Supplement of Geosci. Model Dev., 13, 841–858, 2020
https://doi.org/10.5194/gmd-13-841-2020-supplement

Geoscientific Model Development
Open Access
EGU

Supplement of

Reconstructing climatic modes of variability from proxy records using ClimIndRec version 1.0

Simon Michel et al.

Correspondence to: Simon Michel (simon.michel@u-bordeaux.fr)

Supplementary table 1: Proxy records not in PAGES 2k

N °	Code	Location	Longitude (°E)	Latitude (°N)	First year	Last year	Archive	Proxy type	Related variable	Seasonality	Ref.
1	accr-crete-yr-annual	Crete	-38.50	71.00	1000	1973	Ice core	Snow accumulation	Precip.	Annual	Andersen et al., 2006
2	accr-GISP2-yr-annual	GISP2	-38.50	72.60	1000	1988	Ice core	Snow accumulation	Precip.	Annual	Cuffey et al., 1995
3	d18O-agass-79-annual	Agassiz	-77.00	80.70	1000	1972	Ice core	$\delta^{18}O$	SAT	Annual	Fisher et al., 1995
4	icecore-GISP2-ssNA-annual	GISP2	-39.00	73.00	1000	1986	Ice core	Sea salt Na	SLP	DJF	Meeker and Mayewski, 2002
5	speleo-crystal-d18O-annual	Crystal cave	-121.00	36.90	1000	2007	Speleothem	$\delta^{18}O$	SAT	Annual	McCabe-Glynn et al., 2013
6	tree-alps-Tjjas-annual	European Alps	9.00	46.00	1000	2004	Tree ring	Tree ring MXD	SAT	JJAS	Büntgen et al., 2012
7	tree-AR050-stah-annual	Black Swamp	-91.30	35.15	1019	1980	Tree ring	Tree ring width	SAT	Annual	Stahle, 1996a
8	tree-AR052-stah-annual	Mayberry Slough	-89.00	35.50	1000	1990	Tree ring	Tree ring width	SAT	Annual	Stahle and Cleaveland, 2005a
9	tree-CA051-tosh-annual	San Gorgonio	-116.82	33.40	1000	1970	Tree ring	Tree ring width	SAT	Annual	Tosh, 1994
10	tree-ca605-king-annual	Mammoth Peak	-119.00	41.00	1000	1996	Tree ring	Tree ring width	Precip.	NDJFM	Bunn et al., 2005
11	tree-CA640-graum-annual	Hamilton	-118.92	39.00	1000	1988	Tree ring	Tree ring width	Precip.	NDJFM	Bunn et al., 2005
12	tree-co572-woodho-annual	Lily Lake	-105.60	40.30	1000	1998	Tree ring	Tree ring width	SAT	Annual	Woodhouse and Brown, 2006
13	tree-FL001-stah-annual	Choctawhatchee River	-85.92	30.45	1000	1992	Tree ring	Tree ring width	SAT	Annual	Stahle and Cleaveland, 2005b
14	tree-forfjo-cloud-annual	Forfjorddalen	15.73	68.80	1000	2001	Tree ring	Tree ring $\delta^{13}C$	Cloud %	JJA	Young et al., 2012
15	tree-LA001-stah-annual	Big Cypress	-92.97	32.25	1000	1988	Tree ring	Tree ring width	SAT	Annual	Stahle, 1996b
16	tree-mor-pdsi-annual	Morocco	-5.00	33.75	1049	2001	Tree ring	Tree ring width	PSI	FMAMJ	Esper et al., 2007
17	tree-mt113-wagon-annual	Yellow Mountain Ridge (Entire Bark Trees)	-109.80	45.60	1000	1998	Tree ring	Tree ring width	Precip.	NDJFM	Graumlich et al., 2003
18	tree-NM584-touch-annual	Mesa Alta	-106.60	36.20	1000	2007	Tree ring	Tree ring width	Precip.	ONDJFMAMJ	Touchan et al., 2011
19	tree-NV516-grayb-annual	Hill 10842	-114.20	38.90	1000	1984	Tree ring	Tree ring width	SAT	Annual	Graybill, 1994a
20	tree-nv517-grayb-annual	Spring Mountains Lower	-114.70	34.30	1000	1984	Tree ring	Tree ring width	SAT	Annual	Graybill, 1994b
21	tree-SCpla-precip-annual	S. Colorado Plateau I	-109.30	37.50	1000	1987	Tree ring	Tree ring width	Precip.	October-July	Salzer and Kipfmueller, 2005
22	tree-SCpla-temp-annual	S. Colorado Plateau II	-110.70	36.50	1000	1996	Tree ring	Tree ring width	SAT	Annual	Salzer and Kipfmueller, 2005
23	tree-siber-temp-annual	Taimyr-Putoran	103.00	71.29	1000	1996	Tree ring	Tree ring width	SAT	Annual	Naurzbaev et al., 2002
24	tree-swit177-schwein-annual	Lauenen + div. Stao	6.50	46.42	1000	1976	Tree ring	Tree ring width	SAT	JJA	Schweingruber, 1988
25	tree-UT508-grayb-annual	Wild Horse Ridge	-110.10	40.00	1000	1985	Tree ring	Tree ring width	SAT	Annual	Graybill, 1994c
26	tree-UT509-grayb-annual	Mammoth Creak	-112.67	37.65	1000	1989	Tree ring	Tree ring width	SAT	Annual	Graybill, 1994d
27	tree-albermale-trw-annual	Albermale Sound	-76.00	36.00	934	2005	Tree ring	Tree ring width	PDSI	July	Stahle et al., 2013
28	tree-arjeplog-bi-annual	Arjeplog	17.90	66.50	1200	2010	Tree ring	Tree ring BI	SAT	JJA	Björklund et al., 2014
29	tree-jamtland-mxd-annual	JLamtland	15.00	63.10	800	2011	Tree ring	Tree ring MXD	SAT	AMJJAS	Zhang et al., 2016
30	tree-colzad-trw-annual	Col du Zad	-5.10	33.00	984	1984	Tree ring	Tree ring width	PSI	FMAMJ	Esper et al., 2007
31	tree-forf-mxd-annual	Forfjorffalen-x	15.70	68.80	978	2005	Tree ring	Tree ring MXD	SAT	AMJJAS	McCarroll et al., 2013
32	tree-khibiny-bi-annual	Khibiny	33.50	67.50	821	2005	Tree ring	Tree ring BI	SAT	JJA	McCarroll et al., 2013
33	tree-laanila-mxd-annual	Laanila	27.30	68.50	800	2005	Tree ring	Tree ring MXD	SAT	JJA	McCarroll et al., 2013
34	tree-manitoba-trw-annual	S Manitoba	-97.10	49.50	1409	1998	Tree ring	Tree ring width	Precip.	Annual	George and Nielsen, 2002
35	tree-mesoamerica-trw-annual	Mesoamerica	-100.00	20.00	800	2008	Tree ring	Tree ring width	PDSI	June	Stahle et al., 2011
36	accr-NGRIPs-yr-annual	NGRIP-s	-42.00	76.00	800	1995	Ice core	Snow accumulation	Precip	Annual	Andersen et al., 2007
37	tree-potoriv-trw-annual	Potomac River	-77.50	39.30	950	2001	Tree ring	Tree ring width	Stream flow	MJJAS	Maxwell et al., 2011
38	tree-quebecx-mxd-annual	Quebec-x	-77.50	39.30	1373	1988	Tree ring	Tree ring MXD	SAT	MJJAS	(Not found)
39	tree-SCengland-trw-annual	SC England	-1.40	51.50	950	2009	Tree ring	Tree ring width	Precip.	MJJ	Wilson et al., 2013
40	tree-sodankyla-thi-annual	SodankylaŁ	27.00	67.00	800	2007	Tree ring	Tree height increment	SAT	JJA	Lindholm and Jalkanen, 2011
41	tree-southfin-mxd-annual	Southern Finland	28.50	61.50	800	2000	Tree ring	Tree ring MXD	SAT	MJJAS	Helama et al., 2014
42	tree-SWturkey-trw-annual	SW Turkey	31.00	37.00	1339	1998	Tree ring	Tree ring width	Precip.	MJ	Touchan et al., 2003
43	tree-tyrol-mxd-annual	Tyrol	12.50	48.00	1053	2003	Tree ring	Tree ring MXD	SAT	JAS	Esper et al., 2006
44	d18O-NIshelf-yr-annual	North Icelandic Shelf	66.53	-18.20	953	2000	Ice core	$\delta^{18}O$	SAT	Annual	Reynolds et al., 2016

Supplementary table 2: Ten largest volcanic eruptions from Ortega et al., 2015

Volcano	Country	Date
Unknown		1229
Samalas	Indonesia	1257
Unknown		1285
Huaynaputina	Peru	1600
Parker	Phillippines	1640
Serua	Indonesia	1693
Unknown		1809
Tambora	Indonesia	1815
Cosiguina	Nicaragua	1834
Krakatau	Indonesia	1883

Supplementary table 3: Eleven largest volcanic eruptions from Sigl et al., 2015

Volcano	Country	Date
Unknown		1108
Unknown		1171
Unknown		1230
Samalas	Indonesia	1258
Kuwae	Vanuatu	1458
Huaynaputina	Peru	1601
Parker	Phillippines	1641
Serua	Indonesia	1695
Lakagigar	Iceland	1783
Unknown		1809
Tambora	Indonesia	1815

Supplementary table 4: Variance of the NAO

Method	Mean test variance [range]	Mean train variance [range]	Variance 1000-1855	Variance 1856-1970 for Enet and PCR (1856-1972 for RF and PLS)
RF	0.24 [0.08,0.47]	0.74 [0.64,0.84]	0.16	0.72
Enet	0.26 [0.09,0.96]	0.4 [0.27,0.7]	0.18	0.33
PCR	0.35 [0.06,0.77]	0.56 [0.37,0.86]	0.56	0.62
PLS	0.33 [0.08,1.04]	0.57 [0.4,0.8]	0.17	0.43
NAO (1000-1970)	1.44 [1.23,1.61]	1.46 [0.72,2.27]	1.44	
NAO (1000-1973)	1.44 [1.27,1.57]	1.57 [0.73,2.33]	1.46	

Supplementary table 5: R packages used in ClimIndRec

Package	Reference
glmnet	Friedman et al. 2010
pls	Mevik et al., 2007
randomForest	Liaw and Wienner 2002
ncdf4	Pierce, 2017
stringr	Wickham 2017

Supplementary material 1: Regression methods

We present each method in two steps: model fitting (for training) and reconstruction (for testing). For each method the proxy matrix is denoted as $X \in \mathbb{R}^{n \times p}$ the proxy predictor set and the climate index as $Y \in \mathbb{R}^n$. $X_{(rec)} \in \mathbb{R}^{m \times p}$ is the testing dataset from which a $\mathbb{R}^m$ reconstruction is build using the regression method

1.1 Principal Component Regression (PCR)

1.1.1 Modeling

The Principal Component Regression (Hotelling, 1957) method consists in finding the best linear combination between Y and the Principal Component of X. The Principal Component Analysis (PCA) consists in applying an orthogonal transformation of an initial set of variables, potentially correlated between them, into another set of linearly uncorrelated variables: the Principal Component (Pearson, 1901 ; Hotelling, 1933).

The first step consists in building an orthogonal basis where X will be projected. We define $S \in \mathbb{R}^{p \times p}$, as the empirical estimator of the covariance matrix of X:

$$S = \frac{1}{n} X^T X \in \mathbb{R}^{p \times p} \tag{9}$$

We calculate the orthogonal basis formed by the vectors $v_1, \ldots, v_p$ by diagonalizing S:

$$v_1 = \arg \max_{\substack{v \in \mathbb{R}^p \\ \|v\|=1}} v^T S v \tag{10}$$

$$v_2 = \arg \max_{\substack{v \in \mathbb{R}^p \\ \|v\|=1 \\ \langle v^T v_1 \rangle = 0}} v^T S v \tag{11}$$

$$\ldots \tag{12}$$

$$v_p = \arg \max_{\substack{v \in \mathbb{R}^p \\ \|v\|=1 \\ \langle v^T v_1 \rangle = 0 \\ \ldots \\ \langle v^T v_{p-1} \rangle = 0}} v^T S v \tag{13}$$

$$\tag{14}$$

where $\|v\| = \sqrt{\sum_{j=1}^{p} (v^j)^2}, \forall v \in \mathbb{R}^p$. It is equivalent to maximizing step by step the empirical variance of the projection of X on each orthogonal axis. Indeed, $\forall v \in \mathbb{R}^p$:

$$v^T S v = \frac{1}{n-1} v^T X^T X v = \frac{1}{n-1} (Xv)^T (Xv) = Var_{emp}(Xv) \tag{15}$$

The vectors $(v_k)_{1 \leq k \leq p}$ are called the Empirical Orthogonal Functions (EOFs). It corresponds to the eigenvectors of the covariance matrix and each contains a given part of the spatial variability of the proxy dataset. We attribute them the eigenvalues $(\lambda_k)_{1 \leq k \leq p}$, which corresponds to the initial variance of X translated by each orthogonal projection in the new basis:

$$\lambda_k = Var(Xv_k) = v_k^T S v_k \qquad \forall 1 \leq k \leq p \tag{16}$$

The Principal Components $(u_1, \ldots, u_p)$ are then the projections of X on the EOFs. We denote $V = (v_1, \ldots, v_p)$. We then calculate the Principal Component matrix $U = (u_1, \ldots, u_p)$, defined as:

$$U = XV \in \mathbb{R}^{n \times p} \tag{17}$$

Now, we regress Y on the $q \leq p$ (see subsection 3.1.3) first Principal Component. These q Principal Component are merged in a submatrix of U: $\mathcal{U} = (u_k)_{1 \leq k \leq q}$. The model is given by:

$$Y = \mathcal{U}\beta + \epsilon \tag{18}$$

Where ϵ is a white noise vector of size n.

The best estimator for $\beta = (\beta_1, \ldots, \beta_q)$, is given by the Ordinary Least Squares (OLS) estimator which minimizes $\|\hat{\epsilon}\| = \|Y - \hat{Y}\|$:

$$\hat{\beta}_{OLS} = (\mathcal{U}^T \mathcal{U})^{-1} \mathcal{U}^T Y \tag{19}$$

1.1.2 Reconstruction

We project the testing matrix $X_{(rec)}$ on the pre-calculated orthogonal basis V:

$$U_{(rec)} = X_{(rec)} V \in \mathbb{R}^{m \times p} \tag{20}$$

We then obtain the reconstruction by applying the estimated coefficient vector on the sub-matrix $\mathcal{U}_{(rec)} = (U^1_{(rec)}, \ldots, U^q_{(rec)}) \in \mathbb{R}^{m \times q}$:

$$\hat{Y}_q = \mathcal{U}_{(rec)} \hat{\beta}_{OLS} \in \mathbb{R}^m \tag{21}$$

1.1.3 Control parameters

Here, q is the unique control parameter to be tuned. Here the control parameter vector θ is unidimensional and takes its values in the discrete set $\{i\}_{1 \leq i \leq p}$ and its tuning is obtained by KFCV.

1.2 Partial Least Squares Regression

The PCA keeps most of the initial variance in X in a lower number of vectors. But EOFs $v_1, \ldots, v_p$ are constructed without taking into account any information about the predictand Y. Another possible approach is thus to determine the orthogonal basis in which the empirical covariance between Y and the projection of X on that former is maximized. This is the Partial Least Squares regression (PLSr) method (Wold, 1983).

The first latent variable (LV), denoted $\xi_1 = \sum_{j=1}^{p} v_{1,j} X^j = Xv_1$, where $X \in \mathbb{R}^{n \times p}$ and $v_1 \in \mathbb{R}^p$ is the linear combination of the initial variables $X^1, \ldots, X^p$ such as:

$$v_1 = \arg \max_{\substack{u \in \mathbb{R}^p \\ \|v\|=1}} Cov(Y, Xv), \tag{22}$$

In a similar approach to the PCR, the second LV is $\xi_2 = \sum_{j=1}^{p} v_{2,j} X^j = Xv_2$, orthogonal to ξ_1, such as:

$$v_2 = \arg \max_{\substack{v \in \mathbb{R}^p \\ \|v\|=1 \\ \langle \xi^1, Xv \rangle = 0}} Cov(Y, Xv) \tag{23}$$

And so on, until we have $r \leq p$ LVs. The LV matrix is denoted $\Xi = [\xi_1, \ldots, \xi_p]$. Here, $v_1, \ldots, v_p \in \mathbb{R}^p$, are analogous to the EOFs in PCA, and are here called loadings. The latent variables $\xi_1, \ldots, \xi_r$ respectively correspond to the projection of X on the r loadings.

Finding the loadings is not as trivial as for PCR. Indeed the empirical covariance matrix is not necessary definite positive and thus cannot be diagonalized. We solve this problem by using the algorithm 1 named PLS1. Analogously to the PCR, the method provides various alternative reconstructions depending on the value of r, which corresponds to the number of LVs kept as regressors.

Algorithm 1

procedure PLS1
  $X_0 \leftarrow X$
  for $h = 1, \ldots, r$
    $v_h \leftarrow \frac{X_{h-1}^T Y}{\|X_{h-1}^T Y\|^2}$
    $\xi_h \leftarrow X_{h-1} v_h$
    $X_h = X_{h-1} - \frac{\xi_h \xi_h^T}{\|\xi_h\|^2} X_{h-1}$ *(deflation phase)*
end procedure

Now we regress Y on the $r \leq p$ first LVs. These r LVs are merged in a submatrix of Ξ: $\Psi = (\xi_k)_{1 \leq k \leq r}$. The model is given by:

$$Y = \Psi\beta + \epsilon \tag{24}$$

Where ϵ is a white noise vector of size n.

The best estimator for $\beta = (\beta_1, \ldots, \beta_q)$, is given by the Ordinary Least Squares (OLS) estimator which minimizes $\|\hat{\epsilon}\| = \|Y - \hat{Y}_{q_{KF}}\|$:

$$\hat{\beta}_{OLS} = (\Psi^T \Psi)^{-1} \Psi^T Y \tag{25}$$

1.2.1 Reconstruction

The reconstruction is done in the same way as for PCR. Using the first seen data matrix $X_{(rec)}$ (section 2.4), we project the latter on the pre-calculated orthogonal basis V:

$$\Xi_{(rec)} = X_{(rec)} V \in \mathbb{R}^{m \times p} \tag{26}$$

The reconstruction is obtained by applying the estimated coefficient vector on the sub-matrix $\Psi_{(rec)} = (\xi^1_{(rec)}, \ldots, \xi^r_{(rec)}) \in \mathbb{R}^{m \times r}$:

$$\hat{Y}_r = \Psi_{(rec)} \hat{\beta}_{OLS} \in \mathbb{R}^m \tag{27}$$

1.2.2 Control parameters

For the PLSr method, r is the unique control parameter to be tuned. Analogously to the Principal Component Analysis, the tuning of that latter is obtained by KFCV.

1.3 Elastic Net regression

1.3.1 Modeling

Without using orthogonal transformation of the initial variables as in PCR and PLSr, the most simple predictive model is the multiple linear regression model:

$$Y = X^1\beta_1 + \cdots + X^p\beta_p + \epsilon \tag{28}$$

Where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and $Cov(\epsilon_i, \epsilon_j) = 0$ if $i \neq j$.

The reconstruction of Y, given p proxy records $X^1, \ldots, X^p$ is obtained by the equation:

$$\hat{Y} = X^1\hat{\beta}_1 + \cdots + X^p\hat{\beta}_p \tag{29}$$

$\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)$ are the regression coefficients, which are obtained by the OLS predictor. However, this usual regression model is known to often result in a poor reconstruction accuracy due to the several assumptions made on the original data (Poole, 1971), which are often not verified, such as homoscedasticity and errors normality. Several studies developed regularized (or penalized) regression methods to overcome the OLS defaults. Here we focus on the Elastic Net regression (Zou and Hastie, 2005), which is a combination of the Ridge regression (Hoerl, 1970) and the Lasso regression (Tibshirani et al., 1996). All these methods have been developed to avoid the high variability of the OLS predictor when the number of predictors is relatively high. The Ridge regression shrinks towards zero the estimated coefficients associated to predictors unlinked to the predictand. No predictor selection is made by this method, but the shrunken estimated coefficients modulate the importance of these in the model. By contrast, the Lasso also reduces the variability of the estimates, but in this case by shrinking to zero the estimated coefficients associated to unreliable variables. Hence, a selection is made by rejecting variables associated to coefficients shrunk to zero.

The idea of a regularized (or penalized) regression is to add a threshold constraint using the l_k norm of β: $\|\beta\|_k^k = \sum_{j=1}^k \left|\beta_j\right|^k$. With $k = 1$ in Lasso regression, and $k = 2$ in Ridge regression. The penalized loss functions are given by:

$$L^{ridge}(\beta) = \|Y - \sum_{j=1}^{p} \beta_j X^j\|^2 + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \tag{30}$$

$$L^{lasso}(\beta) = \|Y - \sum_{j=1}^{p} \beta_j X^j\|^2 + \lambda_1 \sum_{j=1}^{p} \left|\beta_j\right| \tag{31}$$

$$L^{enet}(\beta) = \|Y - \sum_{j=1}^{p} \beta_j X^j\|^2 + \lambda_1 \sum_{j=1}^{p} \left|\beta_j\right| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \tag{32}$$

λ_1 penalizes the sum of the absolute values of the regression coefficients while λ_2 penalizes their summed squares. Here, $\lambda_1, \lambda_2 > 0$.

Let $w = (w_j)_{1\leq j\leq p} = (sgn(\beta_j))_{1\leq j\leq p}$, where sgn is the sign function. The loss functions can then be denoted as:

$$L^{ridge} = \|Y - X\beta\|^2 + \lambda_2\beta^T\beta \tag{33}$$

$$L^{lasso} = \|Y - X\beta\|^2 + \lambda_1 w^T\beta \tag{34}$$

$$L^{enet} = \|Y - X\beta\|^2 + \lambda_1 w^T\beta + \lambda_2\beta^T\beta \tag{35}$$

The estimated regression coefficients obtained by minimizing the Lasso and the Ridge loss functions are:

$$\hat{\beta}^{lasso} = (X^TX)^{-1}(X^TY - \frac{\lambda_1}{2}w) \tag{36}$$

$$\hat{\beta}^{ridge} = (X^TX + \lambda_2 I)^{-1}X^TY \tag{37}$$

The Elastic Net regression coefficients are then estimated by minimizing L^{enet}:

$$\hat{\beta}^{enet} = (X^TX + \lambda_2 I)^{-1}(X^TY - \frac{\lambda_1}{2}w) \tag{38}$$

An alternative way to write this equation as a linear combination of $\hat{\beta}^{lasso}$ and $\hat{\beta}^{ridge}$ is:

$$\hat{\beta}^{enet} = (X^TX + (1-\alpha)\lambda I)^{-1}(X^TY - \frac{\alpha\lambda}{2}w) \tag{39}$$

where $\alpha \in [0, 1]$. If $\alpha = 1$, a Rigde regression is applied, and if $\alpha = 0$, we apply a Lasso regression.

1.3.2 Reconstruction

The reconstruction is obtained by applying the estimated regression coefficients $\hat{\beta}^{enet}$ on the validation variables $X^1_{val}, \ldots, X^p_{val}$:

$$\hat{Y}_{\lambda,\alpha} = \sum_{j=1}^{p} X^j_{(val)}\hat{\beta}^{enet}_j \tag{40}$$

1.3.3 Hyperarameters

For Enet method, the control parameters are λ and α such that $\theta = (\alpha, \lambda)$. The latter controls the relative balance between the Lasso and Ridge regularization, while the former controls the overall intensity of regularization as λ_1 (resp. λ_2) in Lasso (resp. Ridge regularization). A high α suggests a dense model with many but small non-zero coefficients. A low α suggests a sparse model with many zero coefficients. In our case, since we want a general methodology performant for each random split, we apply two simultaneous and nested KFCV to find the best estimated pair $(\hat{\lambda}, \hat{\alpha})$.

Since λ and α take respectively their values in the continuous sets $\mathbb{R}^p$ and $[0, 1]$, we have to discretize their respective intervals for the parameter estimation. The finer these discretizations are, the more reliable the parameters will be, at the expanse of the computational time. In the actual version of ClimIndRec (version 1.0) the set where α is tuned is $(0.1i)_{0\leq ileq10}$ and the set where λ is tuned is $(0.2i)_{0\leq i\leq 25}$. It can be easily changed by modifying the code main R script of ClimIndRec.

1.4 Random Forest regression

The random forest has been introduced by Breiman (2001) as a learning method for regression. The method relies on using randomization to minimize the reconstruction uncertainty given by regression trees. Random forests encompass a large variety of regression methods (Breiman, 2001). Here, we present the most classical kind of random forests known as random-input random forests (Breiman, 2001).

1.4.1 Modeling

First we have to define regression trees. We denote each set of predictand/predictors by $\{Y_i, X_i)_{1\leq i\leq n}\}$ where $X_i = (X_i^1, \ldots, X_i^p)$, is the ensemble of proxy records for the i^{th} time step, and Y_i the corresponding values of the climate index at the same time step, $\forall 1 \leq i \leq p$. All the observations, $\{Y_i, X_i)_{1\leq i\leq n}\}$, $\forall 1 \leq i \leq p$, are put on the root of the tree. The first step consists in cutting that root in two child nodes. A cut is defined as:

$$\left\{X^j \leq d\right\} \cup \left\{X^j \geq d\right\} \tag{41}$$

where $j = \{1, \ldots, p\}$ and $d \in \mathbb{R}$. Cutting a node with $\left\{X^j \leq d\right\} \cup \left\{X^j \geq d\right\}$ means that all observations with a j^{th} variable lower than d are placed in the left child node. Hence, all observations with a j^{th} variable greater than d are placed in the right child node. The method selects the best pair (j, d) which minimize a loss function. Here, we aim at minimizing the variance of the child nodes. The variance of a given node t is defined as:

$$\sum_{i:X_i\in t} (Y_i - \bar{Y}_t)^2 \tag{42}$$

where $\bar{Y}_t$ is the averaged Y_i in the node t.

The same procedure is then applied recursively to the next child nodes using the same variables until a certain stop criterion is reached. The procedure automatically stops if each node contains a unique observation. Hence, the maximal depth of a regression tree is $n - 1$. An illustration of such tree is presented in Fig 5.

A random-input regression tree is used here. This is a particular case of regression trees, in which a set of $m < p$ variables is randomly preselected before applying the regression tree. A large number K of random-input trees is computed. For each tree, we randomly select $m < p$ variables with probability $\frac{1}{p}$ and we apply the method until it reaches its maximal depth.

1.4.2 Reconstruction

The reconstruction is obtained by splitting each testing series in the different trees. In each tree, the estimation attributed to an observation is the empirical average of Y inside the node where the corresponding observation ends up, given the cut made on the corresponding predictors. For each testing series, the K reconstructions are averaged to give the final reconstruction.

1.4.3 Parameters

A priori, this method requires the optimization of two parameters: the number of trees K and the number of variables selected for each tree m. In practice K does not require tuning, as long as the number of trees is sufficiently high given p, which guarantees convergent results for any value of m (Breiman, 2001). m is then the only parameter to optimize. The KFCV is then applied on m with a high K (here set to 1000), to select empirically the most efficient model.

Supplementary material 2: Statistical test for correlation significance

The statistical test we use in all the study as been firstly proposed by Bretherton et al. (1999) to avoid the individual autocorrelation effects on the correlation between two series. This is done by adjusting the degree of freedom. However, a simplification of this test has been proposed by McCarthy et al. (2015) by only using the first order autocorrelations to modify the degree of freedom.

Let $X = (X_t)_{t\in T}$ and $Y = (Y_t)_{t\in T}$ two time series of same length. The correlation between the two series is given by:

$$r = cor(X, Y) = \frac{cov(X, Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}} \tag{43}$$

We denote $a_1^{(X)}$ and $a_1^{(Y)}$ the first order lag of the respective autocorrelation functions of X and Y. The effective number of degrees of freedom (Bretherton et al., 1999) is then given by :

$$N_{eff} = N_{obs} \cdot \frac{1 - a_1^{(X)} \cdot a_1^{(Y)}}{1 + a_1^{(X)} \dot{a}_1^{(Y)}} \tag{44}$$

The statistics is then calculated as :

$$t^{stat} = \sqrt{N_{eff}} \cdot \frac{r}{\sqrt{1 - r^2}} \tag{45}$$

For $\alpha \in]0, 1[$, the statistic t^{stat} is compared to the $1 - \frac{\alpha}{2}$ order quantile of a Student distribution with N_{eff} degrees of freedom.

Supplementary material 3: Uncertainties

The uncertainties we provide are calculated as in Ortega et al. (2015) but using the residuals calculated for the final moel. These regression uncertainties are represented by the standard errors (s.e.) of the regression, calculated as the root of the sum of the squared residuals over the training periods divided by the degree of freedom:

$$s.e = \sqrt{\frac{\sum_{i=1}^{n}(Y - \hat{Y})^2}{n-2}} \tag{46}$$

Where n_{train} is the length of the learning sample, Y the true values of the NAO index over the learning period, and $\hat{Y}$ the fitted NAO by the regression model over the training period. An uncertainty band $2 * s.e.$ is calculated which is our estimate of the total uncertainty range of the reconstruction.

Supplementary material 4: Superposed Epoch Analysis

For both sets of volcanic eruptions (3.2.2), each member of the composite is extracted from year N-1 to year N+4 from the corresponding NAO reconstruction, N standing for the year of the eruption. Hence, they are centred to the year of the eruption and averaged at each time step to perform the composite timeseries presented in Fig 12. The significance levels are calculated using a Monte Carlo approach as proposed by Rao et al. (2019). We randomly draw 10000 composite of 11 (for Sigl. et al, 2015, see Fig 12) and 10 (for Ortega et al, 2015, see Fig 12) N-1 to N+4 sub-timeseries of the corresponding NAO index. As for the real composite, each member of the composites is centred to the year N and are averaged over time. Hence, for each time step, we keep the 90^{th} percentile among the 10000 sorted composite values as the 90% confidence level presented Fig 12.