstructions to the calibration period, as for the shortest windows more proxies are available but the regression models are built with shorter time series and therefore fewer degrees of freedom . These 26 temporal windows are used in combination with three AMV indices[2,8,28] and the four regression methods[40,61-63]. Detailed description of these regression methods is given in ref. 38 and Supplementary Information. All these setups are tested by only using proxies available and significantly correlated at the 95% confidence level with the respective AMV index. We thus end up with $26 \times 3 \times 4 = 312$ final reconstructions that are compared in this study using the CE metric.

For the NASST reconstruction, 84 setups are compared, by shuffling the same 26 temporal windows and the same 4 regression methods as the AMV indices.

Computation of a reconstruction and evaluation:

We define the reconstruction period as Γ, defined by N annual time steps, and the common period of the proxy records and the AMV index as T, in this case defined by $n < N$ annual time steps such that $T \subset \Gamma$. We then define the AMV index as $Y \in R^n$ and the matrix of the

available proxy records as $X \in R^{N\times p}$. We finally denote as$x \in R^{n\times p}$ the sub-matrix of X that contains the proxy records values over the time frame T. X can then be denoted as $X = [(X_t^j)_{t\in\Gamma}]_{1\le k\le p}$ and $x = [x^j]_{1\le j\le p} = [(X_t^j)_{t\in T}]_{1\le j\le p}$.

We then randomly split T in $R = 30$ pairs of training/testing samples respectively denoted, $\forall 1 \le r \le R$, by $\{x_{(train)}^{(r)}; Y_{(train)}^{(r)}\}$and $\{x_{(test)}^{(r)}; Y_{(test)}^{(r)}\}$. Here, the training sample size is set to be 80% of the length of $T$and, by extension, the testing sample size is 20% of the length of T.

For statistical modelling, we use $Y_{(train)}^{(r)}$as predictand and $x_{(train)}^{(r)}$as predictors. For a given regression method denoted M, we apply KFCV (*cf*. Methods, section "K-Fold cross-validation (KFCV)") to each training set $\left\{x_{(train)}^{(r)}; Y_{(train)}^{(r)}\right\}$ as a metric to find the optimal set of parameters associated to the training sample and M.

$M$and the associated optimal set of control parameters are then applied to $X^{(r)}$ in order to reconstruct $Y^{(r)}$ on both the testing period, giving $\hat{Y}_{(test)}^{(r)}$, and the reconstruction period, giving $\hat{Y}_{(rec)}^{(r)}$. This involves that $\hat{Y}_{(test)}^{(r)} = (\hat{Y}_{(rec)}^{(r)})_{t\in T}$.The validation score associated to the r^{th} training sample is then calculated using the Coefficient of Efficiency (CE)[39] over the r^{th} testing sample:

$$s^r = CE\left(\hat{Y}_{(test)}^{(r)}, Y_{(test)}^{(r)}\right) = 1 - \frac{\sum_{i=1}^{m} (Y_{i(test)}^{(r)} - \hat{Y}_{i(test)}^{(r)})^2}{\sum_{i=1}^{m} (Y_{i(test)}^{(r)} - \underline{Y}_{(test)}^{(r)})^2}, \text{ with } \underline{Y}_{(test)}^{(r)} = \frac{1}{m}\sum_{i=1}^{m} Y_{i(test)}^{(r)}$$

Where m is the length of the testing sample.

This validation score gives an estimation of the accuracy of the statistical model when reconstructing the observed variability not included in the reconstruction period. $CE < 0$ means that the sample average of the testing period is more reliable than the output given by the statistical model[39]. Contrarily, $CE > 0$, means that the statistical model gives a more

reliable reconstruction than the empirical average of the testing sample[39], the associated reconstruction is thereby considered as reliable in this study.

The reconstruction for a given AMV index Y performed on a given time frame Γ using a given statistical regression method M is obtained by applying it with KFCV[38] (Methods, next section) over the whole learning sample. This reconstruction is thus associated to a global validation score, calculated as the mean of the individual validation scores obtained over the random splits: $s = avg(\{s_r\}_{1 \leq r \leq R})$.

K-Fold cross-validation (KFCV):

Each method requires an optimization of its own set of control parameters θ. To estimate the optimal set of control parameters θ_{opt} on a given training set $\{X_{train}, Y_{train}\}$, we use the KFCV approach[64,65]. Cross validation methods are in general widely used as model validation and selection techniques.

The KFCV splits the observations into a partition of K groups of the same sizes (or with approximately the same size if the length of the training set is not divisible by K). $\forall 1 \leq k \leq K$, we denote $\{X_{(k)}, Y_{(k)}\}$, which contain only observations for the k^{th} drawn sample. We denote $\{X_{(-k)}, Y_{(-k)}\}$ the $K-1$ other sets. For all possible values of $\theta \in \Theta$, we scan the K models based on the sets $\{X_{(-k)}, Y_{(-k)}\}$. The empirical optimal set of control parameters is obtained by minimizing the averaged Root Mean Squared Errors (RMSE) on the K splits by considering all possible values of θ. The optimal KFCV set of control parameters $\theta_{KF}$is determined by:

$$\hat{\theta}_{opt} = \theta_{KF} = arg\, \frac{1}{K} \sum_{k=1}^{K} RMSE(Y_{(k)}, \hat{Y}_{(k),\theta})$$

Nested reconstruction:

In this study, the best reconstruction found (defined as the one yielding the best CE metric) is the reconstruction of the AMV_F index with the random forest method over the period 850-1987, using only the proxy records significantly correlated at the 95% confidence level with the AMV_F index over the training periods (see Extended Data Fig. 2).

Using these methodological choices (calibration period, AMV definition and reconstruction method), we have performed a set of 1021 nested reconstructions for the periods 850-1987 to 1869-1987, subsequently using increments of one year for the inferior boundary, which allow us to use an increasing number of proxy records to reconstruct the most recent years. The nested reconstruction (*e.g.,* the reconstruction presented in this study), is obtained by concatenating the first year in each of the 1021 reconstructions[34].

Random forest variable importance:

The weights of the proxy records used for the nested reconstruction are presented in Fig. 2c. Those weights have been calculated using the random forest variable importance[40]. Different importance metrics exist, and for this study we have selected the commonly used Mean Decrease in Impurity (MDI), also known as Gini importance. The MDI of a given proxy record is calculated as the sum of the number of splits where it is used across the K trees (see Supplementary Information for details on the regression methods), proportionally to the numbers of split samples in all trees[40] (*cf*. Supplementary Information). For Fig. 2c, the MDI for each proxy is aggregated over the 1021 reconstructions using a weight of n/N, where n is the number of available proxies for a given time step, and N the total number of proxies used at the end for the reconstruction (*i.e*., N=55). Finally, Fig. 2c is computed by calculating the importance of each proxy as a fraction of the previously calculated importance and the total importance over the 1001 reconstructions. The same is done in Fig. 3b, but each importance is also averaged over the PPE performed on the 12 members of CESM1-LME used.

NASST reconstruction:

To reconstruct the NASST, the same procedure as for the AMV is applied to select the optimal reconstruction approach, with the only difference that only one index definition is considered, as no method to remove the forced variability is applied. This means that the optimal model selection is made for 312/3=104 compared reconstructions. The final NASST reconstruction is obtained using nested PCR reconstructions for the 1021 periods 850-1977 to 1869-1977, with an increment of one year for the inferior boundary.

Composites of ocean proxy records time series:

Since corals often are very short records and ocean sediment cores have too low temporal resolution (preventing them to meet the requirement of being annually resolved), there is a shortage of ocean records contributing to the reconstructions, which is almost exclusively based on terrestrial records. Interestingly, the low-frequency part in the annually resolved reconstructions can be verified against the ocean records from the Ocean 2K database in Fig. 3 that have not been used in the reconstruction. To avoid overfitting, correlations significance shown in Fig. 3a are only calculated for the preindustrial period (before 1870) with an AR1 correction for the correlation tests to avoid falsely detected significance due to the low resolution of some proxies (See “Statistical information” method section). For the same overfitting reasons, the composite average time series are performed by multiplying by -1 ocean proxies which have negative correlations with the AMV over the historical period only. There is an exception for some ocean records (<10 from Fig 3a) that do not overlap with our reconstruction over the historical period. These time series are therefore multiplied by -1 if their correlations are negative over their overlapping period with the AMV reconstruction.

Two-way multimember pseudo-proxy experiments:

782 For the pseudo proxy experiment, we use 12 last millennium large ensemble members of the
783 National Center of Atmospheric Research (NCAR) Community Earth System Model 1 (CESM1).
784 Since the calculation of a trend for a given NASST time series is time-dependent, we
785 distinguish the calculation of the model AMV over the preindustrial (PI) period (pre-1870,
786 AMV_{PI}) and the historical one (AMV_H), notably because anthropogenic forcings were small
787 during the PI period as compared to the recent one (historical). For AMV_H, we calculate the
788 AMV_F, similarly to the real experiment. For calculating AMV_{PI}, the same is applied but by
789 estimating the NASSTs relationships with the global SST over the PI period.
790 The "Proxy records database" method section emphasizes the fact that the proxy records do
791 not target the same climate variables and seasonalities. For this reason, pseudo-proxies mimic
792 the real-world proxies by taking, in the model, the variable and season (or annual values) with
793 which the real proxies exhibit the largest absolute correlations (see Extended Data Table 3)
794 with the closest grid points from the CRUTS4 dataset. Gaps and missing values in real proxy
795 records are also reproduced in the pseudo-proxy time series.
796 For both PPE cases presented below, and for the sake of reducing computational costs, nested
797 reconstructions have been made with a 20 years time step for the inferior boundary (from
798 850-1987 to 1850-1987) instead of the 1 year one which is used for the real experiment.
799
800 *Proxy to model:*
801 The pseudo-proxy experiment is first used to reconstruct a model AMV reconstruction
802 using the same proxy records and RF method as for the real experiment, and for each time
803 step of the nested reconstruction. Therefore, the reconstruction scores presented in Fig. 4.1
804 for each CESM1 member are those obtained over the 1020 nested reconstruction timeframes.
805 The correlations are those calculated between the RF-based reconstruction of the model AMV

and AMV_{PI} . For each member, the proxy records' weights are calculated in the same way as for Fig 2c, and an ensemble average is presented in Fig. 4b.

Model to proxy:

The second step of the PPE consists in training RF models directly within the CESM1 members, in which the pseudo-proxies are selected using 95% confidence level correlation tests with the model AMV. These trained RF models tailored to the model simulations are then applied to the historical simulatedAMV_F index and compared to the reconstruction using real-world derived weights in Fig. 5. Since real-world proxy records have been measured with specific units (tree ring MXD, ice core $\delta^{18}O$,...), the model pseudo-proxies are rescaled to the mean and the variance of the corresponding real-world proxy. For the same reason, the pseudo-proxy is multiplied by -1 if its correlation with the model AMV has an opposite sign to that of the real-world proxy with the real world AMV. This PPE approach is similar to the model-constrained one from a published NAO reconstruction based on the PCR method[45], which has been adapted to the RF one in this study.

Early warning signal test:

We base our approach on methods for the detection of incoming climate tipping points[50,51], recently applied to detect an AMOC slowdown in a general circulation model[52]. The AMV reconstruction is firstly smoothed using a Kernel Gaussian filtering with a bandwidth of 100 years. The annually-resolved AMV is then regressed onto its long term filtered version. AR(1) coefficients of the residuals from this regression are calculated for different sliding window lengths WL=200, 250, 300, 350, 400 years (Fig. 4b). Kendall τ is calculated for each of the AR(1) coefficient series. Contrary to a former study focusing on early warning signal applied to

model-based investigation of an AMOC collapse[52], we cannot use a model-based estimate of the significance of Kendall τ, which is rather calculated using a gaussian approximation as detailed in the "Statistical information" Methods section.

Boxplots:

For all boxplots of the study, the median is shown as a heavy darkline. Boxplots edges give first and third quartiles. Boxplot "whiskers" gives the full range without including outliers, which are not shown here for better graphical representations. A point from a boxplot is here considered as an outlier when it is outside 1.5 times the interquartile range above the upper quartile and below the lower quartile.

Statistical information:

This section describes the different statistical tests of this study.

- Fig 1b,c: For each grid point, a two-tailed Student test is applied to the regression coefficients between the corresponding climate variable and the AMV indices. The degrees of freedom are corrected using time series autocorrelations as in ref. 17 and 38.
- Fig. 4b: The Kendall rank correlation coefficient, or Kendall τ coefficient, measures the ordinal association between two quantities, here AR(1) coefficients denoted $(x_i)_{1\leq i\leq n}$ here, and time denoted $(y_i)_{1\leq i\leq n}$. The statistic is given by:

$$\tau = \frac{n_c - n_d}{n_0}$$

Where, considering (x_1, y_1), (x_2, y_2), …, (x_n, y_n) the ensemble of joint pairs:

$$n_c = card_{i\neq j}\{\{x_i > x_j \cap\ \ y_i > y_j\} \cup\ \ \{x_i < x_j \cap\ \ y_i < y_j\}\}, (i,j) \in [\![1,n]\!]^2$$

$$n_d = card_{i\neq j}\{\{x_i > x_j \cap\ \ y_i < y_j\} \cup\ \ \{x_i < x_j \cap\ \ y_i > y_j\}\}, (i,j) \in [\![1,n]\!]^2$$

$$n_0 = \frac{n(n-1)}{2}$$

For large sample (n>50), as in this study, the distribution is approximated with a Gaussian distribution of mean 0 and variance $\frac{2(2n+5)}{9n(n-1)}$, under the null hypothesis H_0: "$\tau = 0$"which is tested against the alternative hypothesis H_1: "$\tau \neq 0$". The p-value (shown in Fig 4b) of the test is deduced from the quantile of this distribution.

- Correlation tests: The same bilateral Student test for correlation than ref. 17 and 38 is applied for the whole study, with corrected degrees of freedom using time series autocorrelation. The p-values of all correlations presented in this study are also based on this test, including Fig 3a that includes tests with low-resolution proxy records.

60. Smith, D. M., Booth, B. B. B., Dunstone, N. J., Eade, R., Hermanson, L., Jones, G. S., Scaife, A. A., Sheen, K. L., Thompson, V. Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nat. Clim. Change*. **6**, 936-940 (2016).
61. Hotelling, H. The relations of the newer multivariate statistical methods to factor analysis. *British Journal of Statistical Psychology*. **10**, 69-76 (1957).
62. Wold, H. System analysis by Partial Least Squares**.** *IIASA Collaborative paper. IIASA Laxenburg.* (1983)
63. Zou, H., Hastie, T. Regularization and variable selection via the elastic net. *J. Roy. Stat. Soc*. **67**, 301-32 (2005).
64. Stone, M. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*. **36**, 111-147 (1974).

65. Geisser, S. The predictive sample reuse method with applications. *Journal of the Royal Statistical Society.* **70**, 320-328 (1975).

Figures

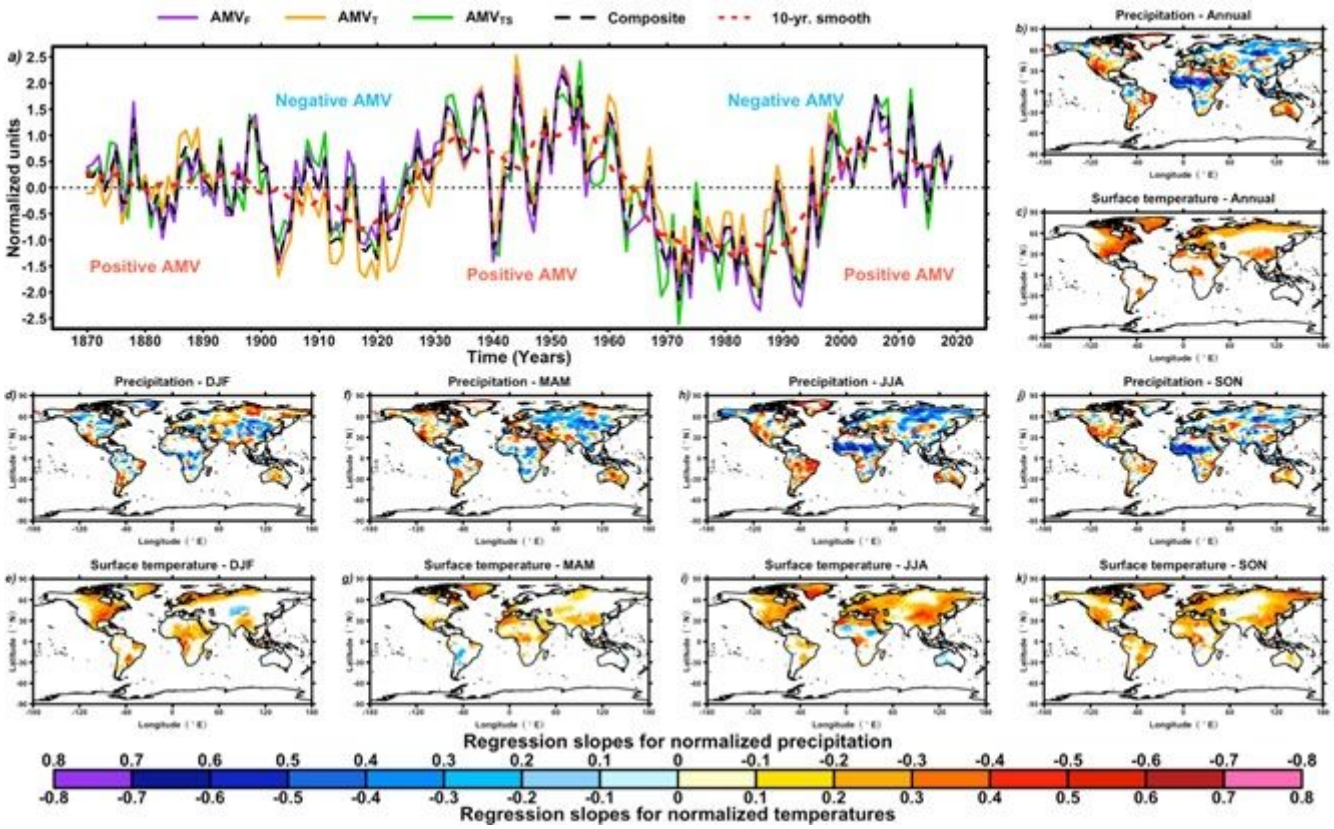

Figure 1

Climatic impacts of the AMV over the historical era. a) Historical evolution of AMV indices investigated in this study for the period 1870-2017 calculated using the HadISST dataset36 (cf. Methods). b,d,f,h,j) Map of averaged regression coefficients between the 10-years smoothed composite of the three AMV indices from a) and CRUTS430 precipitation data for the period 1901-2017. Maps are respectively relating Annual, DJF, MAM, JJA and SON regression coefficients. c,e,g,i,k) Map of regression coefficients between the composite of the three AMV indices from a) and CRUTS430 surface temperature data for the period 1901-2017. Maps are respectively relating Annual, DJF, MAM, JJA and SON regression coefficients. For c-e-), white grid points indicate that regression coefficients are not significantly different than 0 at the 90% confidence level, using a two-tailed student test with corrected degrees of freedom17,38 (Methods).

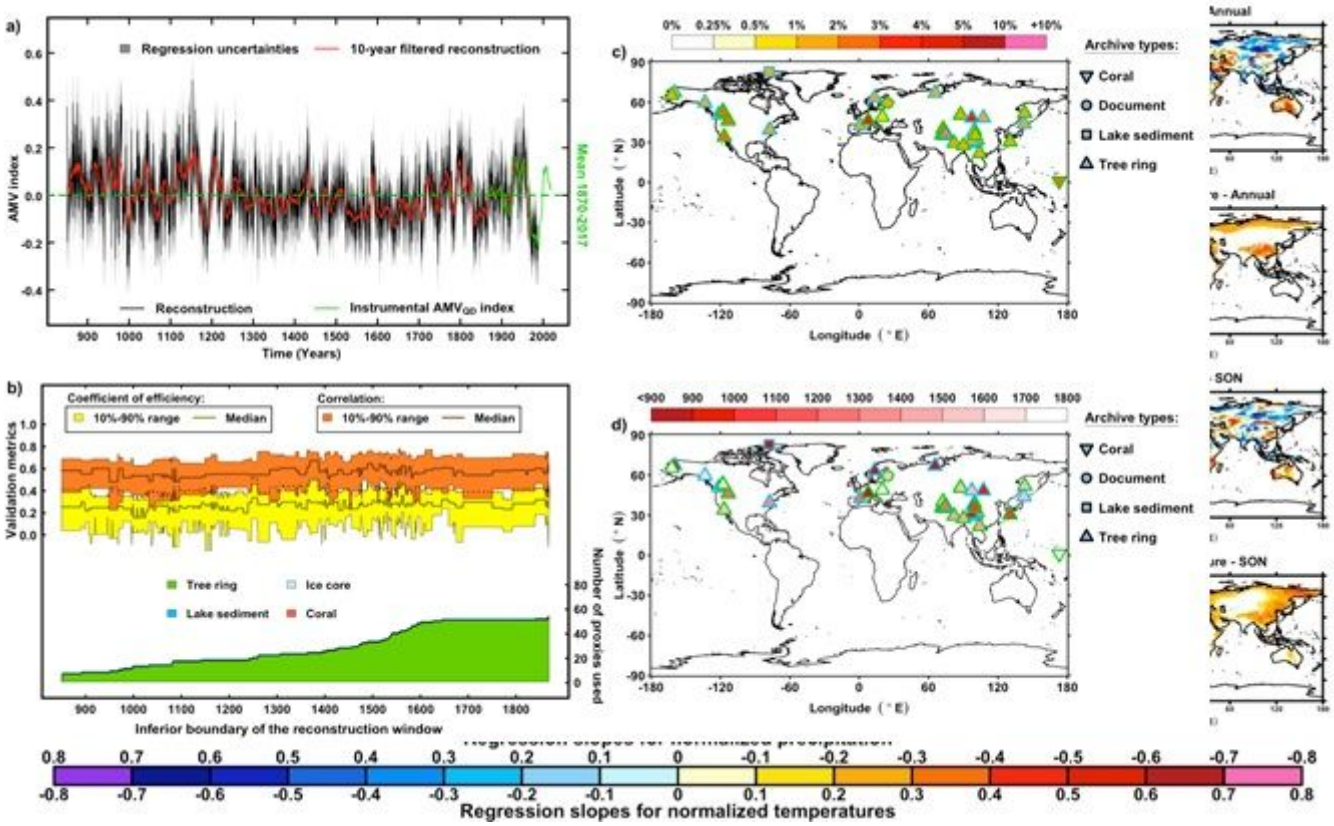

Figure 2

Nested reconstruction of the AMV and related proxies. a) Black line: Annually-resolved nested reconstruction of AMV_F (°C) using random forest40 (cf. Methods). Red line: 10 years kernel smooth of the annually-resolved nested reconstruction (black line). The regression uncertainties of the annually-resolved nested reconstruction (black line) are defined for each time step of the nested reconstruction as ± 2 standard error of the regression. Green line is the time series of the instrumental AMV_Fcalculated from historical SST data30 b) Validation metrics (CE in yellow and correlation in orange) obtained for 30 training-testing splits, and proxy records types availability for the nested AMV_F reconstruction (bottom). c) Proxies weights from the random forest method, relative to the proxy records temporal availability (see Methods) d) Temporal coverage of the availability of the proxy records.

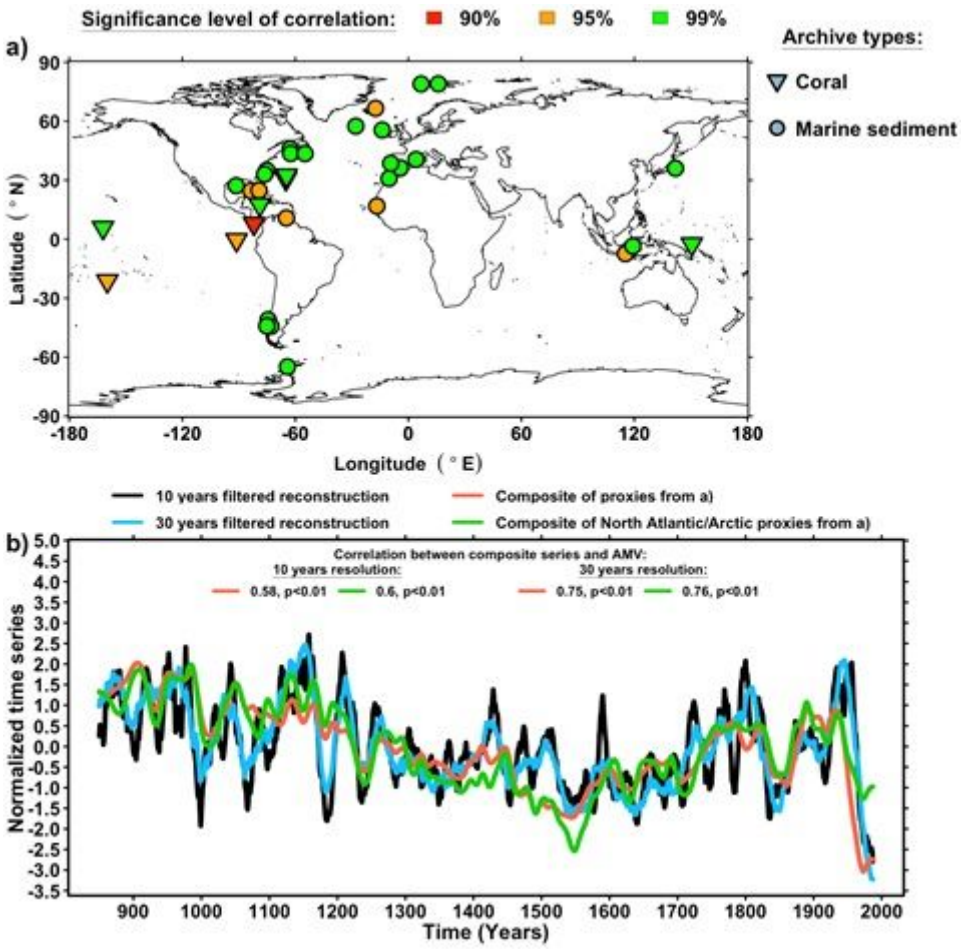

Figure 3

Comparison with independent ocean proxy records. a) 37 Ocean 2k proxy records35 significantly correlated at least at the 90% confidence with the AMV over the pre-industrial period (i.e., prior to 1870). b) 10-years (black) and 30-years (blue) kernel smooth of the AMV reconstruction.

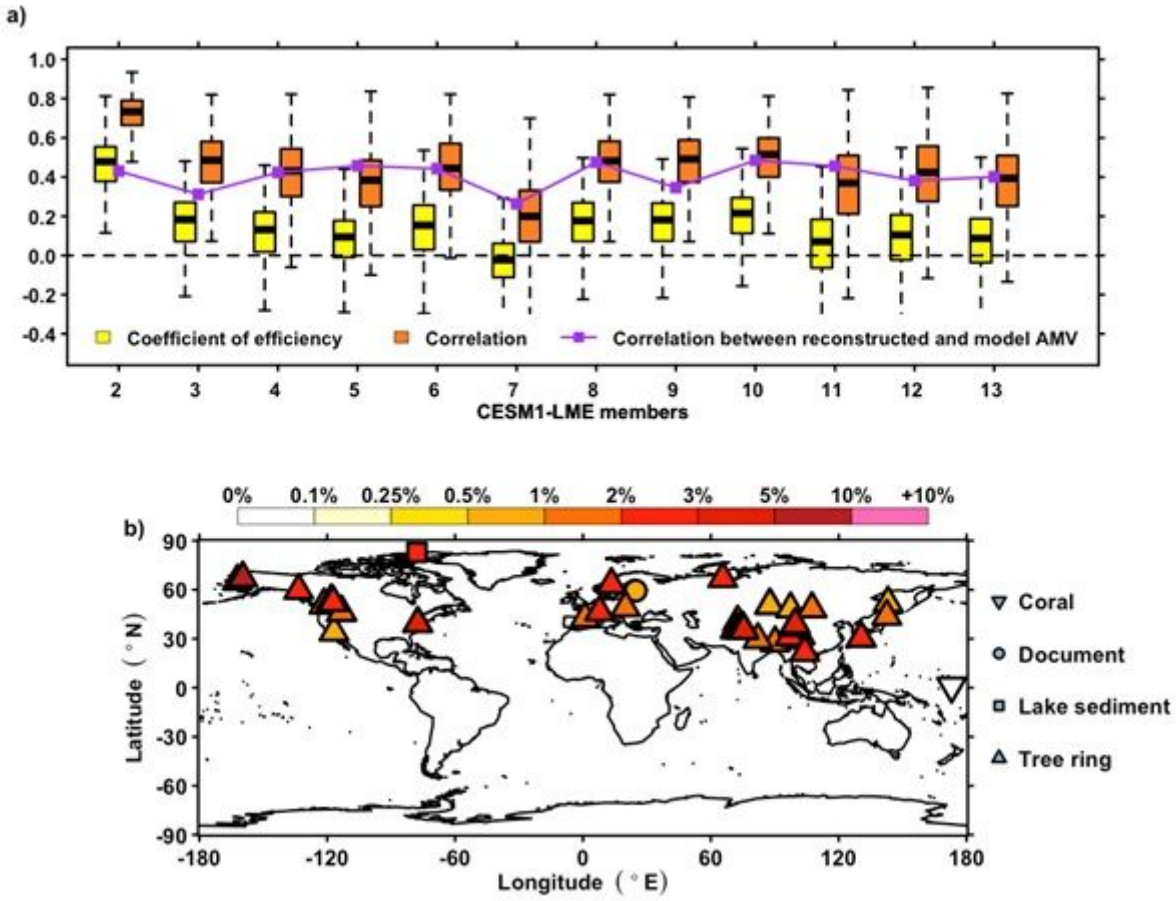

Figure 4

Proxy to model PPE validation. a) CE scores (yellow boxplots), correlation scores (orange boxplots) and correlation between the model AMV and the reconstructed AMV within the model simulations (Methods) (purple line) for 12 members of CESM1-LME. b) Weights of the proxy records from the model simulations from 12 members of CESM1-LME (Methods).

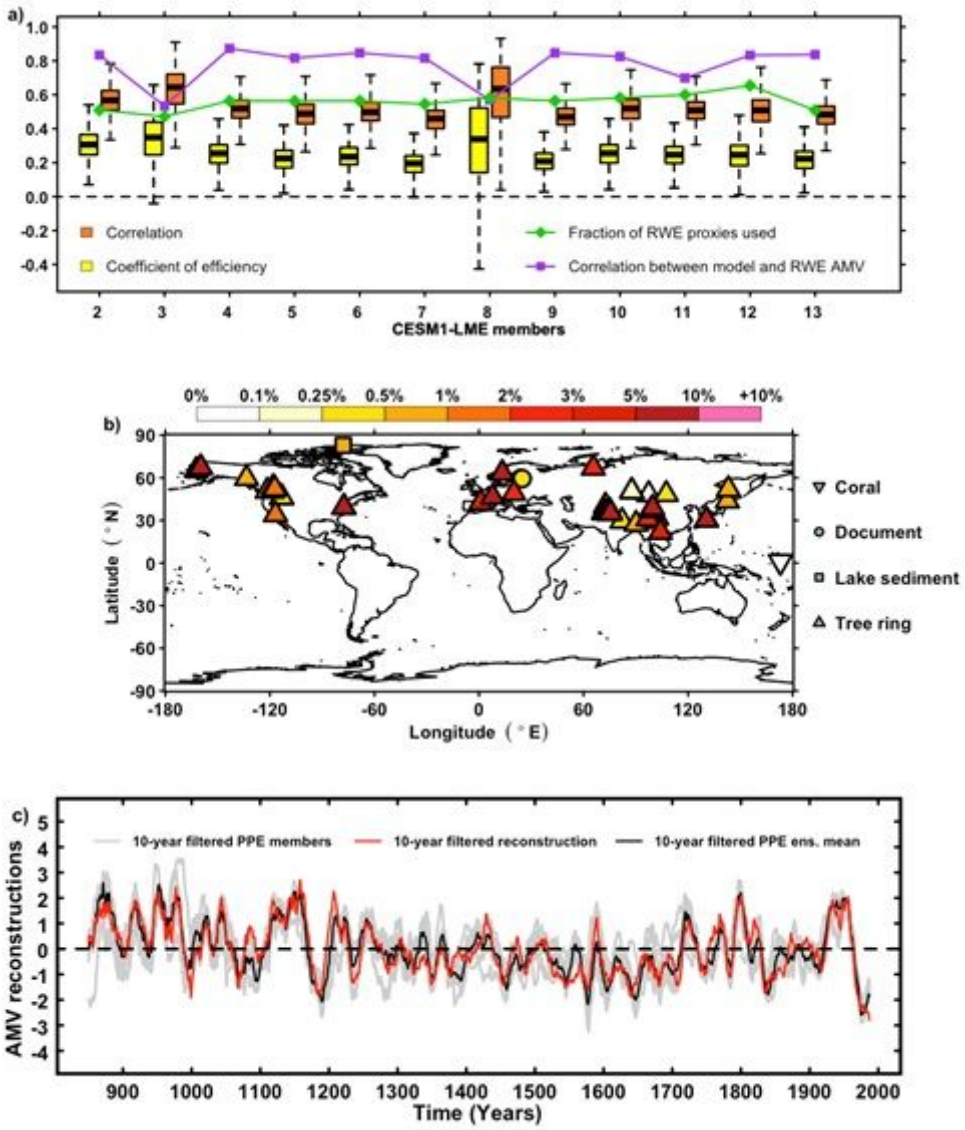

Figure 5

Model to proxy PPE validation. a) CE scores (yellow boxplots), correlation scores (orange boxplots) and correlation between the real-world AMV_F reconstruction and model-constrained44 AMV_F reconstructions (Methods) (purple line) for the 12 members of CESM1-LME. Green line indicates the fraction of proxy records from the real experiments used in the PPE experiment (see Methods) b) Weights of the proxy records from the real-world experiment for RF trained within the model simulations (Methods). c) Grey lines: 10-years kernel smooth of the 12 model-based experiment, based on each CESM1-LME member. Black: 10-years kernel smooth of the ensemble average of the 12 model-based reconstructions. Red: 10-years kernel smooth of the reconstruction from the real experiment.

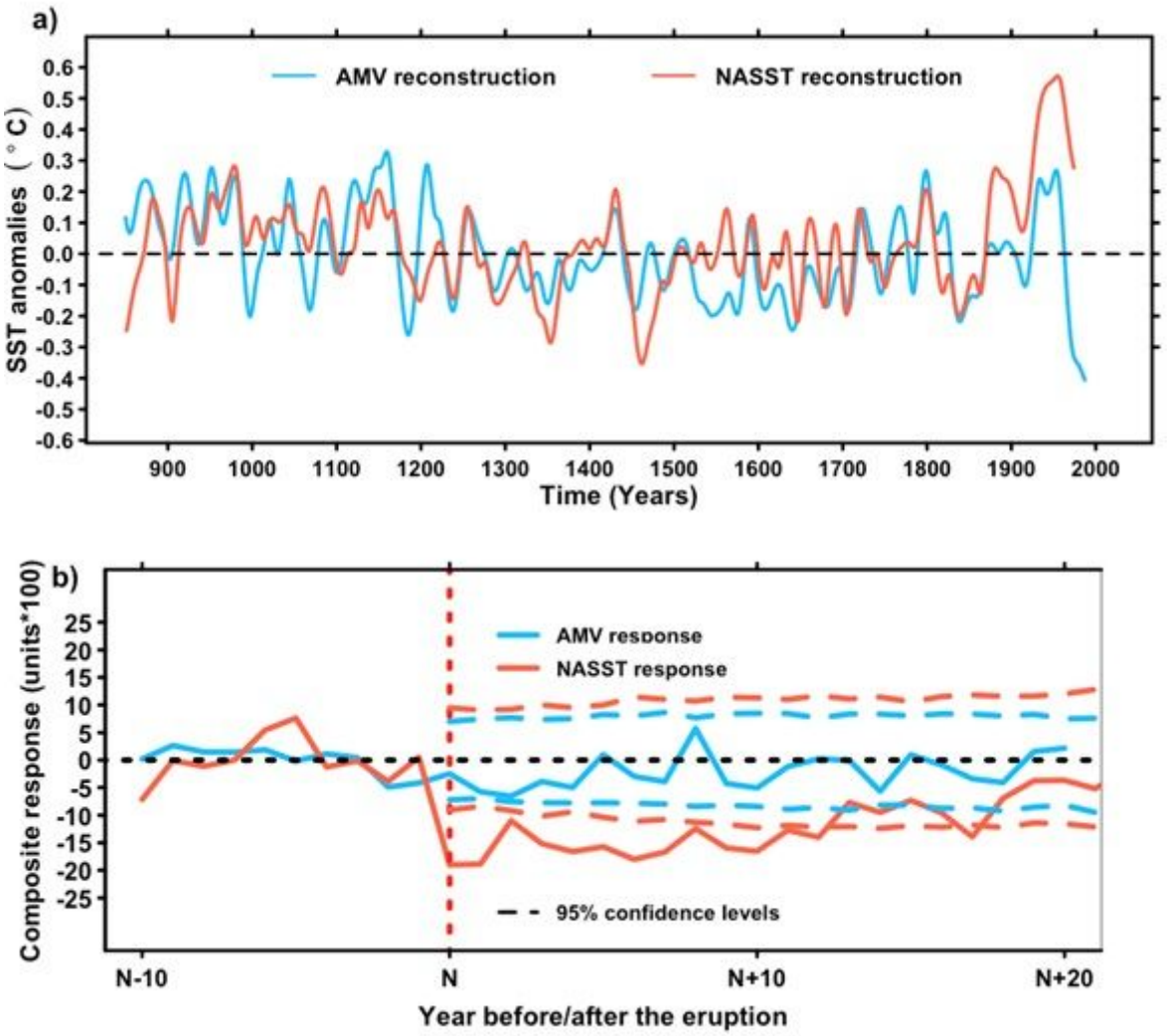

Figure 6

Comparison of the AMV with NASST and volcanic forcing. a) Final reconstructions of AMV and NASST. b) Superposed epoch analysis47 for responses of the AMV and NASST reconstructions to the ten largest eruptions46 of the last millennium (see Extended Data Table 4). Composite series are performed for 31 years, for which the 11th are the actual years of the eruptions. Each individual response is centered to its values 10 years before the eruption (from N-10 to N-1) before computing the composite time series. 95% confidence levels have been calculated using a Monte-Carlo approach47.

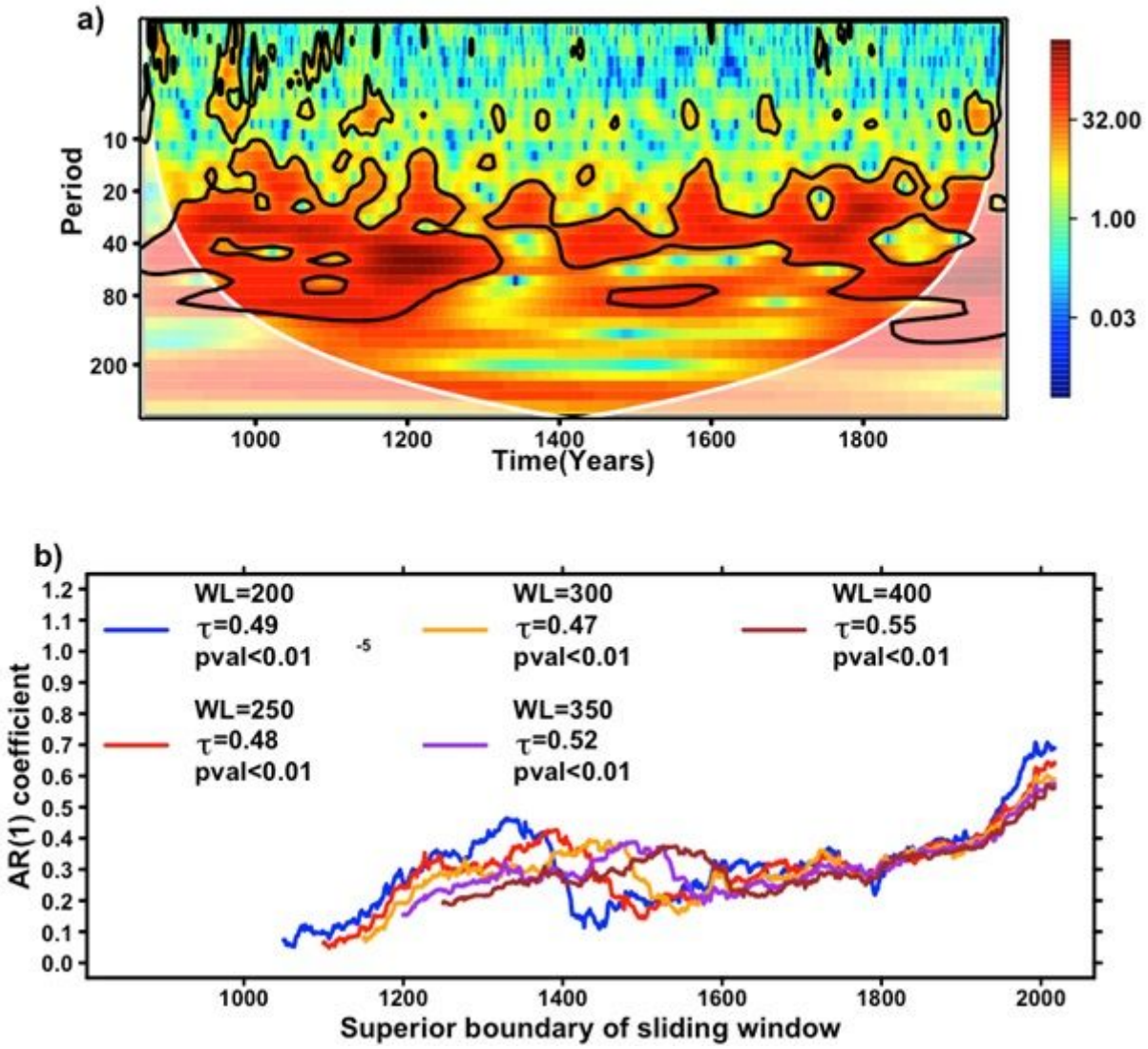

Figure 7

Spectral analysis and early warning signal. a) Discrete wavelet transform of the nested AMV reconstruction from this study. Contours provide a 90% confidence level. The white line and the light white-shaded area below indicate the cone of influence. The cone of influence gives the spectrum borders where the edge effect (i.e., the time boundary effect) becomes too important, which cannot be robustly interpreted. b) Early warning signal test50-52 of the nested AMV_F reconstruction (Methods) based on AR1 coefficients, for different window lengths (WL). For each WL, sliding AR1 coefficient are computed and a Kendall τ statistics between time and the sliding AR1 time series are calculated. Significance is approximated using Gaussian distributions because of the large length (>50) of the AR1 coefficients (see Methods).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ExtendedDataFig1.docx
- ExtendedDataFig2.docx
- ExtendedDataFig3.docx
- ExtendedDataFig4.docx
- ExtendedDataFig5.docx
- ExtendedDataTable1.docx
- ExtendedDataTable2.docx
- ExtendedDataTable3.docx
- ExtendedDataTable4.docx
- SupplementaryInformationREFSForExtDataTab1.pdf
- SupplementaryInformationRegressionMethods.pdf