Auxiliary Material

1. Model data description:

We used model output from phase 5 of the Coupled Model Intercomparison Project (CMIP5). A full list of modeling groups participating in CMIP5 is given at http://cmip-pcmdi.llnl.gov/cmip5/docs/CMIP5_modeling groups.pdf. At the time our research was performed, the CMIP5 archive was not fully populated with model results. We analyzed results from 12 different CMIP5 models, contributed by 10 different research groups (see Table S1). We mostly use models having an ensemble size of at least four members. Results with models having less members lack robustness in their determination of the forced responses, in particular in the case of low signal to noise ratio. As noted from the main text, we analyzed different types of simulation:

1. Simulations with estimated historical changes in human and/or natural external forcings. This includes the ALL ensemble (12 models) forced by all anthropogenic and natural forcings, the GHG ensemble (6 models) by greenhouse gases only, the AER ensemble (4 models) by aerosols only and the NAT one (6 models) by natural forcings only.

2. Pre-industrial control simulations with no changes in external influences on climate, which provide information on internal climate variability.

Details of all the simulations used are provided in Table S2. Only four models provided an ensemble with aerosol-only forcing: can_esm2, csiro_mk3.6.0, giss-e2-r and gfdl_cm3. Using the linear additivity assumption, we reconstruct an aerosol-only ensemble for hadgem2_es, giss_e2_h and cnrm_cm5. For hadgem2-es, we use the residual of the ALL ensemble after removing the contribution of the first signal-to-noise maximizing EOF of an ensemble with all forcings except anthropogenic tropospheric aerosols [Booth et al 2012]. For the other models (cnrm-cm5, giss-e2-h) , we estimate it by using the residuals of the ANT ensemble members after removing the best-estimate of the response to GHG forcing. The latter is taken as the contribution of the first signal-to-noise maximizing EOF of the GHG ensemble. Note here that we assume that changes due to other anthropogenic forcings possibly included in the ANT ensemble (such as land use or ozone) can be neglected compared to the aerosol response.

Figure S1 shows that the observed SSTs are always within the historical ensemble (with all external forcings) spread. It also indicates that substantial uncertainty remains regarding the forced response, particularly that of anthropogenic aerosols (compare hadgem2-es and gfdl-cm3 with the multi-model mean). Note that the forced response spread includes both forcing and structural uncertainties. The latter is related to model errors in representing the mechanisms underlying the forced response.

2. Filtering, analysis of variance and signal-to-noise maximizing EOF:

a. Decadal filter

Observed and simulated temperature data have been low-pass filtered in order to emphasize decadal time scale. The binomial filter used has 13 weights 1/576 [1-6-19-42-71-96-106-96-71-42-19-6-1]. For yearly data the half-amplitude point is about a 12-year period, and the half-power point is 16 years.

b. ANOVA analysis

The ANOVA analysis gives us at the grid-point scale the partition of decadal variance between forced and internal variability. It requires a couple of assumptions [Rowell and Zwiers 1999]. Regarding the 19th and 20th centuries, it seems reasonable to assume that the forced and internal components are uncorrelated (independence assumption). A stationary forcing assumption is also required. While some of the forcings are not stationary (e.g GHG forcing), we assume that the analysis can reasonably be applied using only the 1850-1960 period where the anthropogenic forcing can be considered as a relatively small perturbation.

c. Signal-to-noise maximizing EOF

The method applies a spatial prewhitenening transformation to the model data thus removing the spatial correlations in internal climate variability present in any ensemble mean with a finite number of members. We apply it here to low-pass filtered annual mean surface temperature from various CMIP5 historical simulation multi-model ensembles. While a global domain is usually used within the EOF analysis to estimate the forced response (see Ting et al. [2009]; Ting et al. [2011]), here we choose a North Atlantic domain (the same as that of the AMV-NA index) in order to better discriminate between the regional fingerprints of the various forcings. The pre-whitening step requires a truncation level k to confine the analysis to well sampled directions of the internal variability phase space. Using a simple cumulative signal to noise variance ratio diagnostic [Venske et al., 1999], we choose a truncation of 15 for all the surface temperature analysis and test the sensitivity of results to this choice (see below). The signal to noise maximizing EOF decomposition ranks eigenvectors by signal-to-noise ratios (and not by variance) leading to a possible degeneracy between different forcings with similar signal-to-noise ratios. An additional step of variance-maximizing rotation is then needed to resolve any remaining degeneracy [Allen and Smith, 1997]. We systematically apply this rotation step to any multi-signal analysis to remove any potential degeneracy. We apply the signal to noise maximizing EOF to all selected historical ensembles and models. Figure S2 shows the time evolution and spatial pattern of the North Atlantic surface temperature forced response to all combined forcings for twelve CMIP5 models. Note that while models reasonably agree as to the time evolution of the forced response, its spatial pattern exhibit large regional differences among the different models. Some models (e.g gfdl-cm3) show strong sensitivity to stratospheric aerosols emitted during volcanic eruptions.

Figure S3 address the sensitivity of the detected forced response to the truncation used to select the highest-ranked noise EOFs when constructing the pre-whitening operator. We only show results for the cnrm-cm5 model as other models show similar behavior. The temporal characteristics of the forced response show only marginal sensitivity to variations in the truncation level k, particularly for the first mode (and the second mode as well if k greater or equal to 10). As the spatial structure of the dominant forced response is estimated by computing regression coefficient maps of surface temperature regressed onto the first mode time series, it is therefore stable against the choice of truncation.

Supplementary tables

Table S1: Modeling center information and official acronyms of the CMIP5 models used in this study.

Model	Country	Modeling center
1 gfdl_cm3	USA	Geophysical Fluid Dynamics Laboratory
2 can_esm2	Canada	Canadian Centre for Climate Modelling and Analysis
3 ccsm4	USA	National Center for Atmospheric Research
4 cnrm_cm5	France	Centre National de Recherches Météorologiques / Centre Européen de Recherche et Formation Avancées en Calcul Scientifique
5 csiro_mk3.6.0	Australia	Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence
6 giss_e2_r	USA	NASA Goddard Institute for Space Studies
7 giss_e2_h	USA	NASA Goddard Institute for Space Studies
8 hadgem2_es	UK	Met. Office Hadley Centre
9 hadcm3	UK	Met. Office Hadley Centre
10 ipsl_cm5a_lr	France	Institut Pierre-Simon Laplace
11 miroc5	Japan	Japan Agency for Marine-Earth Science and Technology Atmosphere and Ocean Research Institute (the University of Tokyo) and National Institute for Environmental Studies
12 mri_cgcm3	Japan	Meteorological Research Institute

Table S2: Type of ensembles of simulations used. For each model, the ensemble member identifiers (which indicates the ensemble size), either simulations start and end dates (historical) or length (preindustrial, in years) are provided. A cross indicates that the data was not present or not used (too few members). Details about CMIP5 are available at http://cmip.pcmdi.llnl.gov/cmip5/documents.html.

Model	Historical	Historical GHG	Historical Nat	Historical Misc	Historical AA	piCTRL
gfdl_cm3	r[1-5]i1p1 1850-2005	r[1,3,5]i1p1 1860-2005	r[1,3,5]i1p1 1860-2005		r[1,3,5]i1p1 [g] 1860-2005	r1i1p1 500 yrs
can_esm2	r[1-5]i1p1 1850-2012	r[1-5]i1p1 1850-2012	r[1-5]i1p1 1850-2012	r[1-5]i1p4 [a] 1850-2012	X	r1i1p1 995yrs
ccsm4	r[1-6]i1p1 1850-2005	X	X	X	X	r1i1p1 1300yrs
cnrm_cm5	r[1-10]i1p1 1850-2012	r[1-6]i1p1 1850-2012	r[1-6]i1p1 1850-2012	r[1-10]i1p1 [b] 1850-2012	X	r1i1p1 1000yrs
csiro_mk3.6.0	r[1-10]i1p1 1850-2005	r[1-5]i1p1 1850-2012	r[1-5]i1p1 1850-2012	r[1-5]i1p1 [c] 1850-2012	r[1-5]i1p1 1850-2012	r1i1p1 500yrs
giss_e2_r	r[1-5]i1p1 1850-2012	r[1-5]i1p1 1850-2012	r[1-5]i1p1 1850-2012	r[1-5]i1p109 1850-2012 [d]	r[1-5]i1p107 1850-2005	r1i1p1 1200yrs
giss_e2_h	r[1-5]i1p1 1850-2012	r[1-5]i1p1 1850-2012	r[1-5]i1p1 1850-2012	r[1-5]i1p109 1850-2012 [e]	X	r1i1p1 1470yrs
hadgem2_es	r[1-4]i1p1 1860-2004	r[1-4]i1p1 1860-2012	r[1-4]i1p1 1860-2012	r[1-4]i1p1 [f] 1860-2004	X	r1i1p1 474yrs
hadcm3	r[1-10]i1p1 1860-2004	X	X	X	X	X
ipsl_cm5a_lr	r[1-4]i1p1 1850-2005	X	X	X	X	r1i1p1 1000yrs
miroc5	r[1-4]i1p1 1850-2012	X	X	X	X	r1i1p1 670yrs
mri_cgcm3	r[1-5]i1p1 1850-2005	X	X	X	X	r1i1p1 500yrs

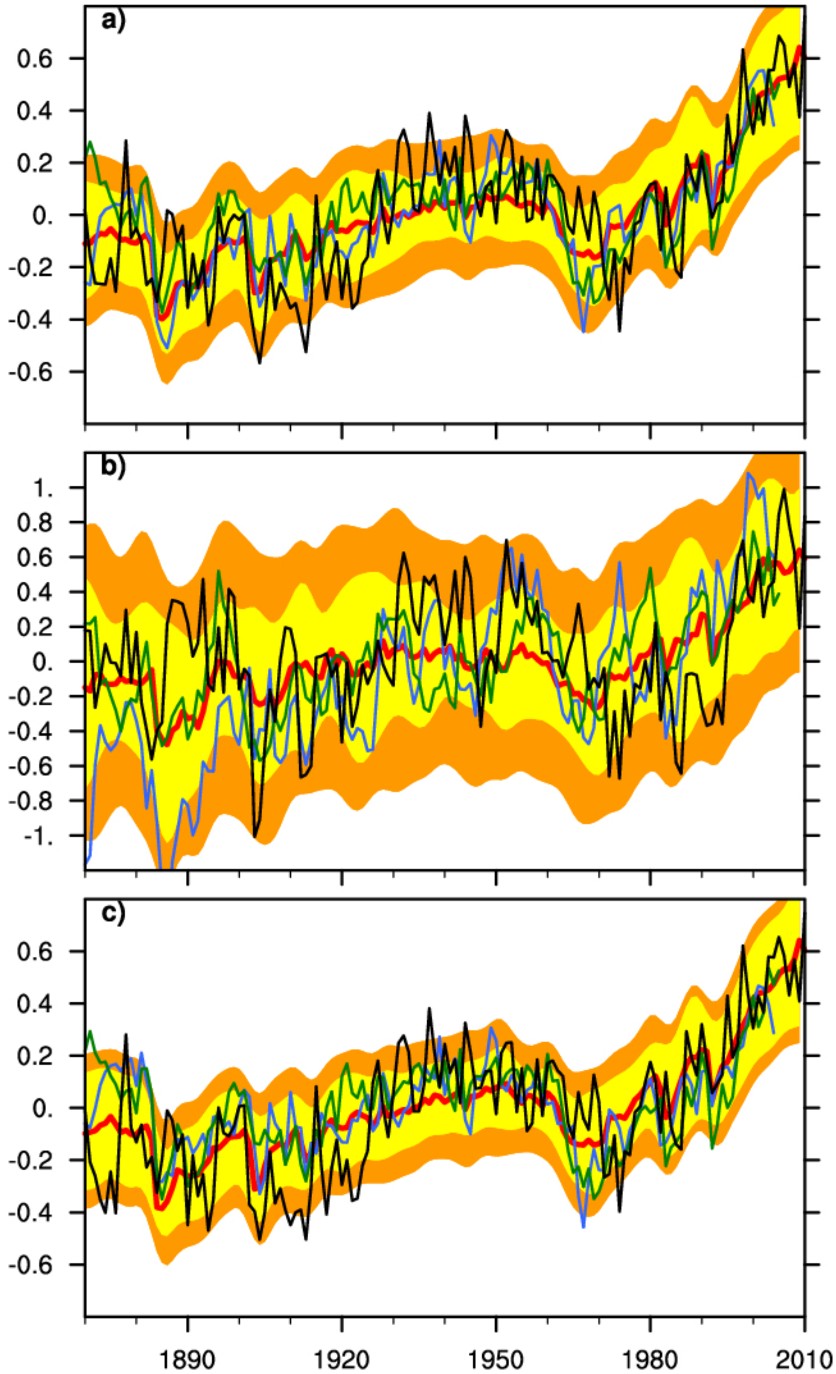
Footnotes: **a.** : only anthropogenic aerosol forcings **b.,c.,d.,e.** : all anthropogenic forcings **f.** : all anthropogenic forcings except anthropogenic aerosols kept constant at their 1860 values, data provided by Paul Halloran, private communication **g.** : data provided by Yi Ming, private communication.

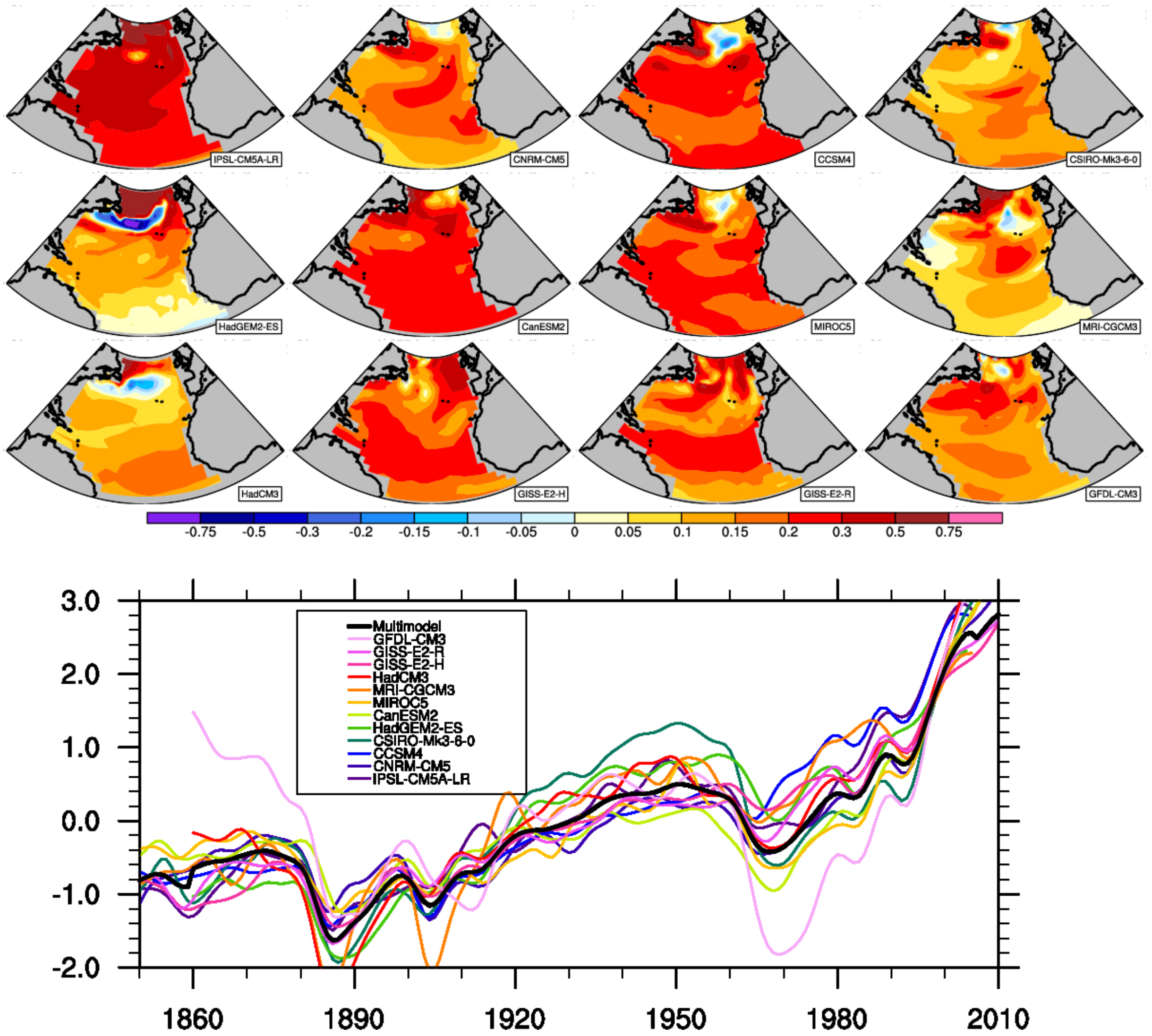
Supplementary figure captions

Figure S1: Evolution of the AMV indexes: anomalous annual mean NASST averaged over a) 0-60°N, b) 45°N-60°N, and c) 0-45°N, from observations (black line), multi-model mean ALL simulations (red line), the hadgem2-es (blue line) and gfdl-cm3 (green line) ALL ensemble means. The base period used to estimate anomalies is 1901-2000. Yellow shading represents ±1.65 standard deviation (s.d) of the models ensemble mean spread and gives an estimate of the forced response uncertainty. Orange shading represents ±1.65 s.d of the individual members spread from all models. The shading time series are lightly smoothed for better viewing..

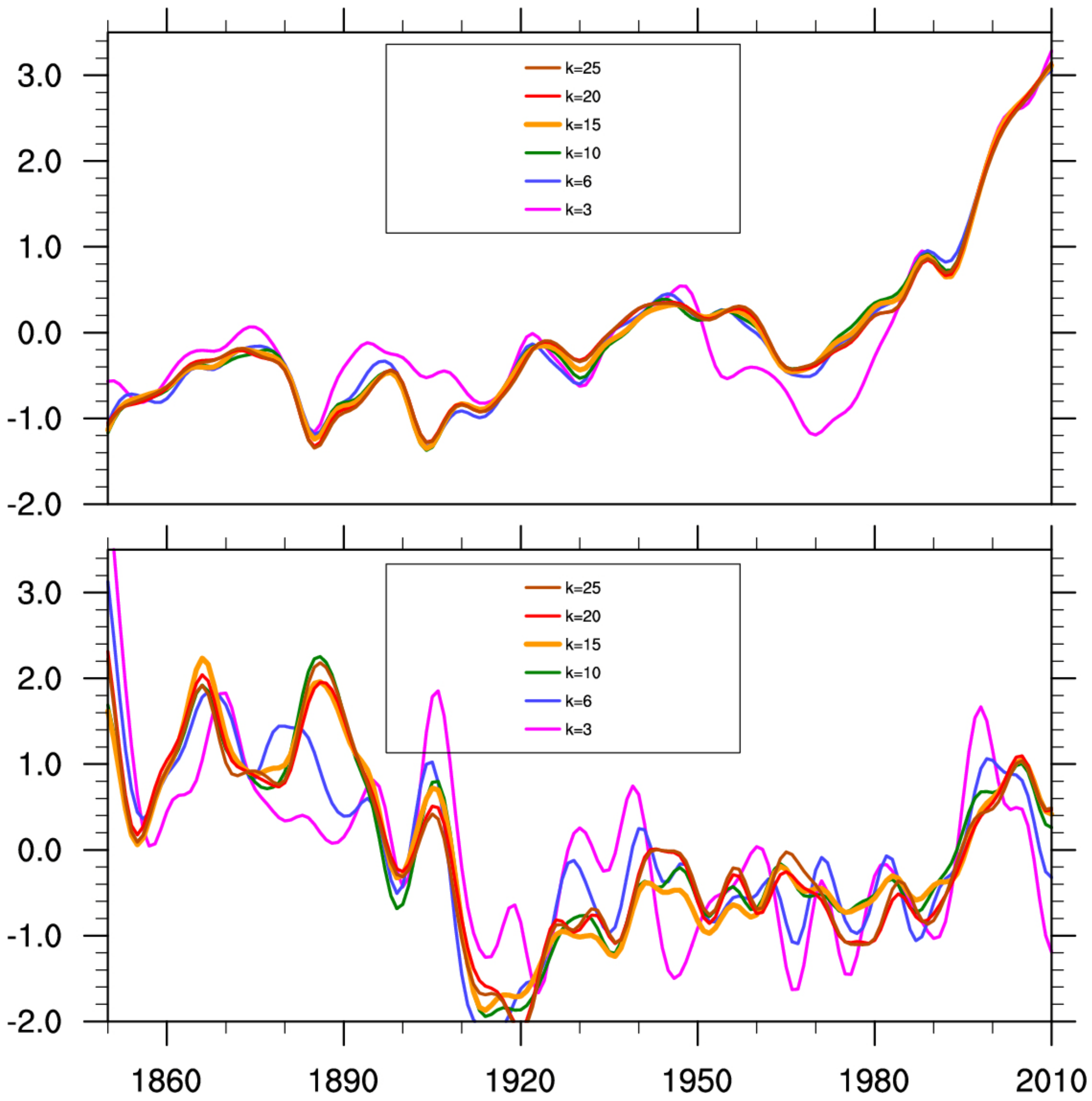
Figure S2: North Atlantic surface temperature forced response using the ALL historical experiments: a) spatial structure for twelve CMIP5 models. b) time evolution of the North Atlantic surface temperature forced response taken as the first mode standardized signal-to-noise principal component (PC1). The black line shows the twelve-model average first principal component. The spatial response is estimated using regressions of annual mean lowpass-filtered surface temperature (first filtered through a projection on the first 15 noise EOFs, see Venzke et al. 1999) on PC1. In a), Unit is K per PC1 standard deviation.

Figure S3: Sensitivity of signal-to-noise maximizing principal components to truncation onto noise EOFs for the cnrm-cm5 North Atlantic surface temperature analysis: first (upper panel) and second (lower panel) principal components. Color lines correspond to different truncations.


a)
0.6
0.4
0.2
0.
-0.2
-0.4
-0.6
b)
1.
0.8
0.6
0.4
0.2
0.
-0.2
-0.4
-0.6
-0.8
-1.
c)
0.6
0.4
0.2
0.
-0.2
-0.4
-0.6
1890
1920
1950
1980
2010



IPSL-CM5A-LR
CNRM-CM5
CCSM4
CSIRO-Mk3-6-0
HadGEM2-ES
CanESM2
MIROC5
MRI-CGCM3
HadCM3
GISS-E2-H
GISS-E2-R
GFDL-CM3
-0.75 -0.5 -0.3 -0.2 -0.15 -0.1 -0.05 0 0.05 0.1 0.15 0.2 0.3 0.5 0.75
Multimodel
GFDL-CM3
GISS-E2-R
GISS-E2-H
HadCM3
MRI-CGCM3
MIROC5
CanESM2
HadGEM2-ES
CSIRO-Mk3-6-0
CCSM4
CNRM-CM5
IPSL-CM5A-LR
3.0
2.0
1.0
0.0
-1.0
-2.0
1860
1890
1920
1950
1980
2010



k=25
k=20
k=15
k=10
k=6
k=3
3.0
2.0
1.0
0.0
-1.0
-2.0
k=25
k=20
k=15
k=10
k=6
k=3
3.0
2.0
1.0
0.0
-1.0
-2.0
1860
1890
1920
1950
1980
2010