The Spatiotemporal Evolution of the Temperature Forced Variance Ratio over the Common Era

Master's Thesis

Faculty of Science, University of Bern

handed in by
Louis Frey
2019

Supervisor:
Dr Raphael Neukom
Oeschger Centre for Climate Change Research and Institute of Geography

Co-Supervisor:
Prof Dr Martin Grosjean
Oeschger Centre for Climate Change Research and Institute of Geography

Co-Supervisor:
Dr Jörg Franke
Oeschger Centre for Climate Change Research and Institute of Geography

Table of Contents

List of Figures

List of Tables

Abbreviations

20CRv2	Twentieth Century Reanalysis data set, version 2
AICC	Akaike's corrected Information Criterion
AMO	Atlantic Multidecadal Oscillation
AMV	Atlantic Multidecadal Variability
Analogue	Analogue Method Reconstruction
ANOVA	Analysis of Variance
AOD	Aerosol Optical Depth
AOGCM	Coupled Atmosphere–Ocean General Circulation Model
ARMA	Autoregressive Moving Average
BCE	Before Common Era
CCA	Canonical Correlation Analysis Reconstruction Method
CCSM4	Community Climate System Model version 4
CDO	Climate Data Operator Software
CE	Common Era
CESM	Community Earth System Model - Last Millennium Ensemble
CFR	Climate Field Reconstruction
CMIP5	Coupled Model Intercomparison Project Phase 5
CMIP6	Coupled Model Intercomparison Project Phase 6
CPS	Composite Plus Scale Reconstruction Method
CRUTEM4	Climatic Research Unit Temperature database, version 4
DA	Data Assimilation Reconstruction Method
ENSO	El Niño / Southern Oscillation Phenomenon
FVR	Forced Variance Ratio

GFS	Global Forecast System of the U.S. National Centers for Environmental Prediction
GHG	Greenhouse Gas
GM	Global Mean
GM-EM	Area-Weighted Global Mean and Ensemble Mean
GM-MEM	Global-Mean Model Ensemble Mean
GraphEM	Graphical Model with Expectation Maximization Algorithm Reconstruction Method
HadCRUT4	Hadley Centre and Climatic Research Unit Temperature data set, version 4
HadCM3	Hadley Centre Coupled Model, version 3
HadSST3	Hadley Centre Sea Surface Temperature, version 3
ISPDv2	International Surface Pressure Databank, version 2
NAO	North Atlantic Oscillation
PCA	Principal Component Analysis
PC	Principal Component
PCR	Principal Component Regression Reconstruction Method
PDO	Pacific Decadal Oscillation
PMIP3	Paleoclimate Modelling Intercomparison Project Phase 3
MEM	Model Ensemble Mean
SAM	Southern Annular Mode
SATIRE-H	Spectral and Total Irradiance Reconstruction, version H
ST	Surface Temperature
TSI	Total Solar Irradiance

Abstract

A quantification of the relative importance of internally generated and externally forced variability is of keen interest in the climate sciences. For example, the phenomenon of the global warming hiatus calls for such a quantification. Furthermore, also the subjects of detection and attribution of climate change, climate scenario projection uncertainty and the identification of optimal locations for paleoclimate reconstruction could be informed by such a quantification. Even though several studies were concerned before with the relationship between internal and external variability, none of them provided a spatially explicit quantification of that relationship with global coverage and temporal coverage exceeding the instrumental period. Such a quantification is made in the present thesis, and is termed Forced Variance Ratio (FVR), the ratio of forced to total variance. FVR is computed for instrumental and reanalysis, climate model and Climate Field Reconstruction (CFR) surface temperature data sets.

This is achieved by using a linear regression approach, where surface temperature time series are regressed against time series representing forcing. The resulting coefficient of determination R^2 is then the fraction of variance in temperature that is explained by the forcing predictor, i.e. the FVR. Those calculations are performed for all grid points on a regular 5° longitude by 5° latitude grid, resulting in global maps of FVR. Moreover, the calculations are also done for sequences of moving windows, yielding the temporal evolution of those maps.

Results indicate that Global Mean (GM) FVR exhibits substantial variability in time. The two most distinctive features of that variability are a significant drop of FVR in the 16^{th} and 17^{th} centuries and a consecutive strong rise in the 19^{th} and 20^{th} centuries, most likely related to anthropogenic GHG forcing. Mostly upward trends that have been detected in time series of GM FVR are likely spurious. Broad agreement in GM FVR could be found among data sets, with the FVR values of the different data sets mostly falling within the same ranges. The different methodologies were found to produce somewhat inconsistent values of GM FVR, due to multicollinearity and large sample sizes, among others. The zonal-mean FVR of the instrumental and reanalysis data sets agreed in maxima in the Arctic and minima in the mid-latitudes of both hemispheres. The two climate model data sets consistently exhibited a global maximum of zonal-mean FVR in the tropics, local minima in the mid-latitudes and local maxima in the high latitudes of both hemispheres. On the other hand, the CFR data sets revealed mostly uniform distributions of zonal-mean FVR. The spatial pattern of FVR was found to exhibit a complex temporal evolution. PCA of FVR fields revealed globally uniform fields of PC1 loadings, a result found to be exceedingly robust with regard to time scale, methodology and data set. This result indicates that FVR varies in a globally uniform manner, on a first order. The spatial pattern of FVR itself was found to strongly depend on methodology, yet explanations for many of the differences could be provided. Most robust estimates of GM FVR are 0.3–0.4 over the last millennium on the climatological time scale and small (GM FVR < 0.05) over the period 1401–2000 on the annual time scale.

1 Introduction

1.1 Relevance and Motivation

The relationship between internal variability and external forcing is of central interest in the climate sciences. An example is the 21st century global warming hiatus, a slowdown in observed global warming (Easterling & Wehner 2009). This phenomenon has provoked a lively debate about whether such a slowdown is due to internal variability or might be related to external forcing (Steinman et al. 2015). This debate has far-reaching implications for anthropogenic climate change and its mitigation (e.g. Kosaka & Xie 2013; Crowley et al. 2014; Trenberth 2015).

Also, the relative importance of internal and external variability has implications for detection and attribution of climate change. On the global scale, IPCC 2013 states that a response of the climate system to external forcing could be successfully detected and attributed to anthropogenic causes. However, on the regional scale, detection and attribution of climate change is more difficult (Stott & Tett 1998) because the contribution of internal variability is higher (Hawkins & Sutton 2009). However, it is expected that detection of anthropogenic climate change should become possible on smaller scales as the strength of the anthropogenic warming signal increases (Stott & Tett 1998). Hence, information on which regions display higher influence of external forcing can contribute to our understanding of where anthropogenic climate change is easier or harder to detect.

Similar considerations apply to projections of climate change and their associated uncertainties. This topic is of keen interest for strategies to mitigate and adapt to climate change (Deser et al. 2012). Uncertainty in climate change projections has three major components: climate model uncertainty, uncertainty in future forcing and internal variability (Hawkins & Sutton 2009; Hawkins & Sutton 2011). Generally, the importance of internal variability decreases with projection lead time, while the importance of forcing, respectively climate scenario projection uncertainty increases. With smaller spatial scale, the importance of internal variability increases, while the importance of scenario projection uncertainty remains substantial (Hawkins & Sutton 2009; Hawkins & Sutton 2011). Therefore, spatial information about the relative importance of internal and external variability could enhance our understanding of the spatial properties of uncertainty of climate projections.

Finally, spatially explicit information about the relative importance of internal and external variability is of interest to paleoclimatology. Bradley 1996 investigated climate model data to identify optimal locations for the reconstruction of past temperature. More precisely, he was interested in locations where the local temperature signal correlated well with the GM temperature. It seems likely that the forced response is reflected most strongly in both the GM signal and in signals from individual locations with a strong influence of forcing. Therefore, those strongly forced locations can also be expected to correlate well with the GM signal and are, therefore, well suited for providing proxies for GM paleoclimate reconstruction.

1.2 Definition and State of Research

All the examples given above in 1.1 point to the importance of climatic variability, internal and external. But what do we understand by those terms? IPCC 2013 gives the following definition:

> "Climate variability refers to variations in the mean state and other statistics (such as standard deviations, the occurrence of extremes, etc.) of the climate on all spatial and temporal scales beyond that of individual weather events. Variability may be due to natural internal processes within the climate system (internal variability), or to variations in natural or anthropogenic external forcing (external variability)."

The traditional examples of internal variability are the modes of internal climate variability, such as the El Niño / Southern Oscillation Phenomenon (ENSO) (Bjerknes 1966; McPhaden et al. 2006). On the other hand, external variability is due to forcing factors external to the climate system, such as solar variability, volcanic eruptions and anthropogenic Greenhouse Gas (GHG) emissions (IPCC 2013). Since internal and external variability are so central concepts in the climate sciences, a quantification of their relative importance is of keen interest. Such a quantification is undertaken in this thesis and is termed FVR, in analogy to Lyu et al. 2015. The FVR is defined here as the ratio of forced variance to total variance, according to the following equation:

$$FVR = \frac{Var(X)_{forced}}{Var(X)_{total}} = \frac{Var(X)_{forced}}{Var(X)_{forced} + Var(X)_{internal}} \tag{1.1}$$

where $Var()$ denotes the variance and X denotes any climatic variable. However, in this analysis X was constrained to surface temperature.

Several studies have made quantifications of the relationship between internal and external variability before. A first group of studies has made use of (multiple) linear regression calculations to separate internal and external variability (Dai et al. 2015; Frankcombe et al. 2015; Wang et al. 2017; Wang et al. 2018; Faden 2018). A second group of studies implemented regression-based detection and attribution (D&A) approaches (Schurer et al. 2013a; PAGES 2k Consortium 2019). This method was developed by Hasselmann 1979; Bell 1982; Bell 1986; Hasselmann 1993; North et al. 1995; North & Kim 1995; Hegerl & North 1997. Reviews can be found in Stott et al. 2010; Hegerl & Zwiers 2011; Lean 2018. A third group of studies employed a sequential removal approach (Thompson et al. 2009; Crowley et al. 2014). A fourth group of studies performed linear detrending using a Model Ensemble Mean (MEM) (Barkhordarian et al. 2012a; Barkhordarian 2012; Barkhordarian et al. 2012b; Barkhordarian et al. 2013; Barkhordarian et al. 2016; Barkhordarian et al. 2018). Further studies have subtracted multi-model means to obtain internal variability (Kosaka & Xie 2016) or performed a method they termed "low-frequency component analysis" (Wills et al. 2018). A final group of studies has adopted an approach termed "Analysis of Variance", which works within a model ensemble framework (Harzallah & Sadourny 1995; Rowell et al. 1995; Ting et al. 2009; Matsumura et al. 2010; Hu et al. 2012; Lyu et al. 2015).

Even though the studies mentioned above are all concerned with the relation between internal variability and external forcing, some of them have not made an explicit quantification of FVR (Thompson et al. 2009; Schurer et al. 2013a; Crowley et al. 2014; Dai et al. 2015; Kosaka & Xie 2016;

Barkhordarian et al. 2018; Faden 2018). Others have looked only at a limited time span, mostly during the instrumental period (Harzallah & Sadourny 1995; Rowell et al. 1995; Thompson et al. 2009; Ting et al. 2009; Matsumura et al. 2010; Lyu et al. 2015; Kosaka & Xie 2016; Wills et al. 2018). A further group of studies has examined only time series of global (Thompson et al. 2009) or regional (Wang et al. 2018; Schurer et al. 2013a) means with no examination of spatial patterns. The study by Wang et al. 2017 has placed a focus on Atlantic Multidecadal Variability (AMV), also without examination of spatial patterns. Wills et al. 2018 have examined the spatial structure, but their analysis remained limited to the Pacific ocean and the 20$^{\text{th}}$ and 21$^{\text{st}}$ centuries. The studies that made use of the "Analysis of Variance"-approach only work in a model ensemble framework and can thus not be extended to incorporate also CFR, instrumental and reanalysis data sets.

Therefore, a spatially explicit quantification of FVR with global coverage and temporal coverage exceeding the instrumental period has not yet been undertaken. The present thesis aims to address that knowledge gap by spatially explicitly quantifying the FVR globally and over the Common Era (CE).

1.3 Research Questions and Hypotheses

To explore the properties of the global fields of FVR that are produced in this thesis and to guide the research, the following research questions QA1–QB3 are formulated:

QA1 Is the GM FVR stable in time?

QA2 Is the GM FVR consistent across data sets?

QA3 Is the GM FVR consistent across the different regression methods (section 3.2)?

QB1 Is the spatial structure of FVR stable in time?

QB2 Is the spatial structure of FVR consistent across data sets?

QB3 Is the spatial structure of FVR consistent across the different regression methods?

To answer question QA1, time series of GM FVR were examined, both qualitatively (section 4.1) and quantitatively (section 4.2). To answer questions QA2 and QA3, distributions of FVR were compared across data sets and regression methods, also both qualitatively and quantitatively. To answer the questions QB1 to QB3, the fields of FVR were analyzed qualitatively using map plots and Principal Component Analysis (PCA) results. In addition to the analyses described before, an intermediate results chapter (4.5) treats the zonal-mean FVR. For this part, no research questions have been formulated and the research remained rather exploratory.

Suitable hypotheses HA1–HB3 corresponding to research questions QA1–QB3 can be formulated as follows:

HA1 The GM FVR is stable in time.

HA2 The GM FVR is consistent across data sets.

HA3 The GM FVR is consistent across the different regression methods.

HB1 The spatial structure of FVR is stable in time.

HB2 The spatial structure of FVR is consistent across data sets.

HB3 The spatial structure of FVR is consistent across the different regression methods.

In the following, the attempt is made to answer the research questions and to potentially falsify the hypotheses stated above.

1.4 A Note on Terminology

In order to prevent misunderstandings and to enhance clarity, a brief note on the terminology used in this thesis is given in the following. For simplicity of formulation and better readability, the following short hands are used:

- "forcing regressions" ⇔ regressions using reconstructed forcing time series as independent variable
- "MEM regressions" ⇔ regressions using the CESM MEM as independent variable
- "GM-MEM regressions" ⇔ regressions using the CESM GM-MEM as independent variable

Moreover, the prefix "annual" indicates that the regressions have been performed using time series containing annual-mean data. Likewise, the prefix "climatological" signifies that the regressions have been carried out using 31-year Butterworth low-pass filtered time series. Furthermore, caution is called for when interpreting time axes of FVR plots. Namely, the time axes in the figures of FVR and PCA thereof indicate start years of moving windows. Therefore, an indication of 1601 in a time axis refers to a moving window covering the time from 1601 to 1750. This is reflected in the descriptions in the text, so that this period from 1601–1750 would be referred to as "17^{th} and 18^{th} centuries". Additionally, the abbreviation GM FVR refers to the GM of the FVR values computed on a grid of 5° longitude by 5° latitude cells, not to be confused with FVR computed for GM time series. Such a quantification is not undertaken in this thesis.

2 Data

This chapter presents the data sets used in the present study. The ordering of the data sets in the chapter reflects their increasing temporal coverage. Thus, the relatively short instrumental and reanalysis data sets come before the climate model data sets, which roughly cover the last millennium. The Climate Field Reconstruction (CFR) data sets are treated last, since they cover the longest time period, encompassing the entire CE. Additionally, an introductory section treats the criteria for temperature data set selection, and reconstructed forcing time series data sets are presented leading up to the climate model data sets.

2.1 Criteria for Temperature Data Set Selection

The FVR can be calculated for any climatic variable, in principle. Initially, surface temperature (for all data sets), reference height temperature and precipitation (for climate model, instrumental and reanalysis data sets) have been included in the present analysis. However, the FVR data from reference height temperature and precipitation have been excluded during the course of the work. One reason for that was simplification of the analysis. A second reason was that only surface temperature was present in all of instrumental, reanalysis, climate model and CFR data sets and therefore comparability was enabled. Moreover, the FVR values from reference height temperature data were found to be very similar to the ones from surface temperature data.

A further criterion for surface temperature data set selection was that the data sets should have complete global coverage. This was chosen so as to simplify the analysis and enable comparability.

For the climate model data sets, an additional criterion was that they should be large ensembles that go back in time as far as possible. This applied to the Community Earth System Model - Last Millennium Ensemble (CESM), covering the period 850–2005 CE with an ensemble of size 13 and for the Hadley Centre Coupled Model, version 3 (HadCM3) ensemble, which covers the time 1401–2000 CE with an ensemble of size 4. The reason for this criterion was that the model ensemble means were used as predictor data sets in the MEM and GM-MEM regressions (see subsections 3.2.2 and 3.2.3). This was effectively carried out only for the CESM ensemble.

2.2 Instrumental and Reanalysis Temperature Data Sets

2.2.1 HadCRUT4

The Hadley Centre and Climatic Research Unit Temperature data set, version 4 (HadCRUT4) is temperature data set with global coverage, including both the land surface (near-surface air temperature) and the oceans (sea-surface temperature) (Morice et al. 2012). It is a composite of the Climatic Research Unit Temperature database, version 4 (CRUTEM4) for the land surface (Jones et al. 2012) and the

Hadley Centre Sea Surface Temperature, version 3 (HadSST3) data set for the oceans (Kennedy et al. 2011a; Kennedy et al. 2011b). In its original version, it had incomplete spatial coverage, lacking large areas in the Arctic and Antarctic as well as in continental Africa (Morice et al. 2012). This is a potential source of bias, and the incomplete data set did not fulfill the criteria delineated in section 2.1. For these reasons, an interpolated version (Cowtan & Way 2014) that made use of the kriging optimal interpolation algorithm (Cressie 1990) was used in this thesis. HadCRUT4 has a spatial resolution of 5° (longitude) x 5° (latitude) and a monthly temporal resolution. It starts in 1850 CE and the temperatures are given as anomalies to the reference period 1961-1990 (Morice et al. 2012). In the present study, the ensemble median (n = 100) of the HadCRUT4 data set was used.

2.2.2 20CRv2

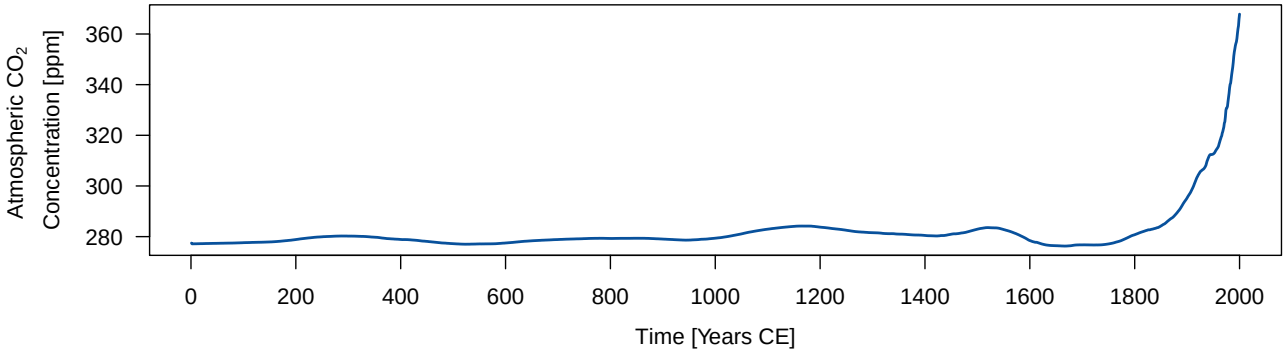
The Twentieth Century Reanalysis data set, version 2 (20CRv2) is a "synoptic-observation-based estimate of global tropospheric variability" (Compo et al. 2011). It was produced using observations of surface pressure from the International Surface Pressure Databank, version 2 (ISPDv2) (Compo et al. 2010; Cram et al. 2015). Also, sea-ice concentrations as well as SST fields were prescribed as boundary conditions (Compo et al. 2011). As a model for the reanalysis, the Global Forecast System of the U.S. National Centers for Environmental Prediction (GFS) was used (Kanamitsu et al. 1991; Moorthi et al. 2001; Saha et al. 2006, among others). This coupled atmosphere-land model allowed for dynamic CO_2 and volcanic aerosols. It provided forecast ensembles that were subsequently used in a deterministic "Ensemble Kalman Filter data assimilation method" (Compo et al. 2011) to produce the 20CRv2 data set. In doing so, the ensemble (n = 56) approach also provided a basis for estimation of uncertainty. The 20CRv2 data are available at 6-hourly temporal and 2° spatial resolution (Compo et al. 2011). In the present study, the ensemble mean of the 20CRv2 surface temperature data set was used.

2.3 Reconstructed Forcing Time Series Data Sets as used in CMIP6

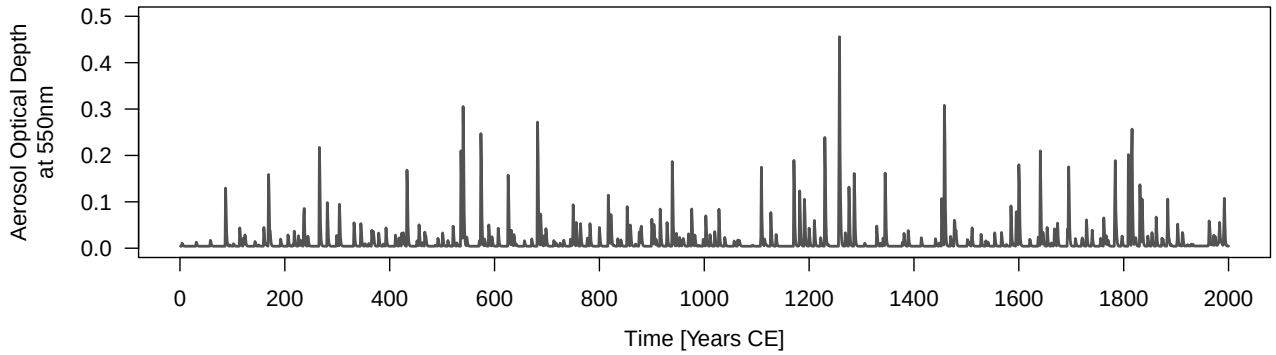
Reconstructed forcing time series data sets have been used in this analysis as predictor variables in the multiple linear regression calculations described in subsection 3.2.1. Possible choices of climate-system external forcings comprise natural forcings such as orbital, volcanic and solar forcings as well as anthropogenic forcings such as aerosol, land use change and GHG forcings (e.g. Schmidt et al. 2011; Jungclaus et al. 2017). For the present analysis, volcanic and solar forcing as well as anthropogenic GHG forcing have been chosen, as they were deemed the most important over the time scale considered here (the CE). In this section, the solar, volcanic and GHG forcings from the Coupled Model Intercomparison Project Phase 6 (CMIP6) are described (Eyring et al. 2016). The forcings used to drive the climate models are subsequently described in the corresponding sections 2.4.1 and 2.4.2.

The solar forcing reconstruction by Vieira et al. 2011 was adopted in this study. This data set was produced in the Spectral and Total Irradiance Reconstruction, version H (SATIRE-H) and is based on dendrochronological records of cosmogenic ^{14}C isotopes. As a volcanic forcing data set, a composite of the volcanic forcing used in CMIP6 (Eyring et al. 2016, spanning the time from 1850 to 2000 CE) and the eVolv2k, version 2 volcanic forcing reconstruction (Toohey & Sigl 2017, spanning the time from 500 BCE to 1900 CE) was used. This volcanic forcing reconstruction is based on measured

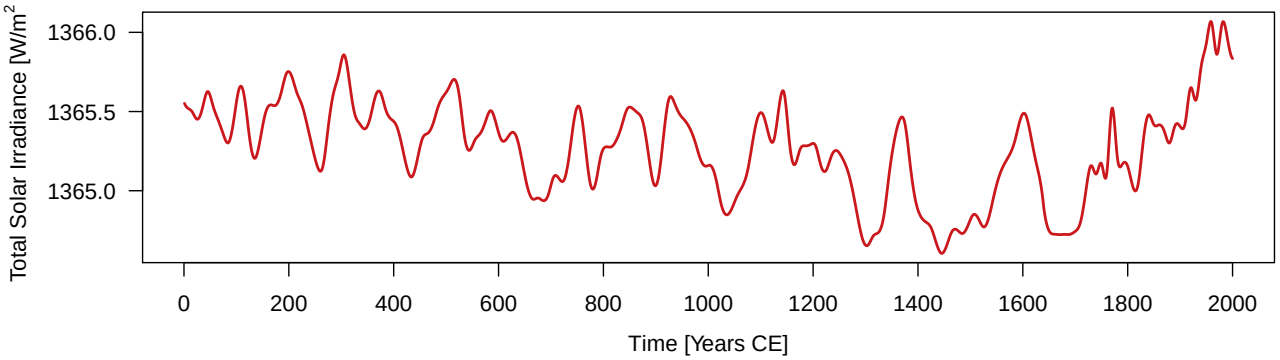
sulfate in ice core records from Greenland and Antarctica. It represents a significant improvement compared to earlier reconstructions (Crowley & Unterman 2013; Gao et al. 2008) used in Coupled Model Intercomparison Project Phase 5 (CMIP5) model simulations (Sigl et al. 2015). As a GHG forcing reconstruction, the CO_2 component of the data set described in Meinshausen et al. 2017 and Jungclaus et al. 2017 has been used in this thesis. It is based on highly resolved ice-core records from Greenland and Antarctica. Figure 2.1 depicts the three climate forcing reconstructions described before.



(a) CO_2 Forcing



(b) Volcanic Forcing



(c) Solar Forcing

Figure 2.1: *Reconstructed forcing time series from CMIP6. (a) Atmospheric CO_2 concentrations in parts per million (Meinshausen et al. 2017), (b) Global-mean annual-mean stratospheric AOD at 550nm wavelength (Toohey & Sigl 2017; Eyring et al. 2016), (c) TSI in W/m^2 (Vieira et al. 2011).*

2.4 Climate Model Data Sets

A climate model is "a numerical representation of the climate system based on the physical, chemical and biological properties of its components, their interactions and feedback processes" (IPCC 2013). Climate models can range in their sophistication according to "such aspects as the number of spatial dimensions, the extent to which physical, chemical or biological processes are explicitly represented or the level at which empirical parametrizations are involved" (IPCC 2013). Here, surface temperature data sets from realizations of two highly sophisticated climate models are used. The first is the CESM, the second the HadCM3. Both models are briefly described in the following.

2.4.1 Community Earth System Model - Last Millennium Ensemble

The CESM is a highly sophisticated climate model – an earth system model (Hurrell et al. 2013). It has been used in an ensemble approach to produce 13 realizations of climate spanning the period from 850 to 2005 CE, the Community Earth System Model - Last Millennium Ensemble (CESM) (Otto-Bliesner et al. 2016). All these 13 realizations have been forced with transient forcing reconstructions from CMIP5, comprising solar, volcanic and GHG forcings, among others (Schmidt et al. 2011). As a solar forcing, the reconstruction by Vieira et al. 2011, already described in 2.3, was used. As volcanic forcing, the CESM adopted a reconstruction based on sulphate deposition in ice cores from Greenland and Antarctica (Gao et al. 2008). As GHG forcing, Otto-Bliesner et al. 2016 made use of the composite reconstruction by Schmidt et al. 2011. This forcing data set combines the three GHG CO_2, CH_4 and N_2O, measurements of which stem from highly resolved ice cores from Antarctica. The CESM simulation was carried out on a 2° horizontal resolution for the atmosphere and land-surface components and 1° horizontal resolution for the ocean and sea-ice components (Otto-Bliesner et al. 2016).

2.4.2 Hadley Centre Coupled Model – Version 3

Like CESM, the Hadley Centre Coupled Model, version 3 (HadCM3) is a highly sophisticated climate model – a Coupled Atmosphere–Ocean General Circulation Model (AOGCM) (Pope et al. 2000; Gordon et al. 2000). In this study, surface temperature data sets from four runs of the HadCM3 are used. The first run covers the period from 800 to 2000 CE, while the other three cover the period from 1401 to 2000 CE. All four runs were forced with transient forcing reconstructions (Schurer et al. 2013a). As a solar forcing, the reconstruction of TSI by Steinhilber et al. 2009 was used. This reconstruction makes use of cosmogenic ^{10}Be radionuclides that were measured in ice cores. As a volcanic forcing, the reconstruction by Crowley & Unterman 2013 was adopted, which is based on sulphate measurements in ice cores from Greenland and Antarctica. As a GHG forcing, the reconstruction by Schmidt et al. 2011 was used, which was already described above (2.4.1). The HadCM3 simulation was carried out on a 3.75° longitude by 2.5° latitude resolution with 19 vertical levels (atmosphere) and 1.25° longitude by 1.25° latitude resolution with 20 vertical levels (ocean) (Tett et al. 2007; Schurer et al. 2013a).

2.5 Climate Proxy Data Sets and Climate Field Reconstruction Data Sets

2.5.1 Climate Proxy Data Sets

Due to the limited temporal coverage of the instrumental period, paleoclimatologists resort to climate proxies in order to reconstruct the climate of the past. The word "proxy" indicates that the information it contains is related indirectly to climate. Climate proxies can be of natural or anthropogenic origin. Anthropogenic climate proxies are archives such as records of phenology, lake levels or archaeological sources. Natural proxies encompass environmental archives such as tree rings, ice cores, speleothems, corals and lake sediments (Brönnimann 2015).

In this study, no climate proxy data sets were used directly. However, a set of CFR data sets (Neukom et al. 2019) that is based on natural climate proxy data was used. The proxies used by Neukom et al. 2019 are contained in the PAGES 2k global temperature-sensitive proxy collection (PAGES 2k Consortium 2017). Figure 2.2 shows an overview of these proxy data.

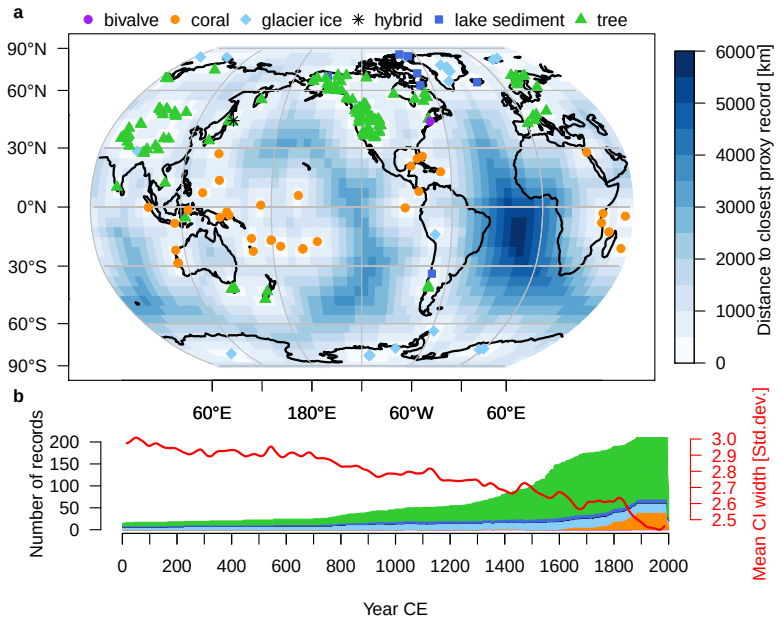


Figure 2.2: *Overview of climate proxy records from the PAGES 2k global temperature-sensitive proxy collection (PAGES 2k Consortium 2017). (a) Spatial distribution of the proxy records by proxy type, and an indication of the distance of each grid cell to the closest proxy record (shading), (b) temporal coverage of the different proxies, by proxy type (colors as in (a)), and the average width of the confidence interval of the CFRs by Neukom et al. 2019. Figure courtesy of Raphael Neukom.*

2.5.2 Climate Field Reconstruction Data Sets

Using the environmental proxies described before (PAGES 2k Consortium 2017), Neukom et al. 2019 reconstructed globally complete Surface Temperature (ST) fields for the CE using six different Climate Field Reconstruction (CFR) methods. An overview of those CFR methods is given in the following. In all six CFR methods, a version of HadCRUT4 (Morice et al. 2012) infilled with GraphEM was used as an instrumental reconstruction target. Moreover, ensembles of size 100 with different realizations of

the errors were produced for every CFR method by Neukom et al. 2019 and used in this thesis.

The conceptually simplest CFR method used by Neukom et al. 2019 is the Composite Plus Scale Reconstruction Method (CPS). Here, the average of the proxy time series is taken to form a composite time series. This composite is then scaled to match the standard deviation and mean of the reconstruction target. Neukom et al. 2019 performed the CPS in a gridpoint-wise manner to reconstruct the surface temperature field. CPS has the property not to rely on the spatial covariance structure of the reconstruction target. On the one hand, this "may lead to unrealistic spatial consistency in the reconstructed fields" (Neukom et al. 2019). On the other hand, this method does not rely on the assumption that the spatial covariance structure of temperature (as defined by the reconstruction target) is stable over time (Neukom et al. 2019).

The second CFR technique used by Neukom et al. 2019 is called Principal Component Regression Reconstruction Method (PCR). In this method, PCA is used to reduce the dimensions of the proxy records and the instrumental reconstruction target. Subsequently, regression against the proxy Principal Component (PC)s is used to reconstruct the instrumental PCs back in time. Subsequently, ST fields are retrieved by using the PC loading fields. This CFR technique relies on the assumption that the spatial covariance structure of the instrumental target is stable over time (Neukom et al. 2019).

The third CFR technique that Neukom et al. 2019 used is the Canonical Correlation Analysis Reconstruction Method (CCA). It is related to PCR, but here the information from the proxies and the instrumental reconstruction target are used simultaneously by Canonical Correlation Analysis. This method relies on the assumption that most of the variance from the proxies and the instrumental target is contained within the leading modes of the CCA (Neukom et al. 2019).

The fourth CFR method used by Neukom et al. 2019 is called Graphical Model with Expectation Maximization Algorithm Reconstruction Method (GraphEM). It makes use of graph theory, more precisely of Gaussian graphical models (also called Markov random fields). These make use of the "conditional independence structure of the climate field" to estimate "the inverse of the covariance matrix", which is then iteratively determined (Neukom et al. 2019).

A further CFR technique that Neukom et al. 2019 used is called Data Assimilation Reconstruction Method (DA). Unlike the previously described CFR methods, this method takes the covariance structure of temperature from a climate model simulation and not from the instrumental reconstruction target. Hence, the covariance structure is assumed to be stable over time. As a climate model simulation, the ensemble member 10 of the CESM is used. Real proxies and pseudoproxies from the CESM are used and weighted using a Kalman filter. This weighting has the property that high uncertainty in the proxies leads to low variability in the CFR (Neukom et al. 2019).

The last CFR method that Neukom et al. 2019 used is the Analogue Method Reconstruction (Analogue). Like DA, this CFR method makes use of the covariance structure of temperature from climate model simulations. For that, Neukom et al. 2019 used the ensemble of simulations from the Paleoclimate Modelling Intercomparison Project Phase 3 (PMIP3). The proxy records are linearly fitted to the instrumental reconstruction target to create a local temperature reconstruction. The temperature fields from the climate models are then compared to those local temperature reconstructions and the five fields with the best fit are subsequently averaged to form the reconstructed temperature field in a given year. Therefore, the temporal evolution of temperature is defined by the proxy records, while the spatial structure is derived from the climate model simulations (Neukom et al. 2019).

3 Methods

3.1 Data Pre-Processing

To homogenize the surface temperature data sets described in chapter 2, a number of data pre-processing steps were carried out using the Climate Data Operator Software (CDO) (Schulzweida 2018). First, the 20CRv2 and HadCRUT4 data sets were transformed to a regular 5° longitude by 5° latitude grid, using bilinear interpolation. The climate model data sets used here were kindly provided by Michael Faden in an already interpolated form. The CFR data sets used were originally produced on a 5° longitude by 5° latitude grid (Neukom et al. 2019), so no interpolation was necessary.

Second, annual mean values were computed from monthly mean values for the instrumental, reanalysis and climate model data sets. Because the CFR data sets reflect April – March annual mean temperature, this partitioning of months was also adopted for the other data sets, to enable comparability. Hence, a reference to year 1600 means the time from April 1600 to March 1601.

In a third step, all the surface temperature data sets were transformed from absolute values to anomalies with respect to the reference period 1961–1990 CE. For the climate model data sets, anomalies were calculated for every ensemble member individually. For the CFR data sets, anomalies were calculated with respect to the ensemble mean, so as to preserve the variance within the ensemble.

For the surface temperature data from CESM, the Model Ensemble Mean (MEM) and the Global-Mean Model Ensemble Mean (GM-MEM) were calculated. This was needed for the regression computations with MEM and GM-MEM (see sections 3.2.2 and 3.2.3).

Finally, the surface temperature data sets as well as the predictor data sets (forcing, MEM and GM-MEM time series) were temporally filtered using a 31-year Butterworth low-pass filter (Butterworth 1930). This was done in order to extract the climatological signal in the time series. The filter length of 31 years was preferred to a filter length of 30 years (the classical time scale of climate according to IPCC 2013) for reasons of symmetry of the filter. This filtering resulted in a loss of effective degrees of freedom and also in a loss of 15 years of data at either end of each time series. In the remainder of the thesis, calculations using the filtered time series are referred to as ”climatological” and calculations using the unfiltered annual-mean time series are referred to as ”annual”.

3.2 Linear Regression Calculations

The central concept in this thesis is to extract the FVR using linear regression calculations. More precisely, surface temperature time series are regressed against time series that represent climatic forcing. The resulting coefficient of determination (regression R^2) represents the fraction of variance in the response time series (surface temperature) that is explained by the predictor (the time series representing forcing) (Wilks 2011). Accordingly, R^2 represents the FVR. Linear regression calculations

have been used before to separate climatic internal variability from external forcing (see section 1.2).

Such linear regression calculations were performed at every grid cell in the surface temperature data sets. This yielded a FVR value at every grid cell, thereby generating spatially explicit quantifications (global maps) of FVR. Moreover, these regression calculations were carried out in moving windows of 150 years length and 10 years leap time. This made the analysis of the temporal evolution of FVR possible, as sequences of FVR maps were produced. The window length of 150 years was chosen so that exactly one moving window would fit into the time span of data availability in the instrumental and reanalysis data sets. The leap time of 10 years was chosen so that the evolution of FVR was temporally sufficiently resolved while still keeping computing effort within reasonable limits. Furthermore, these regression calculations were carried out using both the annual-mean time series (annual regressions) as well as the 31-year Butterworth filtered time series (climatological regressions). This was intended to yield quantitifications of FVR both on the annual and on the climatological time scale. All linear regression models were fitted using minimization of ordinary least squares. Finally, the regression calculations were carried out in three different approaches, which are described in the following subsections.

3.2.1 Multiple Linear Regression with Forcing Time Series

The first approach to calculating the FVR is a multiple linear regression approach. Surface temperature time series are regressed against reconstructed forcing time series, representing solar, volcanic and GHG forcings.

As a preparatory step, the three forcing time series were individually cross-correlated with the surface temperature time series before being used in the regression. The rationale behind this cross-correlation is to take into account a potentially delayed (lagged) response of the climate system to forcing. Similar approaches have been used by Wang et al. 2017, Wang et al. 2018 and Faden 2018. For the cross-correlations, the Pearson correlation coefficient was used (Pearson 1895). The lag of temperature relative to forcing in the cross-correlation was constrained to be between zero and 30 years. Within this time, the highest absolute correlation was looked for, allowing for a correlation of either sign between temperature and forcing. Thus, no assumption was made about the sign of the response of temperature to forcing. Moreover, the cross-correlation calculations were carried out over the entire time series of data availability, not in a moving window approach. Therefore, a constant response time of surface temperature to forcing was assumed. To investigate the effect of the lagged response of temperature, the regression calculations described in the following have been carried out both with a lag of zero years and with the lags from the cross-correlation calculations.

The multiple linear regression approach using reconstructed forcing time series as predictors can be summarized in the following equation:

$$T(t) = \beta_0 + \beta_1 F_{solar}(t - lag_{solar}) + \beta_2 F_{volcanic}(t - lag_{volcanic}) + \beta_3 F_{GHG}(t - lag_{GHG}) + \varepsilon(t) \quad (3.1)$$

where $T(t)$ is the surface temperature response time series, β_n are the regression parameters, $F_i(t)$ are the forcing predictor time series, lag_i are the respective lags from the cross-correlations and $\varepsilon(t)$ is the regression residual time series. This regression approach is referred to as "forcing regressions" in the remainder of the thesis.

3.2.2 Linear Regression with Model Ensemble Mean

The second approach to calculating the FVR makes use of the CESM Model Ensemble Mean (MEM). The rationale is that the forced response of the climate system can be extracted from an ensemble of model runs by taking the MEM (Frankcombe et al. 2015). By taking the MEM, the different signals from internal variability are averaged out and the forced response signal becomes apparent. The surface temperature time series is then regressed against the MEM to derive FVR. Similar approaches have been carried out before (Dai et al. 2015; Frankcombe et al. 2015; Kosaka & Xie 2016). Here, the MEM is calculated at every grid cell and the regression carried out at every grid cell separately. With this approach, the forced response of the climate system is treated dynamically, the different forcings are weighted and the forcing fingerprint is allowed to vary spatially. The linear regression equation with MEM then looks as follows:

$$T(t) = \beta_0 + \beta_1 MEM(t) + \varepsilon(t) \tag{3.2}$$

where $MEM(t)$ is the MEM time series of the CESM model at each grid point. The other variables are analogous to equation 3.1. Note that there is no lag here, because the lagged response of the climate system to forcing is already treated dynamically in the model. This linear regression approach relies on the assumption that the different CESM ensemble members have the same response to external forcing. Another assumption is that the different ensemble members have been sufficiently differently initialized to produce independent realizations of internal variability (Frankcombe et al. 2015). In the remainder of the thesis, this regression approach is referred to as "MEM regressions".

Due to the limited ensemble size (n = 13 in the CESM), not all of the internal variability is averaged out by taking the MEM (Lyu et al. 2015; Barkhordarian et al. 2013). For this reason, the regression R^2 needs to be corrected for ensemble size. This is done using a simple linear transformation of the original R^2 to a value corrected based on the ensemble size (see derivation in the appendix). In the cases where FVR was corrected to below zero, FVR was set to zero (Harzallah & Sadourny 1995; Lyu et al. 2015).

3.2.3 Linear Regression with Global-Mean Model Ensemble Mean

The third approach to calculating the FVR is very similar to the second (3.2.2). The only difference is that here, the surface temperature time series at every grid point are regressed against the CESM GM-MEM instead of the grid-point wise MEM. Like with the regressions with MEM, no lag is needed and R^2 needs to be corrected for ensemble size. This approach is referred to in the remainder of the thesis as "GM-MEM regressions".

3.2.4 Validation of the Linear Regression Calculations

The linear regression calculations described before were assessed for their quality in three different ways. First, they were cross-validated by independently recalculating the regressions for a sub-sample of the data points. Subsequently, the resulting R^2 values were compared to the corresponding ones calculated before.

Second, the regression calculations were repeated with the time series covering the full length of each data set as opposed to the moving window approach. This was done to address concerns about

reduced effective degrees of freedom in the climatological regressions. By taking the entire time series, effective degrees of freedom could be enhanced (due to the longer time series) while still retaining filtered time series for the calculations. Therefore, an overestimation of R^2 due to reduced degrees of freedom could be corrected for by using the entire time series with enhanced degrees of freedom.

Third, the forcing regression calculations were assessed for multicollinearity. Multicollinearity is a phenomenon that occurs in regression calculations when predictors are correlated. If present, it leads to overestimation of regression R^2 (Gujarati & Porter 2010). To assess whether multicollinearity is present, 10 arbitrary moving windows and 100 arbitrary grid points of an example data set (the DA CFR) were repeatedly (n = 100) sampled. For every sample, the corresponding forcing predictors (solar, volcanic and GHG forcing) and response time series (surface temperature) were jointly assessed for multicollinearity using six different indicators (Imdadullah et al. 2016). The indicators used were the determinant $|X'X|$, the Farrar's Chi-Squared, the Red indicator, the sum of reciprocal eigenvalues (sum of inverse lambdas), Theil's indicator and the condition number. Every indicator could either be 0 (multicollinearity not detected) or 1 (multicollinearity detected) for every sample. The results were summarized as percentages of samples in which multicollinearity was detected, in table form (see 4.11.3). This was done for both annual and climatological time series.

3.3 Trend Analysis

From the FVR maps generated with the methods described in 3.2, Global Mean (GM) values were calculated using latitude-dependent area-weighting of the grid cells. For the climate model and CFR data sets, this resulted in time series of GM FVR. For the instrumental and reanalysis data sets, no time series resulted, since they only possessed a single moving window. The time series from the climate model and CFR data sets were subsequently subjected to a trend analysis to answer research question QA1.

For the trend analysis, the three methods ordinary least-squares linear regression, linear regression with Autoregressive Moving Average (ARMA)(p,q) errors (Hyndman & Athanasopoulos 2018) and the Mann-Kendall trend test (Kendall 1938; Mann 1945; Kendall 1975) were considered. Examination of example time series of GM FVR revealed high autocorrelation at several lags. Therefore, it was decided to carry out the trend analysis using linear regression with ARMA(p,q) errors. The linear regression model applied (Hyndman & Athanasopoulos 2018) is of the form:

$$FVR(t) = \beta_0 + \beta_1 t + \varepsilon(t) \tag{3.3}$$

where $FVR(t)$ is the FVR time series, β_n are the regression parameters, t is time and $\varepsilon(t)$ are ARMA(p,q) errors. Here, the regression parameter β_1 represents the magnitude of the linear trend. The order of the ARMA(p,q) errors was determined in a stepwise procedure by optimization of Akaike's corrected Information Criterion (AICC); p and q were limited to the range of integers between 0 and 5 (Hyndman & Khandakar 2008). The p-values of the trends were calculated from the trends and their associated standard errors.

The trend tests have been carried out over both the entire time series of FVR available for each data set and also for the pre-industrial time only. For the pre-industrial trend tests, the time series

have been truncated after the moving window start time 1601, so that the analysis remained limited to the time before 1750. All trend tests have been calculated on the GM FVR data from CFR and climate model ensemble members as well as on the ensemble mean.

3.4 ANOVA and Student's t-Test

In order to answer research questions QA2 and QA3, the means of the distributions of FVR needed to be compared between data sets and between regression methods. To this end, Analysis of Variance (ANOVA) and two-sample Student's t-tests were used.

In a first step, ANOVA was used to determine if there are differences between groups (data sets or regression methods). In a second step, all combinations of two groups were subjected to pairwise (not paired!) two-sample Student's t-tests to examine them for differences in their mean values. This was done by treating the variances of the respective samples as unequal and by using the Welch-Satterthwaite approximation to degrees of freedom (Satterthwaite 1946; Welch 1947).

3.5 Principal Component Analysis

PCA (Obukhov 1947; Lorenz 1956; Davis 1978) is a multivariate statistical method widely used in the atmospheric sciences (Wilks 2011). In this thesis, it was applied to the FVR fields from climate model and CFR data sets to extract the dominant patterns of FVR variability. In doing so, the grid cells were weighted according to the square root of the cosine of their latitude and the PCA was carried out on the covariance matrix of the data. The covariance matrix was chosen because all the data were of the same units (FVR, i.e. [0,1]) and hence standardization by using the correlation matrix was not necessary. However, calculations with the correlation matrix were carried out on a sub-sample of the data and results qualitatively compared to the ones from the PCA with the covariance matrix. Since the number of grid points ($72 \times 36 = 2592$) was by far larger than the frequency of FVR (ranging from 46 for HadCM3 to 186 for the CFRs), a computational trick (Storch & Hannoschöck 1984; Wilks 2011) was used to limit calculation effort by using the transpose of the data matrix. PCA was carried out on both ensemble members and ensemble means of the climate model and CFR data sets.

4 Results

This chapter presents the main findings of this thesis. Some interpretations will also be given, at the end of the respective sections and clearly separated from the results.

The FVR data calculated from reference height temperature data sets were very similar to the ones calculated from ST data sets. For this reason, and for the reasons delineated in section 2.1, the results shown here are constrained to FVR data calculated from surface temperature data sets.

In addition to the results shown in this chapter, several other analyses have been carried out. Because they do not contribute directly to answering the research questions laid out in section 1.3, they are not presented here.

This chapter is broadly structured into four parts: The first (sections 4.1 - 4.4) is concerned with the GM FVR. An intermediate part treats the zonal-mean FVR (section 4.5). The third part (sections 4.6 - 4.10) is concerned with the two-dimensional spatial structure of the FVR. The fourth, and last part (section 4.11) treats the results of the validation of the linear regression calculations described in section 3.2.

For the part on the spatial structure, the results shown are not intended to be exhaustive, but rather to give examples of key properties found in the data. Also, for the world maps shown in this part, 120°W has been chosen as a central longitude, so as to display the whole Pacific and Atlantic oceans at the same time. Where sections are subdivided according to data sets, subsections have been ordered so as to reflect the increasing temporal coverage the data types (instrumental and reanalysis data, climate model data, CFR data).

4.1 Global-Mean FVR: Overview

This section will present an overview over all the FVR data produced in this thesis. To do so, the FVR data (which were calculated for every grid point and ensemble member) have been compressed by taking global means and ensemble means for every data set. Figure 4.1 shows these data as time series plots, grouped according to the regression methods described in section 3.2. The FVR data from CFRs and models are shown as lines, whereas the FVR data from the instrumental and reanalysis data sets are shown as points, since for them only a single map of FVR was calculated. The data from both timescales (annual and climatological) are included in the same plots, with the FVR values from the regressions on the annual time scale generally being lower and closer to zero. Figure 4.2 shows the same data in a different way, with the exception that data from annual regressions with MEM and GM-MEM are omitted. Those are very low in general and exhibit little structure. For both figures 4.1 and 4.2, the time axes are given as the starting years of the respective moving windows (see section 3.2).

4.1.1 Description

The FVR values produced from regressions with forcing time series on the climatological time scale are generally the highest, as can be seen from figures 4.1a and 4.2b. Also, for those data, the agreement between FVR values from CFRs and models is generally very high. The two most distinctive features are a significant drop in FVR in the 16^{th} and 17^{th} centuries and a consecutive strong rise in the 18^{th}, 19^{th} and 20^{th} centuries. One exception thereof are the FVR values from the HadCM3, which do not drop in the 17^{th} century. Another noticeable feature are the high FVR values from the CESM in the 12^{th}, 13^{th} and 14^{th} centuries.

The FVR values produced from regressions with forcing time series on the annual time scale are much lower than those on the climatological time scale (figures 4.1a and 4.2a. For most data sets, they are close to zero. A noticeable exception is the DA CFR, which has somewhat higher values and also higher variability. As with the data on the climatological time scale, the most prominent feature is the rise of FVR in the 18^{th}, 19^{th} and 20^{th} centuries, encompassing most data sets. The FVR values from the instrumental and reanalysis data sets also fall within the range of the other data sets. An exception is the CESM, whose FVR values only rise slightly in the 19^{th} and 20^{th} centuries. However, the higher values in the 12^{th}, 13^{th} and 14^{th} centuries described for the data on the climatological time scale can also be observed here.

For the FVR values from regressions with MEM, it can be observed that they are typically lower than those from regressions with forcing time series (figures 4.1b and 4.2c). This applies to both regressions on the climatological and the annual time scale. Also, the FVR data from MEM regressions on the climatological time scale are much more different among the data sets than the corresponding ones from regressions with forcing time series. Another important difference is that the aforementioned drop in FVR values in the 17^{th} century here is prolonged, and even more accentuated, in the 18^{th} and even 19^{th} centuries for FVR values from CFR data. However, this is not true for the FVR data from the two model data sets, whose development after about 1600 is very similar and more closely related to the one from regressions with forcing time series. FVR data from CESM are again special in the 12^{th}, 13^{th} and 14^{th} centuries, with much higher FVR than the other data sets. In the 19^{th} and 20^{th} centuries however, the agreement between the FVR values from all data sets is close. Moreover, the FVR data from all the CFR data sets agree quite well among each other over the whole time.

The FVR values from regressions with MEM on the annual time scale are generally very low, with even the ensemble mean being close to zero for most data sets and over the whole time (figure 4.1b). A notable exception are the FVR data from the CESM, which are substantially higher than those from other data sets. Also here the increase in the 12^{th}, 13^{th} and 14^{th} centuries can be observed. In the 19^{th} and 20^{th} centuries, a small increase in FVR is evident in the data from the CFRs and HadCM3.

The properties of the FVR data generated from regressions with GM-MEM (figures 4.1c and 4.2d) are qualitatively very similar to the ones from regressions with MEM. This holds true for both regressions on the climatological and the annual time scale. The only apparent qualitative difference is that the variability seems to be slightly higher in the FVR data from regressions with GM-MEM as compared to those from MEM. This manifests itself in the time series from MEM being more smooth, while those from GM-MEM being more jagged (respectively patchy in figure 4.2d).

All the climatological regressions seem to have in common a period of enhanced FVR during the

14^{th} to 16^{th} centuries, a significant drop thereafter and a consecutive strong rise in FVR in the 18^{th}, 19^{th} and 20^{th} centuries.

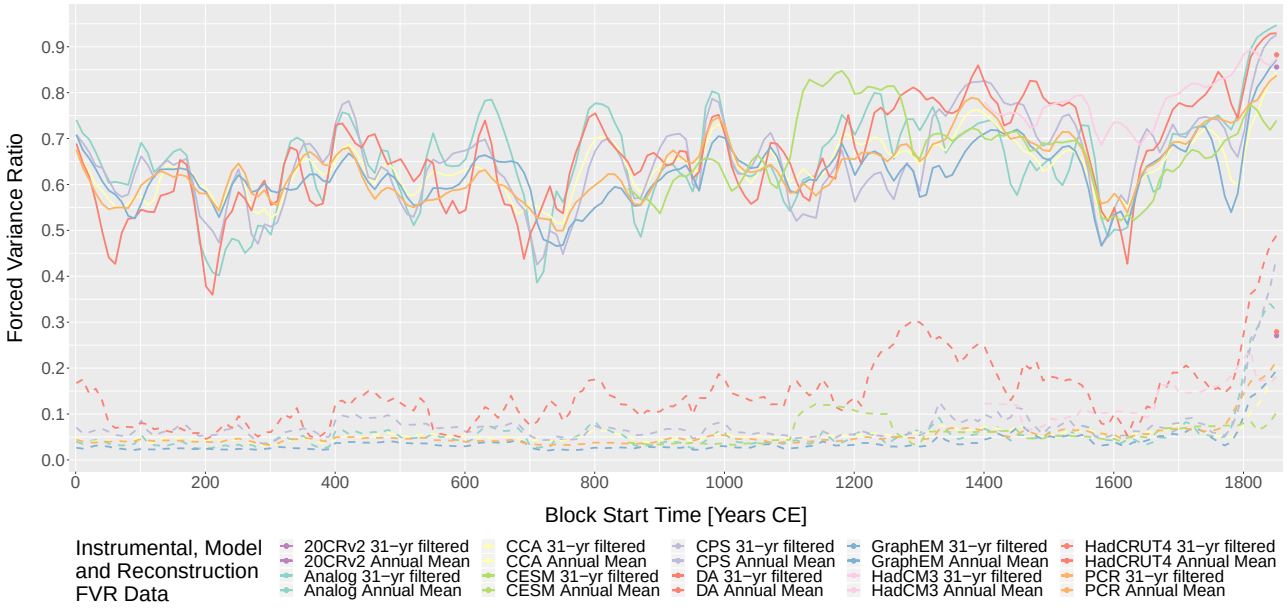
4.1.2 Interpretation

The FVR values from climatological forcing regressions seem to be the highest, as is evident from figure 4.2. This issue shall be examined more closely in section 4.4.

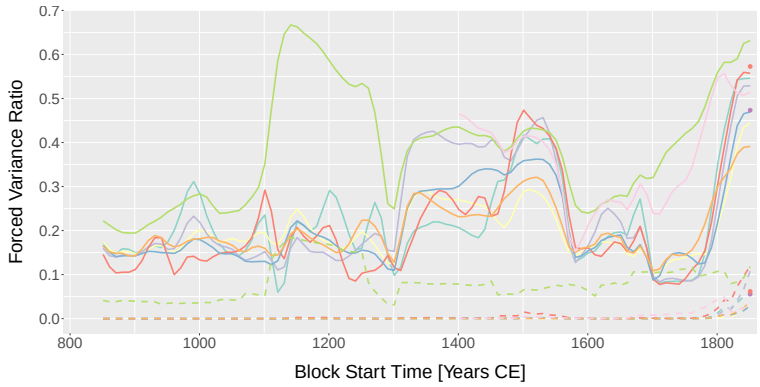
The comparatively high values of the CESM in the 12^{th}, 13^{th} and 14^{th} century are likely due to enhanced forcing as a consequence of the Samalas large volcanic eruption (Lavigne et al. 2013). This hypothesis is corroborated by the observation that the FVR values increase and decrease precisely when the year of the eruption, 1257, enters or leaves the respective moving window. The fact that CESM seems to respond more strongly to this volcanic eruption than do the other data sets could be explained by a higher sensitivity of the model to volcanic forcing.

The simultaneous periods of high and low FVR are consistent with data set and also with regression method (for climatological regressions). Therefore, this result seems to be robust and likely represents a true property of the climate system, rather than being a statistical artifact. A possible explanation for the high values of FVR in the 14^{th} to 16^{th} centuries could be stronger than average volcanic forcing. The period of low FVR in the 17th and 18th centuries can be interpreted as one of low forcing, where none of the three major drivers of climate (the sun, volcanoes and GHG) was very active and climate evolved freely, leading to high amounts of variance accounted for by internal variability. The most likely explanation for the generally high FVR values in the 19^{th} and 20^{th} centuries is enhanced forcing from anthropogenic GHG emissions. Additionally, climate has been shown to be forced by volcanic eruptions in the early 19^{th} century (Brönnimann et al. 2019).

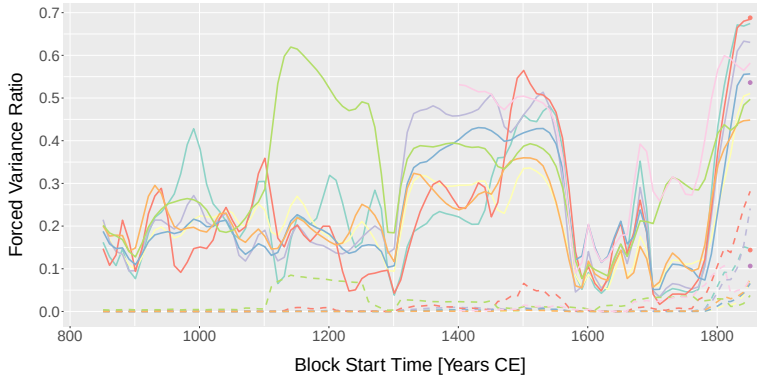
The enhanced variability of FVR in climatological GM-MEM regressions as compared to MEM regressions seems not to have an obvious explanation. However, it could be that due to the large number of different predictors (one for every grid point), the variability in the FVR values from MEM regressions was reduced. On the other hand, the GM-MEM regressions only had a single predictor (the CESM GM-MEM), which could have lead to higher variability.



(a) *Global-Mean Ensemble-Mean FVR from Forcing Regressions*

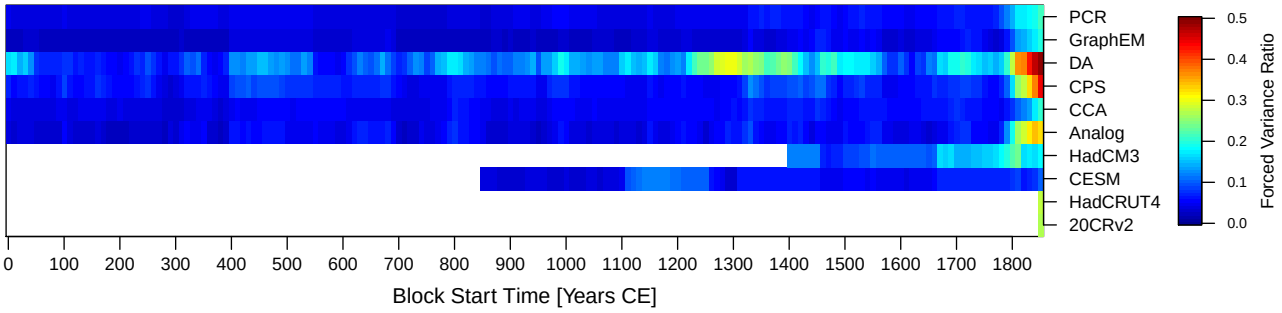


(b) *Global-Mean Ensemble-Mean FVR from MEM Regressions*

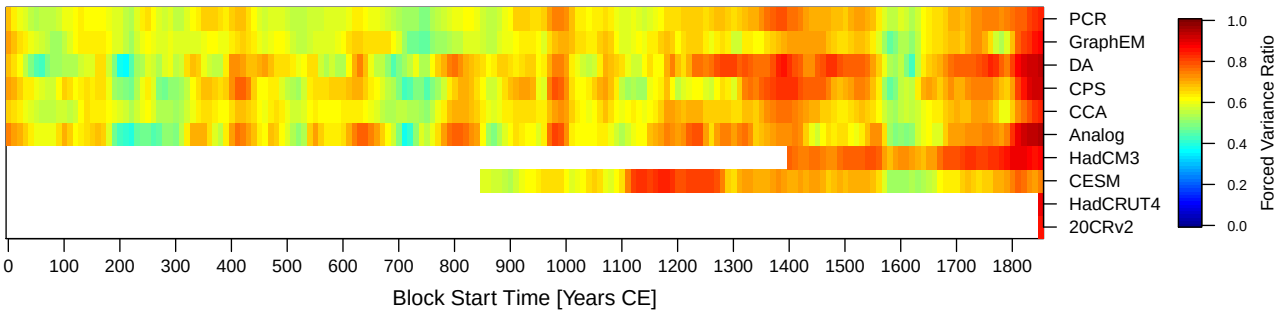


(c) *Global-Mean Ensemble-Mean FVR from GM-MEM Regressions*

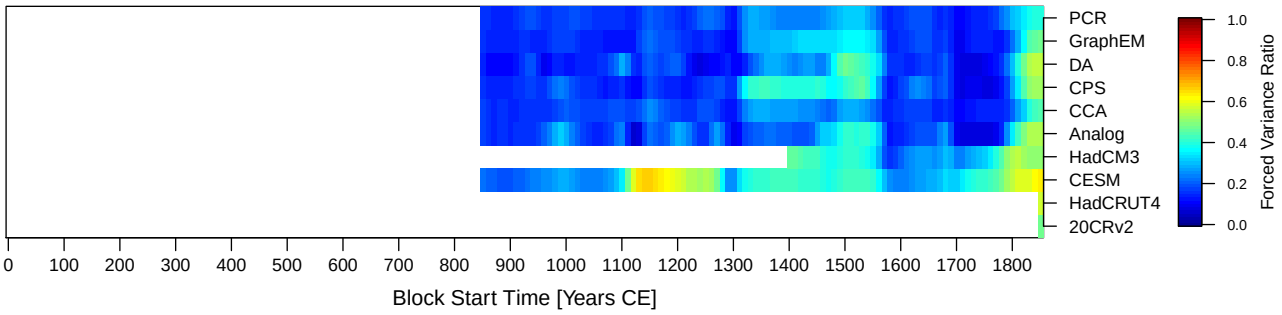
Figure 4.1: *Global-mean ensemble-mean FVR for the three different regression methods and all data sets. (a) Data from forcing regressions, (b) data from MEM regressions, (c) data from GM-MEM regressions. The FVR values obtained from regressions on the climatological time scale are generally found in the upper parts of the plots (solid lines), while those obtained from regressions with annual-mean data are generally found in the lower parts of the plots (dashed lines).*



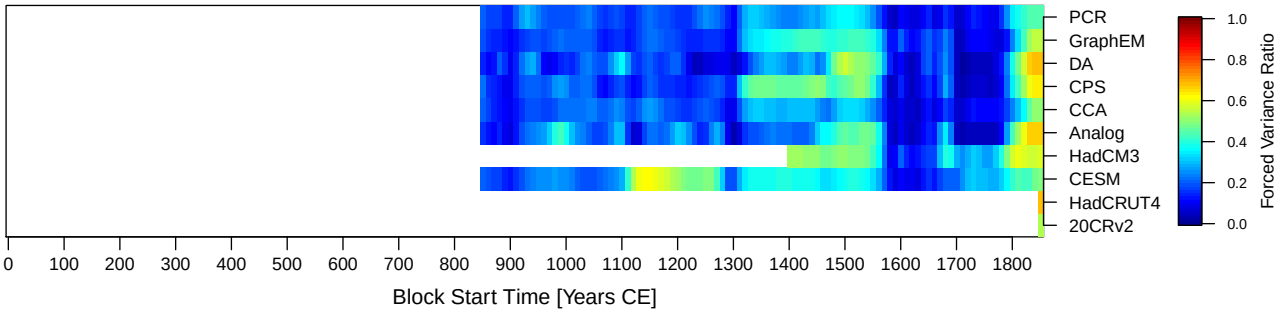
(a) *Global-Mean Ensemble-Mean FVR from Annual Forcing Regressions*



(b) *Global-Mean Ensemble-Mean FVR from Climatological Forcing Regressions*



(c) *Global-Mean Ensemble-Mean FVR from Climatological MEM Regressions*



(d) *Global-Mean Ensemble-Mean FVR from Climatological GM-MEM Regressions*

Figure 4.2: *The same data as in figure 4.1, but without the annual MEM and GM-MEM regressions. Here, data are displayed as heat maps, with the colour indicating the magnitude of FVR. (a) Data from annual forcing regressions, (b) data from climatological forcing regressions, (c) data from climatological MEM regressions, (d) data from climatological GM-MEM regressions. Note the different colour scale on subfigure (a).*

4.2 Global-Mean FVR: Trend Tests

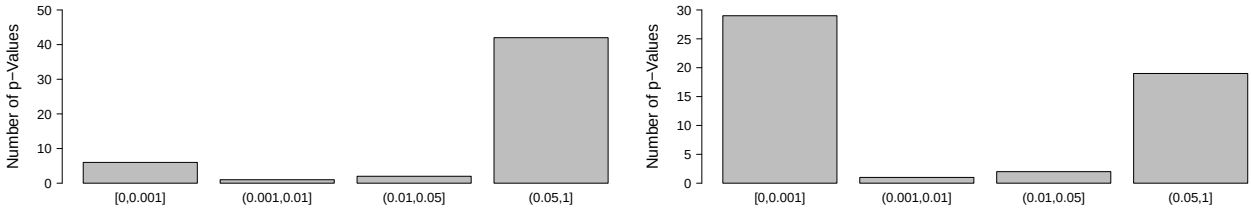
This section gives the results of the trend tests described in section 3.3. The idea behind the trend tests was to determine if there is a systematic increase or decrease of FVR over time, beyond (and taking into account) the variability demonstrated above (section 4.1). The effect of the high-forced industrial period was taken into account by also performing trend tests over the pre-industrial period only. Subsection 4.2.1 is dedicated to the results of the trend tests of the FVR data from the climate models, and subsection 4.2.2 to the ones from the CFRs.

4.2.1 Climate Model Data

For the trend tests of the entire time series, of all the trends from the CESM (n = 78), 35 were found to be significant at the 5% level. For HadCM3 (n = 24), 6 trends were found significant at the 5% level. For both models, all of the trends significant at the 5% level were found to be positive. The average magnitude of the significant trends ($\alpha = 5\%$) was found to be 0.021 (CESM) and 0.026 (HadCM3) units of FVR per 100 years.

For the trend tests of only pre-industrial FVR data, 28 trends were found to be significant at the 5% level for data from the CESM. This corresponds to a percentage of about 36%. Of those, all trends were found to be positive. The average magnitude of the 28 significant trends was found to be 0.025 units of FVR per 100 years. For HadCM3, eight trends were found to be significant at the 5% level, corresponding to a third of all the trends. Of the significant ones, two were found to be positive and six negative. The two positive ones were found to have magnitudes of 0.028 and 0.024 units of FVR per 100 years, respectively. The six negative ones were found to have an average magnitude of -0.145 units of FVR per 100 years.

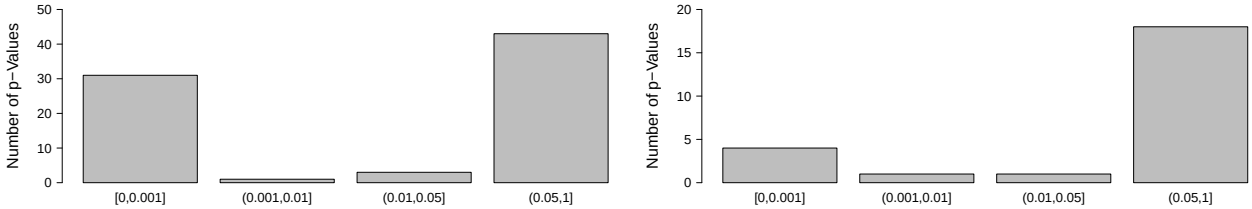
In the following, the results are constrained to the trend tests of the entire time series (including the industrial period). Figure 4.3 shows the distribution of the p-values of the trend tests of the FVR data from the two climate model data sets. P-values from all three regression approaches and the two models have been pooled here. However, the figure differentiates between the p-values obtained from trend tests on FVR data from annual regressions (figure 4.3a) and climatological regressions (figure 4.3b). It appears that for the trend tests from the annual regressions, most trends are not significant at the 5% level. On the other hand, for the trend tests from the climatological regressions, a majority is significant, most even at the 1‰ level.



(a) *Trend p-Values for annual Regressions* ***(b)*** *Trend p-Values for climatological Regressions*

Figure 4.3: *P-Value distributions of the trend tests on FVR data from the climate models. (a) p-values from trend tests on FVR data obtained from regressions on the annual time scale, (b) p-values from trend tests on FVR data obtained from regressions on the climatological time scale. Intervals have been chosen so as to reflect typically used significance levels. Note that zero is included as a lower boundary in the left-most columns.*

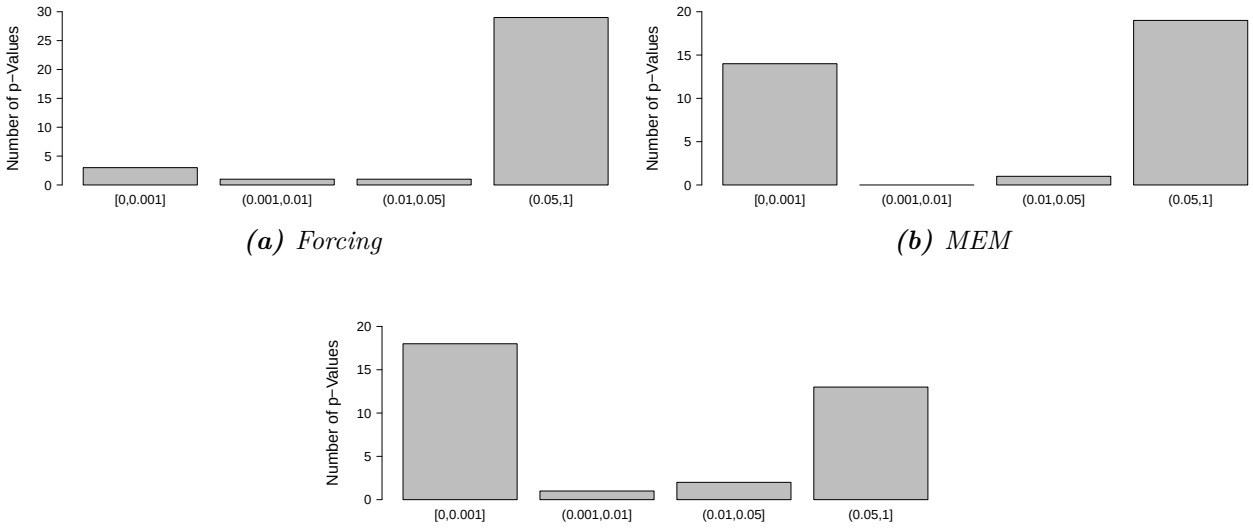
Figure 4.4 shows the same data as figure 4.3, but now separated according to the climate model. It is apparent that CESM has the higher fraction of significant trends, with roughly half of its trends being significant at the 5% level. For HadCM3 on the other hand, most trends are not significant at the 5% level.



(a) *CESM* ***(b)*** *HadCM3*

Figure 4.4: *Comparison of trend p-value distributions for the two model data sets (regression methods and time scales have been pooled).*

Figure 4.5 shows the same data as figures 4.3 and 4.4, but here the data are separated according to the regression method. It is evident that, for the FVR data from forcing regressions, most trends are not significant at the 5% level. For the FVR data from MEM regressions, about half of the trends are significant at the 5% level, with most of those being significant even at the 1‰ level. For GM-MEM, the distribution is similar, but shifted even a bit more towards higher significance.



(a) Forcing *(b) MEM*

(c) GM-MEM

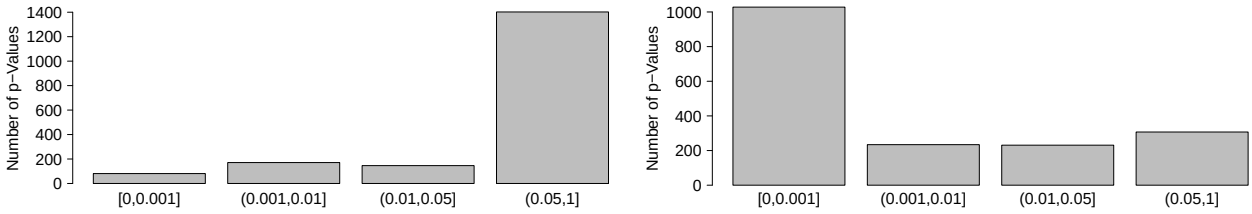
Figure 4.5: *Comparison of trend p-value distributions from climate model FVR data for the different regression methods*

4.2.2 Climate Field Reconstruction Data

For the entire time series (including the industrial period), of all the trends from the CFR FVR data (n = 3600), about half (52.5%) were significant at the 5% level. Of those, all were positive. The average magnitude of the significant trends ($\alpha = 5\%$) was found to be 0.014 units of FVR per 100 years.

For the CFR trend tests of only the pre-industrial FVR values, 35.6% of the trends were found significant at the 5% level. Of the significant ones, virtually all (99%) were found to be positive. The positive significant trends had an average magnitude of 0.017 units of FVR per 100 years, the negative ones an average magnitude of -0.011 units of FVR per 100 years.

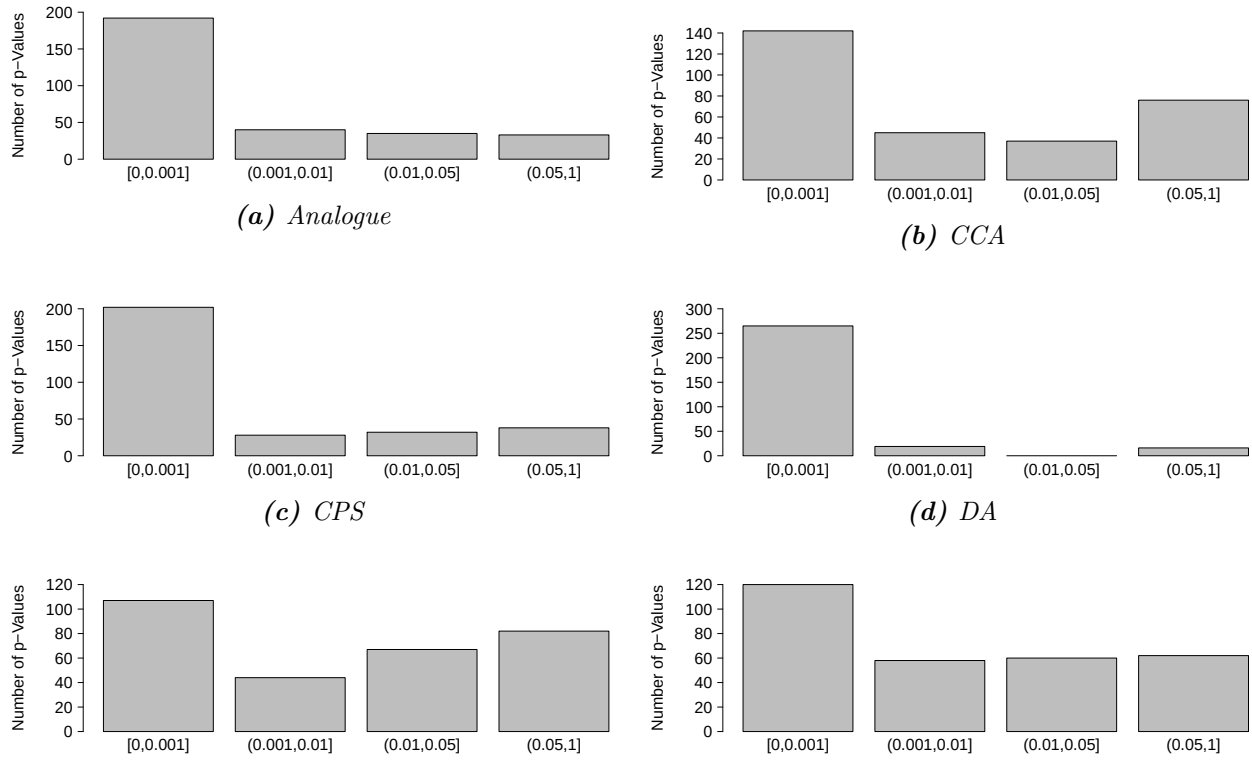
In the following, only results from the trend tests with the entire time series (including the industrial period) shall be presented. Much like figure 4.3, figure 4.6 shows the distribution of the p-values of the trend tests of the FVR data, but here for the CFRs. P-values from all three regression approaches and all six CFR methods have been pooled. Strikingly, most of the p-values from trends of FVR data on the climatological time scale are smaller than 5%, corresponding to a high proportion of tests revealing significant trends. Most p-values thereof are even smaller than 1‰, indicating extremely high significance of a large number of the trends. On the other hand, almost none the p-values from trend tests of FVR data on the annual time scale are smaller than 5%, meaning most of those trend tests reveal no significant trend. Because of that, the further analysis of the distribution of the p-values excludes the results of the trend tests on FVR data on the annual time scale.



(a) Trend p-Values for annual Regressions *(b) Trend p-Values for climatological Regressions*

Figure 4.6: *P-Value distributions of the trend tests on FVR data from the CFRs. (a) p-values from trend tests on FVR data obtained from annual regressions, (b) p-values from trend tests on FVR data obtained from climatological regressions. Intervals have been chosen so as to reflect typically used significance levels. Note that zero is included as a lower boundary in the left-most columns.*

Figure 4.7 shows the same p-values as figure 4.6b, but split according to the CFR method that the underlying FVR data come from. Apparently, the trend test on the FVR data from the Analogue, CPS and DA CFRs have the highest fractions of significant trends, both at the 5% and at the 1‰ level. On the other hand, of the trend tests on the FVR data from the CCA, GraphEM and PCR CFRs, generally fewer trends are significant.



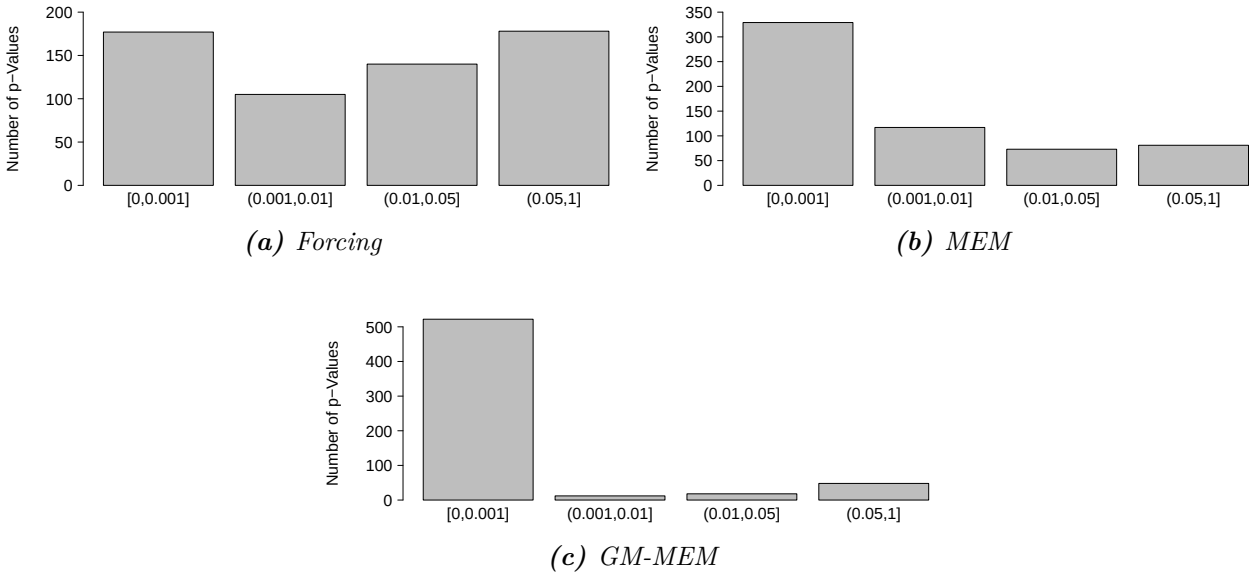
(a) Analogue *(b) CCA*

(c) CPS *(d) DA*

(e) GraphEM *(f) PCR*

Figure 4.7: *Comparison of trend p-value distributions for the different CFR data sets (regression methods have been pooled, only data from the climatological time scale)*

Figure 4.8 shows the same data as figures 4.6b and 4.7. However, here the data are separated according to the regression method with which the FVR data have been produced. It is evident that the FVR data from regressions with forcing time series show much fewer significant trends than those from regressions with MEM and GM-MEM. The most extreme case are the trends from FVR data from regressions with GM-MEM, of which most are significant at the 1‰ level.



(a) Forcing *(b) MEM*

(c) GM-MEM

Figure 4.8: *Comparison of trend p-value distributions from CFR FVR data, for the different regression methods (only data on the climatological time scale are shown).*

4.2.3 Interpretation

For the trend results from the climate model FVR data, the higher number of significant trends in CESM as compared with HadCM3 could be due to the longer time series of the CESM. Therefore, based on this enhanced sample size, it could be easier to detect a significant trend, even when trend magnitudes in HadCM3 are slightly higher. However, the longer time series in CESM also imply that the industrial period with its high values of FVR is less important, relatively. This is also reflected in the slightly higher trend magnitudes in HadCM3. When the industrial period is excluded from the analysis, this has a strong effect of the trends from HadCM3. Now, most of its significant trends are negative. This can also be seen in the evolution of the Area-Weighted Global Mean and Ensemble Mean (GM-EM) FVR from climatological regressions in figures 4.1b and 4.1c, where the FVR values of HadCM3 drop substantially from moving window start time 1401 to 1601 (the period considered for the pre-industrial only trend tests). This has even led to an increment in the number of significant trends in HadCM3. All of these negative trends come from climatological MEM and GM-MEM regressions, corresponding to the above-mentioned figures.

In general, the upward trends in FVR observed could be due to a combination of three factors: The first is a natural increase of FVR over time. The second is the impact of the high-forced industrial period. The third is the increase in quality of reconstructed forcings (in forcing regressions and also in

the forcing drivers of the two climate models) and environmental proxy records (in the CFRs) with time. The latter issue shall be discussed in greater detail in section 5.3. However, the three effects can also be disentangled by looking at the trend tests with only pre-industrial values. Naturally, the effect of the industrial period is eliminated there. Nevertheless, still the large majority of trends is positive, with the exception of the HadCM3. Therefore, it seems likely that either FVR has increased naturally over time or that the improving quality in reconstructed forcing time series and environmental proxy records led to positive trends in FVR. For a natural increase of FVR over time, either the variance of forcing would need to have increased over time, or the variance of internal variability decreased. For both cases, no indications or obvious explanations are at hand.

The lower number of significant trends in the pre-industrial only trend tests could be due either to the shorter time series and reduced sample size or to the exclusion of the industrial period with its high values (leading to significant upward trends), or due to a combination of both.

The trend tests with FVR data from annual regressions generally revealed little significant trends. One explanation for that could be that the FVR values from annual MEM and GM-MEM regressions are generally very low and close to zero during most of the time, therefore not allowing for trends. That would mean that most of the significant trends from FVR data from annual regressions should come from forcing regressions. This was checked for both models and CFRs. For the models, this check revealed that roughly the same amount of significant trends from annual regression data came from forcing regressions and MEM regressions. This makes the above explanation unlikely at least for the models. However, it also makes sense, since the GM-EM FVR time series are not so close to zero, at least for CESM (figures 4.1b and 4.1c). Likewise, subfigure 4.5a reveals that FVR data from forcing regressions in the models generally exhibit few significant trends. For the CFRs, on the other hand, the above explanation seems likely. There, a check revealed that from all the significant trends from FVR data from annual regressions, a large majority came from forcing regressions (241 vs. 30 from MEM and 127 from GM-MEM). This could be due to the better agreement of the CFR temperature data with the CMIP6 forcing as opposed to the CESM MEM and GM-MEM (which are based on different forcings).

For both the models and the CFRs, the magnitude of the trends in general seems realistic. Of course, such an increase (or decrease) in FVR cannot continue indefinitely, since FVR would inevitably reach its boundaries by definition (zero and one). However, for the time periods considered, the numbers presented seem to be realistic. The last moving window analyzed entirely covers the strongly forced anthropogenic period. Slightly higher values are possible due to the still rising anthropogenic forcing. However, the current FVR values are likely close to the maximum that can be expected to occur. Thus future trends strongly depend on the development of anthropogenic forcing.

4.3 Global-Mean FVR: Comparison between Data Sets

This section will present the results of comparisons of GM FVR values between the different data sets. As a visual tool to this end, boxplots are used to depict the median values and spread of FVR of each data set. Note that, in order to enhance contrast and readability, the scales of the different figures have been individually adjusted. Also, the boxplots show the spread of only those moving windows that started in 1401 or after. This was chosen so that the two model and the six CFR data sets are made comparable (the period 1401 – 2000 is covered by all of them, with HadCM3 starting in 1401). However, the consequence is that the FVR data from HadCRUT4 and 20CRv2 are not shown, since they only cover the period 1851 – 2000 and therefore are not readily comparable to the other data sets. In order to quantitatively corroborate the information from the boxplots, the results of the ANOVA and Student's t-tests delineated in section 3.4 are given.

4.3.1 Climatological Time Scale: Results from Forcing Regressions

Figure 4.9 shows the distributions of the GM FVR data from regressions with forcing time series for the different data sets. On a first order, all the data sets seem to agree quite well in the overall magnitude of FVR. This is consistent with the finding described in 4.1.1 and figure 4.1a, where the ensemble mean values of the same data have been shown to closely agree in the magnitude of FVR. However, an ANOVA performed on the data shown revealed that there are significant differences between the data sets ($p < 0.001$). Moreover, pairwise t-tests showed that the group means are significantly different at the 5% level for all data set pairs except for CCA & GraphEM and CESM & GraphEM.

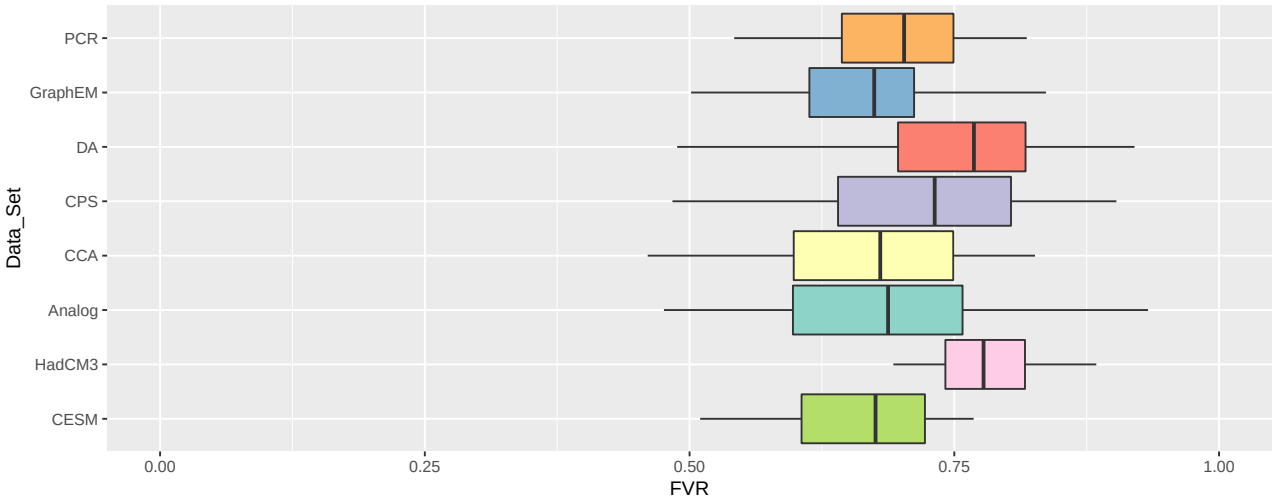


Figure 4.9: *Boxplots of the GM FVR values from climatological forcing regressions, separated according to data set. The data from the different ensemble members and moving windows of each data set have been pooled to form the distributions shown here. Note that only moving windows with start times $\geqslant$ 1401 have been used so as to make the data sets comparable. The central line in the boxes marks the median of the respective distribution, the boxes denote the 25^{th} and 75^{th} percentiles. Whiskers mark the 5^{th} and 95^{th} percentiles of the distributions. Due to their large number, the outliers have not been plotted.*

4.3.2 Climatological Time Scale: Results from MEM Regressions

Figure 4.10 is similar to figure 4.9, but it shows the FVR data generated from climatological MEM regressions. Here, the differences between the data sets seem to be more pronounced, and the two models in particular show higher FVR than the CFRs. Also, the boxes (but not necessarily the whiskers) seem to be wider than in figure 4.9. Also here, an ANOVA reveals that the differences between the data sets are highly significant ($p < 0.001$). Pairwise t-tests indicate that all pairs of data sets are significantly different from each other at the 5% level, with the exception of the Analogue & DA combination.

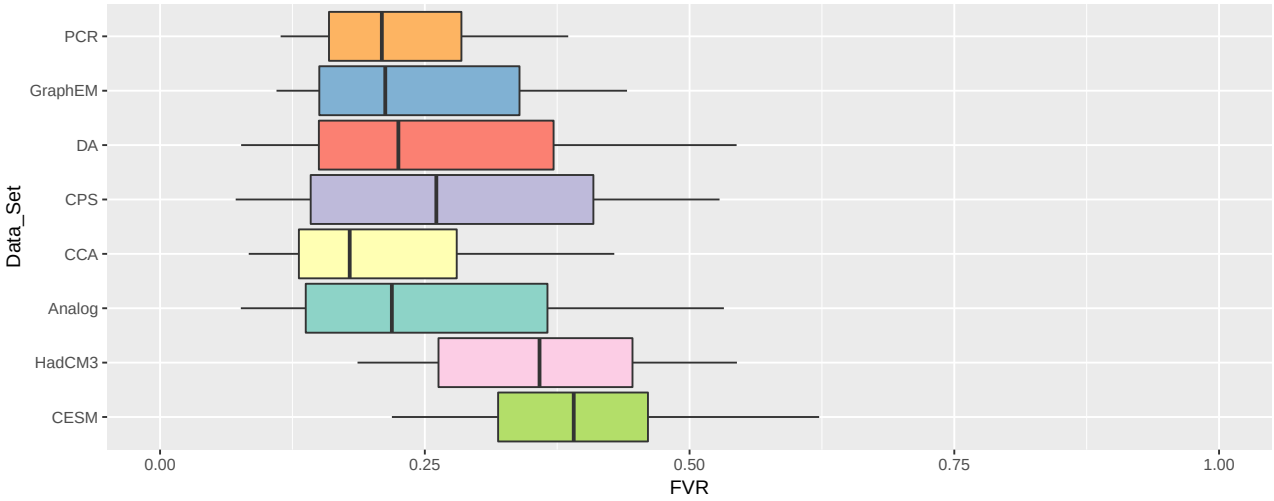


Figure 4.10: *The same as in figure 4.9, but for FVR data produced from regressions using MEM.*

4.3.3 Climatological Time Scale: Results from GM-MEM Regressions

The distribution of GM FVR values from climatological GM-MEM regressions is visualized in figure 4.11. As in figures 4.9 and 4.10, the data are separated according to the data set. Some features here are similar to figure 4.10, namely the models having higher FVR than the CFRs. However, HadCM3 here shows higher FVR than CESM. Among the CFRs, CCA, and partly PCR seem to have substantially lower FVR than the other methods. Indeed, also here an ANOVA reveals significant differences between the data sets ($p < 0.001$). Pairwise t-tests indicate significant differences ($\alpha = 5\%$) between all combinations of data sets, except for CESM & CPS and DA & GraphEM.

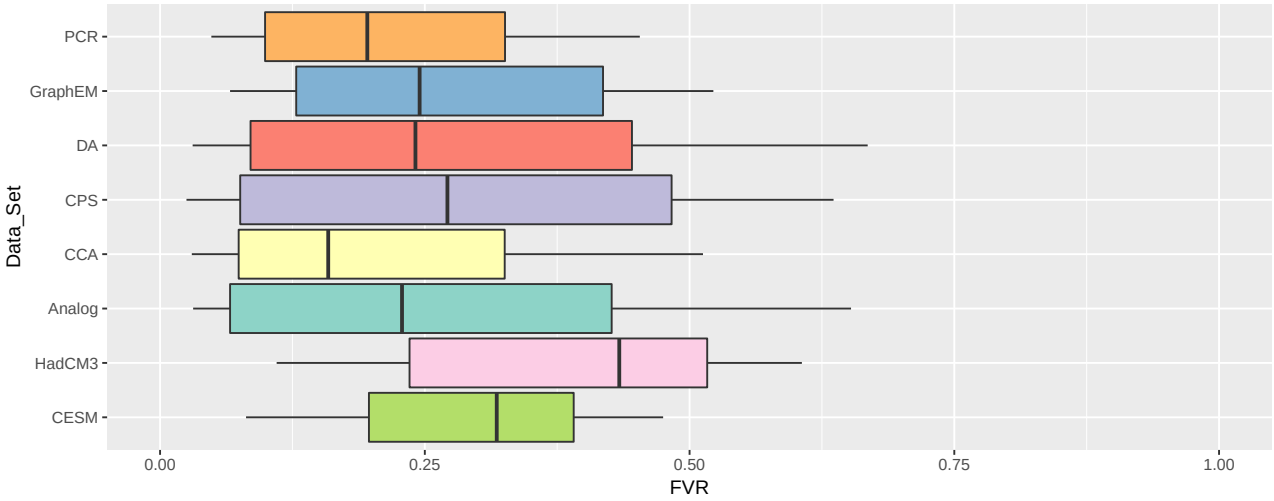


Figure 4.11: *The same as in figure 4.9, but for FVR data produced from regressions using GM-MEM.*

4.3.4 Annual Time Scale: Results from Forcing Regressions

Figure 4.12 is similar to figure 4.9, except that here the FVR data from annual forcing regressions are shown. Note also the different scale as compared to figure 4.9. Here, the median values of the data sets seem to be rather diverse, and especially HadCM3 and DA have much higher FVR than the other data sets. Also, the overall FVR is rather low. The 5^{th} to 95^{th} percentile spread seems quite large in comparison to the interquartile range in some data sets. This equates to the distributions having long tails. Also here, ANOVA reveals significant differences between the data sets ($p < 0.001$). Pairwise t-tests indicate significant differences ($\alpha = 5\%$) between all pairs of data sets except for CESM and GraphEM.

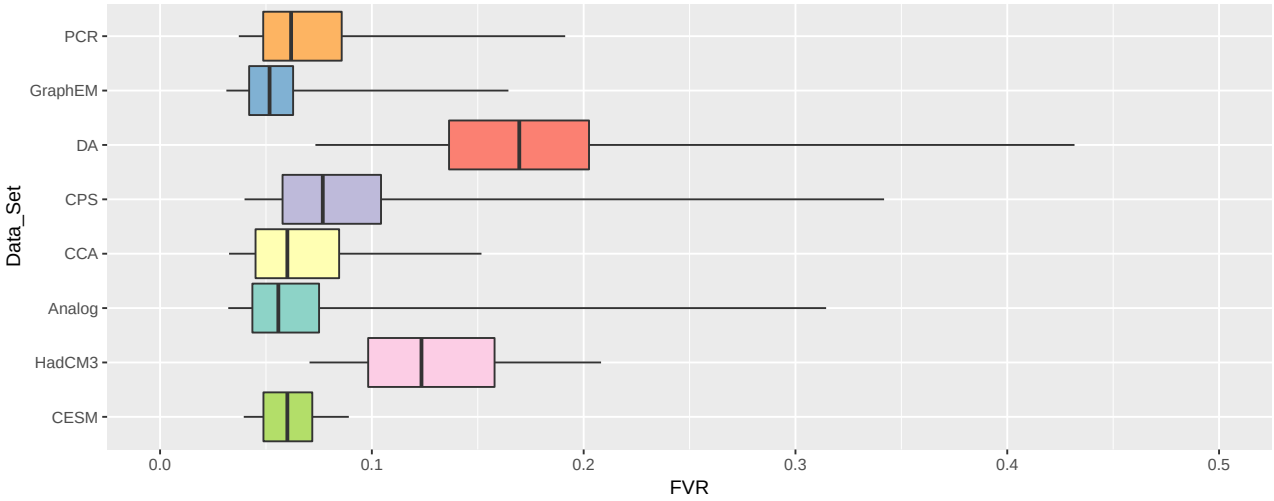


Figure 4.12: *The same as in figure 4.9, but for FVR data produced from annual forcing regressions. Note that the scale is different than in figures 4.9 - 4.11.*

4.3.5 Annual Time Scale: Results from MEM Regressions

Like figure 4.9, figure 4.13 shows the distribution of GM FVR values, but here from annual MEM regressions. Here, the very low FVR values are striking, something already alluded to in the introduction of section 4.1. The only exception thereof are the FVR values from the CESM, but even those range only around 0.1. Overall, there seems to be close agreement between all the data sets, except for CESM. However, an ANOVA shows significant differences between the data sets ($p < 0.001$). Pairwise t-tests reveal significant ($\alpha = 5\ \%$) differences between all pairs of data sets, except for the combination of CCA & GraphEM.

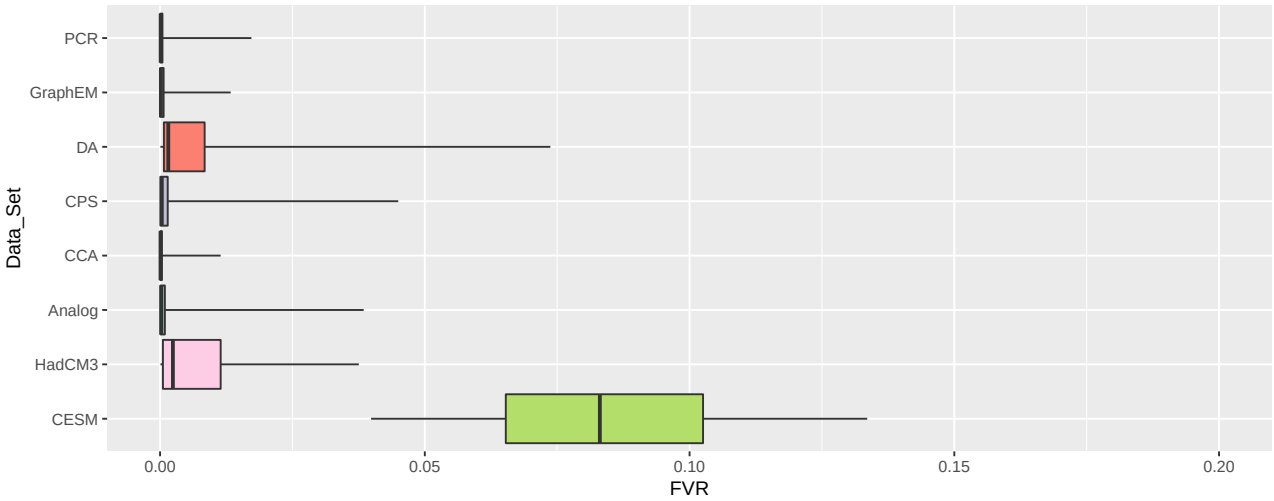


Figure 4.13: *The same as in figure 4.9, but for FVR data produced from annual MEM regressions. Note the different scale as compared to figures 4.9 - 4.12. The comparison of means with ANOVA and t-tests requires the assumption of normal distribution. It needs to be stated that this assumption likely is not fulfilled here and neither in the following figure 4.14.*

4.3.6 Annual Time Scale: Results from GM-MEM Regressions

Figure 4.14 shows the distributions of GM FVR values for annual regressions with GM-MEM, much like in the previous figures. As in figure 4.13, the overall FVR values are very low. There seem to be some differences between data sets, with CESM, HadCM3 and DA having somewhat higher FVR values than the other data sets. Indeed, also here an ANOVA indicates significant differences among data sets ($p < 0.001$). Pairwise t-tests indicate significant differences ($\alpha = 5\%$) in most combinations of data sets, with the exceptions of CESM & CPS, CESM & HadCM3, GraphEM & PCR and HadCM3 & CPS.

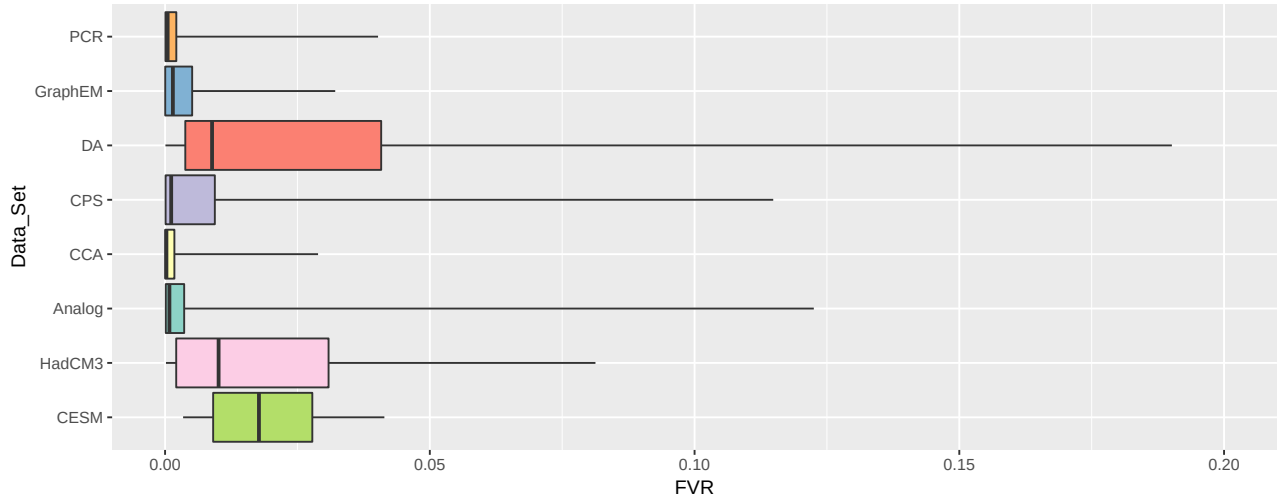


Figure 4.14: *The same as in figure 4.9, but for FVR data produced from annual GM-MEM regressions. Note the different scale as compared to figures 4.9 - 4.12.*

4.3.7 Interpretation

Some data sets exhibited significant differences, even when the FVR values were very close together. This can be seen, for example, in figure 4.13, where most of the FVR values are very close to zero. The differences being significant even when they were very small can be explained by the very large sample sizes of the different data sets. This is most pronounced for the CFR data, which each have samples of size 4600 (46 moving windows from 1401 to 1851 with 100 ensemble members each).

In some instances, the FVR values from the CESM seem to have been enhanced over the ones from the other data sets. This can be seen most clearly in figure 4.13, but also in figures 4.10, 4.11 (although here HadCM3 is even higher) and 4.14. A very likely explanation for this observation is that the agreement between CESM and its own gridpoint-wise MEM is higher than the one of the other data sets with the CESM MEM. The effect seems less pronounced when the GM-MEM is considered (figures 4.11 and 4.14, as compared to figures 4.10 and 4.13).

Also the DA CFR shows somewhat enhanced FVR values as compared to the other data sets. This could be due to its property to generally underestimate the uncertainty in the reconstructed temperature.

Overall, even though there were significant differences detected between most of the data sets, it can be said that the FVR values of the different data sets mostly fall within the same ranges. This is most apparent for the climatological forcing regressions (figure 4.9). Therefore, there seems to be some agreement between data sets on the overall magnitude of FVR.

4.4 Global-Mean FVR: Comparison between Regression Methods

This section aims at comparing the GM FVR values of instrumental and reanalysis data, models and CFRs between the three (six, respectively) regression methods described in section 3.2. Much like in section 4.3, boxplots are used to visualize the median values and spread of FVR for each regression method. This section is structured into three subsections, one for each of instrumental and reanalysis, model and CFR data. In every subsection, the FVR values from the three regression methods are compared for the annual and climatological time scales. To simplify the analysis, and to enable comparison of regression methods (as opposed to data sets, section 4.3), the FVR data from the two model data sets, respectively the six CFR data sets, have been pooled to form single vectors of FVR values for every regression method. The data from the different ensemble members and moving windows have also been pooled. As in section 4.3, the scales of the plots have been individually adjusted to enhance contrast and readability. Similarly, to quantitatively underpin the information from the boxplots, the results of the Student's t-tests and ANOVA (see section 3.4) are given.

4.4.1 Instrumental and Reanalysis Data

Figure 4.15 shows the GM FVR values for 20CRv2 and HadCRUT4, all regression methods and both time scales. Here, in contrast to subsections 4.4.2 and 4.4.3, single values of GM FVR instead of distributions are displayed. The reason is that, for the instrumental and reanalysis data sets, only a single map of FVR was generated for each regression method and time scale, with a single ensemble member and a single moving window. It is most apparent that the climatological regressions produced higher FVR values than the annual regressions. Furthermore, for both climatological and annual regressions, the forcing regressions seem to have produced higher FVR values than MEM and GM-MEM regressions. Of the latter two, GM-MEM regressions generally produced a bit higher FVR values than MEM regressions. Also, HadCRUT4 seems to have consistently higher FVR values than 20CRv2.

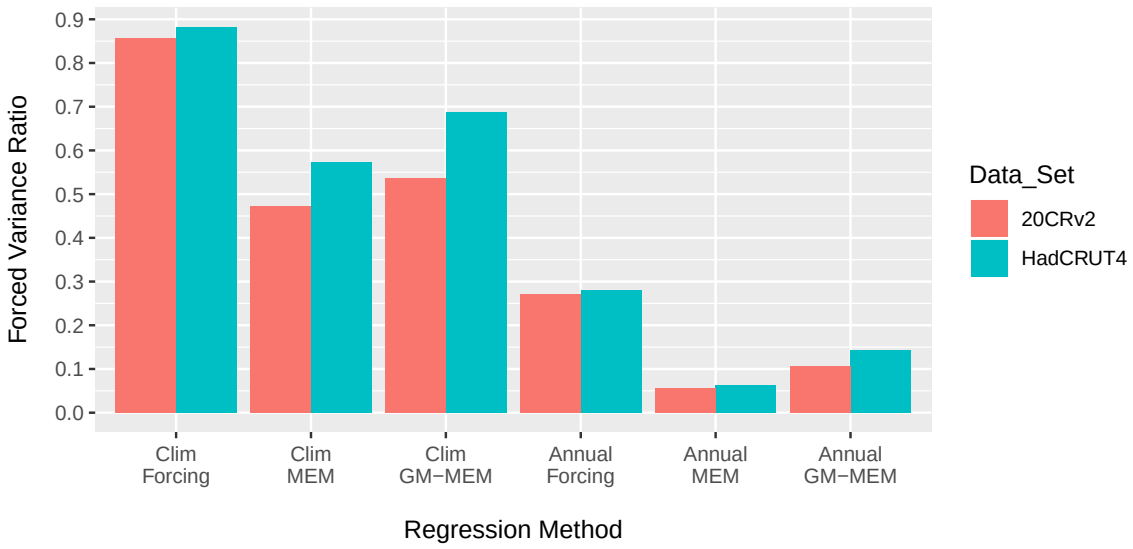


Figure 4.15: *GM FVR values for the instrumental and reanalysis data sets, all three regression methods and both time scales.*

4.4.2 Climate Model Data

Figure 4.16 shows the distributions of GM FVR of the two climate models for the three regression methods on the climatological time scale. The median values range between around 0.3 for GM-MEM regressions and around 0.7 for forcing regressions. The distributions of the FVR values from MEM and GM-MEM regressions show substantial overlap, whereas the one of the FVR values from forcing regressions is quite separate and substantially higher. The ranges between the 5^{th} and 95^{th} percentiles seem to be roughly equal among all three regression methods. The interquartile ranges, on the other hand, are relatively small for forcing regressions and large for GM-MEM regressions. The one for MEM is somewhere in between. An ANOVA reveals significant differences in means among the three regression methods ($p < 0.001$). Likewise, pairwise t-tests between the three regression methods indicate significant difference in means in all three combinations ($p < 0.001$ in all instances).

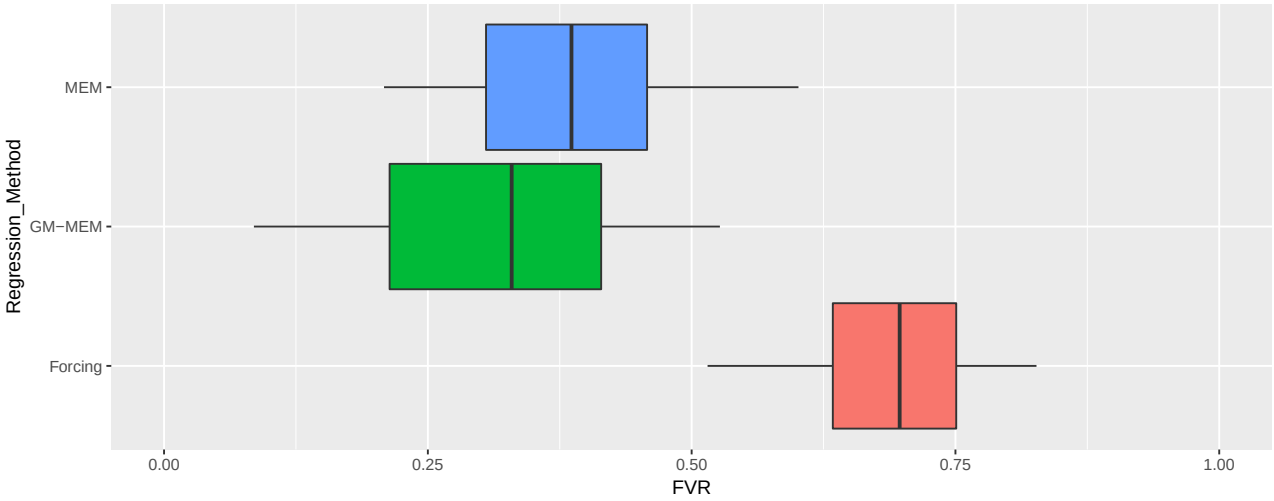


Figure 4.16: *Boxplots of the GM FVR values of the climate models, separated according to the three regression methods. The FVR values from the two climate models have been pooled, as well as the ones from the different model ensemble members and moving windows. The central line in the boxes marks the median of the respective distribution, the boxes denote the 25^{th} and 75^{th} percentiles. Whiskers mark the 5^{th} and 95^{th} percentiles of the distributions. Due to their large number, the outliers have not been plotted.*

The distributions of the GM FVR values from regressions on the annual time scale, separated according to the three regression methods, are shown in figure 4.17. Here, the FVR values are much smaller than for the regressions on the climatological time scale, as shown in figure 4.16. Apparently, GM-MEM regressions produced the lowest FVR values here, while the ones from MEM and forcing regressions are generally higher and of about the same magnitude. An ANOVA indicates significant differences in means among the three regression methods ($p < 0.001$). Similarly, pairwise t-tests of the three regression methods show significant difference in means between all pairs ($p < 0.001$ in all instances).

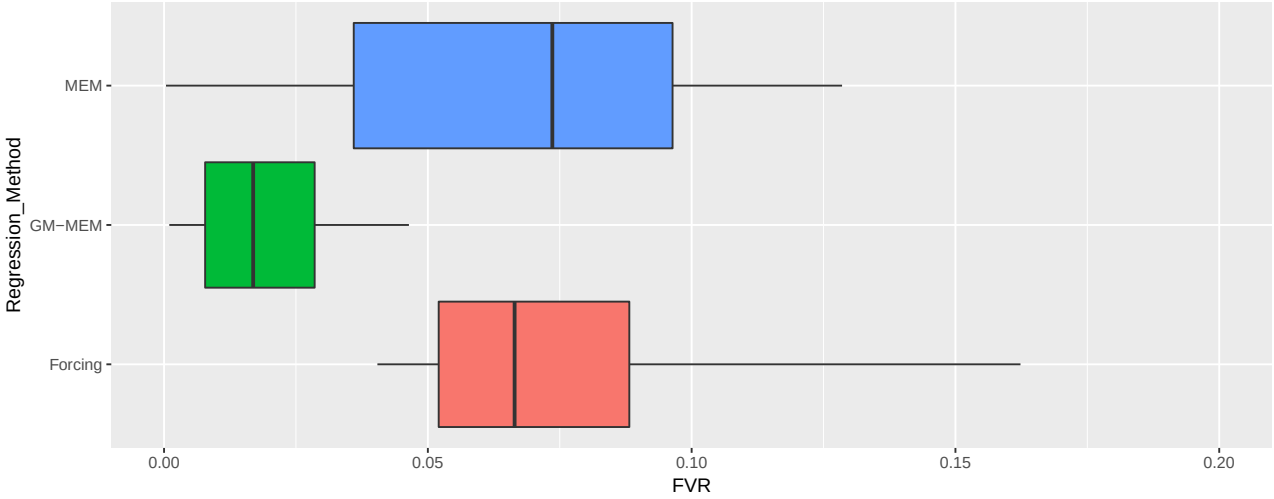


Figure 4.17: *The same as in figure 4.16, but for FVR data from regressions on the annual time scale. Note the different scale on the x-axis as compared with figure 4.16.*

4.4.3 Climate Field Reconstruction Data

Figure 4.18 shows the distributions of GM FVR of the CFR data sets, from regressions on the climatological time scale. As in figures 4.16 and 4.17, the data have been separated according to the three regression methods, to enable comparison between them. Results seem to be qualitatively similar to the ones from the climate models (see figure 4.16), with the FVR values from forcing regressions being substantially higher than the other two. Those are of about the same magnitude. An ANOVA indicates significant differences in means among the three regression methods, being highly significant ($p < 0.001$). Similarly, pairwise t-tests reveal highly significant differences in all three combinations ($p < 0.001$ in all instances).

Figure 4.19 shows the distributions of GM FVR for the CFRs, from annual regressions. Again, the data have been separated according to regression method. Here, MEM and GM-MEM regressions have produced extremely low FVR values, as already mentioned in the introduction of section 4.1 and in subsection 4.3.5. FVR values from forcing regressions are substantially higher, but still rather low, with a median value of around 0.06. Also for these data, an ANOVA revealed highly significant ($p < 0.001$) differences between mean values of the three regression methods. Likewise, pairwise t-tests indicate highly significant differences in all three combinations ($p < 0.001$ in all instances).

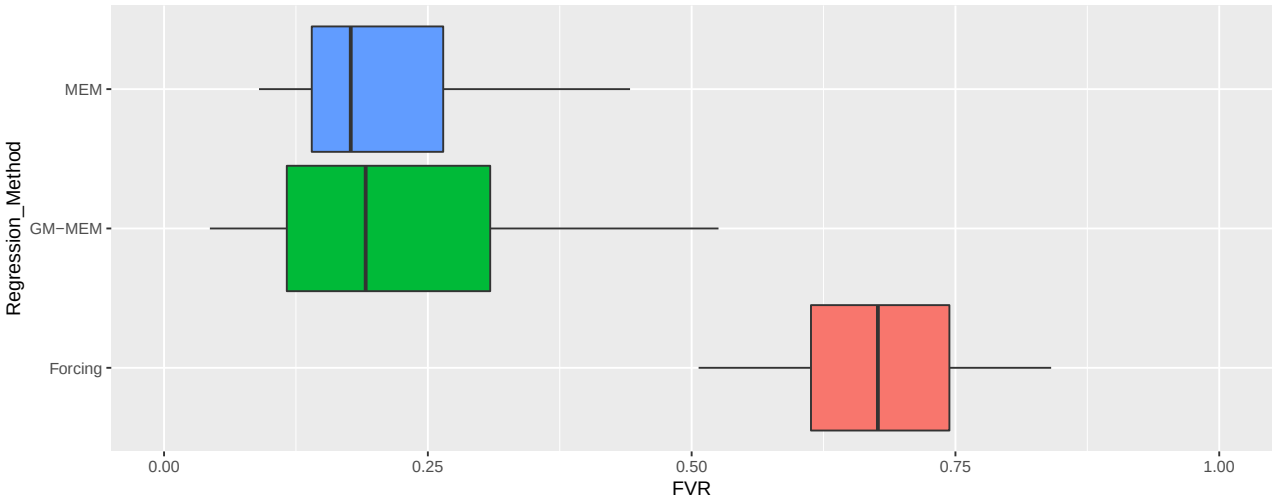


Figure 4.18: *Boxplots of the GM FVR values of the CFRs, separated according to the three regression methods. The FVR values from the six CFR methods have been pooled, as well as the ones from the different ensemble members and moving windows. The central line in the boxes marks the median of the respective distribution, the boxes denote the 25^{th} and 75^{th} percentiles. Whiskers mark the 5^{th} and 95^{th} percentiles of the distributions. Due to their large number, the outliers have not been plotted.*

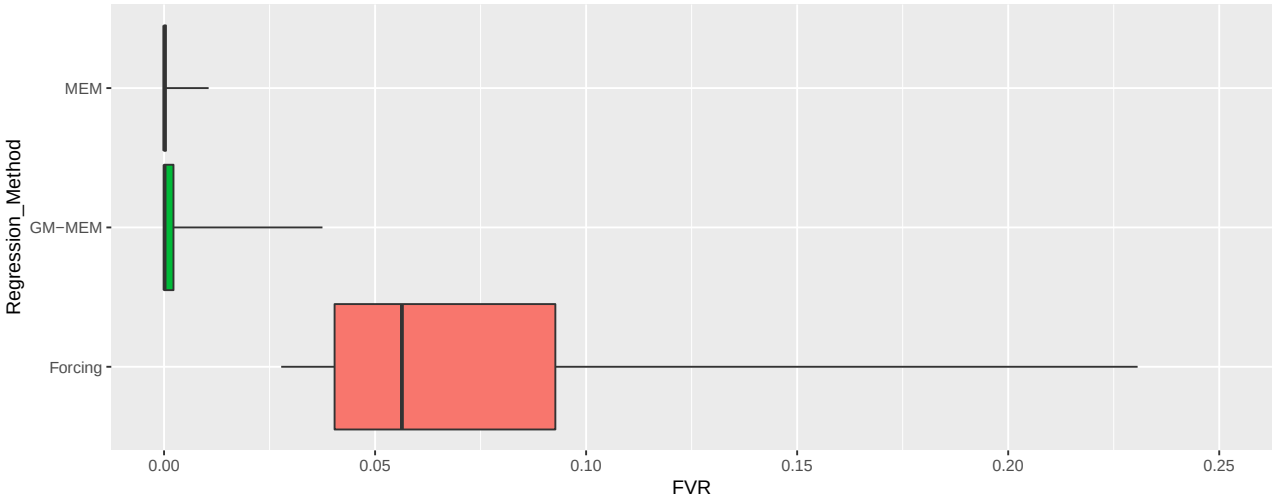


Figure 4.19: *The same as in figure 4.18, but for FVR data from regressions on the annual time scale. Note the different scale on the x-axis as compared with figure 4.18.*

4.4.4 Interpretation

As was already shown in figure 4.2, and is also evident from figures 4.16 and 4.18, the FVR values from climatological forcing regressions are substantially higher than those for the other climatological regressions. Due to this low agreement, the confidence in the results from the climatological forcing regressions is questionable. A possible explanation for the enhanced FVR values from climatological forcing regressions could be multicollinearity, leading to overestimation of R^2 values (Gujarati & Porter 2010). This was tested for (3.2.4), the results are described in 4.11.3.

Also, it is most apparent that GM FVR values are much lower for annual regressions than for climatological ones. At first glance, one might conclude that FVR is lower per se on the annual time scale. This is plausible when bearing in mind that even in annual means, substantial variability in surface temperature can be found (the time scale is still close to the one of weather). This could be regarded as noise in the time series. On the other hand, on the climatological time scale, the inter-annual variability is filtered out and the much more slowly changing climatic signal becomes apparent. For this time scale, the noise component is much lower and, therefore, the signal from forcing should be stronger. However, the explanation for the higher R^2 in climatological regressions may also be a purely statistical one: Due to the filtering of the time series before the regressions, they have become autocorrelated and their effective degrees of freedom were reduced, which leads to higher values of R^2 (e.g. Mukherjee et al. 1998).

Similar to subsection 4.3.7, also here the detection of significant differences even with small absolute differences (e.g. figure 4.18, MEM and GM-MEM) can be explained with the very large size of the samples.

The higher values of MEM as compared to GM-MEM for the model data can be explained by the influence of the CESM, at least for annual regressions, where the difference is large (figure 4.17). There, CESM seems to match better with its own gridpoint-wise MEM than with its GM-MEM. This carries through, even when the data from the two models are pooled, because CESM has more data points by far (598 vs. 184 for HadCM3).

4.5 Spatial Structure of the FVR: Zonal Mean

4.5.1 Instrumental and Reanalysis Data

Figure 4.20 shows the zonal-mean FVR for the 20CRv2 data set, for all three regression methods and both time scales. Generally, the FVR values from climatological regressions are higher than those from annual regressions. The FVR values from climatological forcing regressions are very high throughout all latitudes, with somewhat lower values in the northern hemispheric mid-latitudes. The distribution of FVR values in the climatological MEM and GM-MEM regressions is quite similar, but complex. They both share a maximum in the Arctic and minima in the mid-latitudes of both hemispheres. The FVR data from climatological GM-MEM regressions also show a local minimum in the tropics. In the annual regressions, all three regression methods agree in a maximum of FVR in the Arctic. For annual forcing regressions, there is also a maximum in the Antarctic. All three methods have in common a local minimum in the mid-latitudes of both hemispheres. The distributions of the zonal-mean FVR for HadCRUT4 are qualitatively very similar to the ones from 20CRv2 shown here (results not shown).

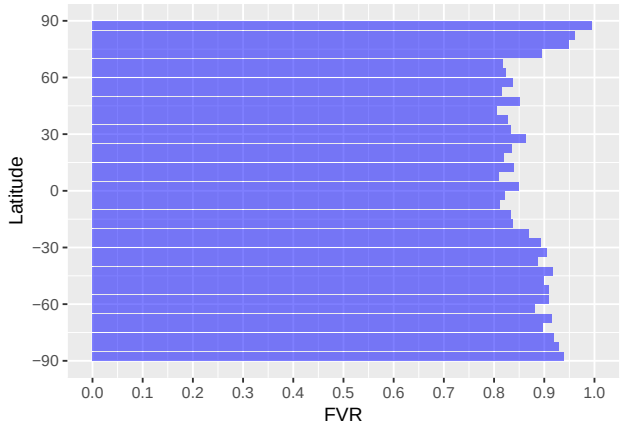
4.5.2 Climate Model Data

Figure 4.21 shows zonal-mean FVR data from the two climate models, much like figure 4.20. However, here also the ensemble spread of the respective models has been visualized. Error bars denote the ensemble mean plus/minus one standard deviation of ensemble members. In addition, figure 4.21 displays the climate model FVR data as a time-mean of the respective data set. That is, for CESM, data from 851–2000 are shown, for HadCM3, data from 1401–2000.

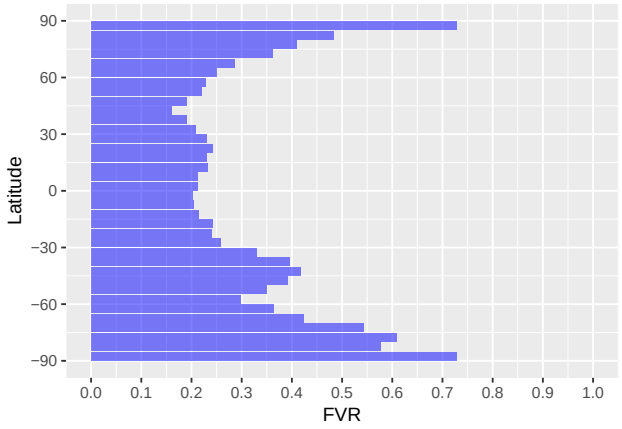
The zonal-mean FVR data from the two climate models show very high agreement. Both models, all three regression methods and both timescales are very similar. All agree in a global maximum of FVR in the tropics, and lower values in the mid-latitudes of both hemispheres. In addition, the FVR data from climatological regressions also show local maxima in the high latitudes of both hemispheres. This can be seen in figure 4.21a. However, it needs to be stated that these polar local maxima are accompanied by high ensemble spread and thus high uncertainty.

To enable comparison of climate model and CFR FVR data, the zonal means have also been computed over the time 1401–2000. For the climate model data, the results were quite similar to those over the entire time, as depicted in figure 4.21. For HadCM3, this is the case by definition, as the ensemble covers only the aforementioned period.

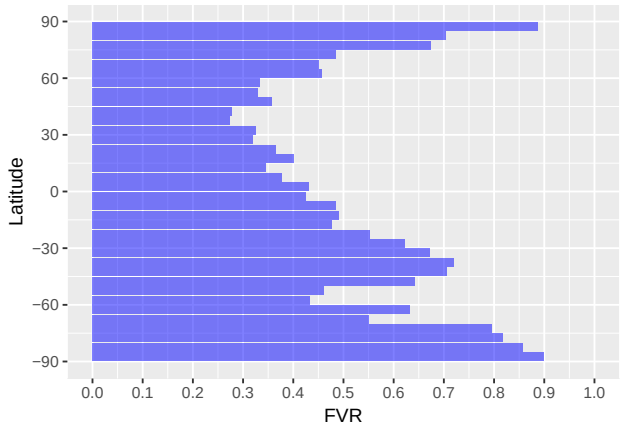
A check of the zonal-mean FVR for the last moving window (1851–2000) was also performed to compare the climate model data to the instrumental and reanalysis data and the CFR data in this time. For climatological forcing and MEM regressions, good agreement could be found between the climate model and the instrumental and reanalysis data. For annual forcing and MEM regressions, agreement could be found between CESM and the instrumental and reanalysis data sets, but not between HadCM3 and the instrumental and reanalysis data sets. For both annual and climatological GM-MEM regressions, some qualitative agreement could be found.



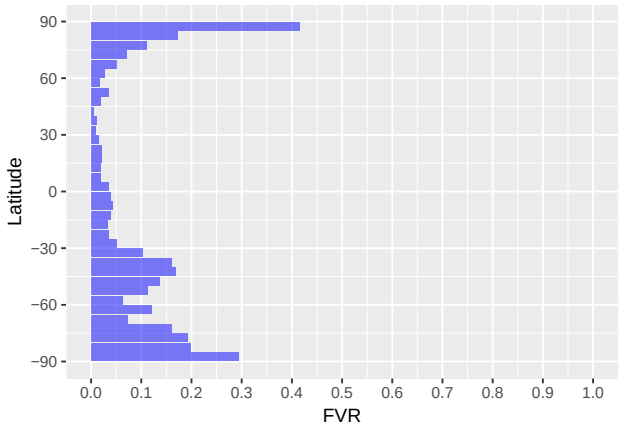
(a) 20CRv2, Climatological Forcing Regressions



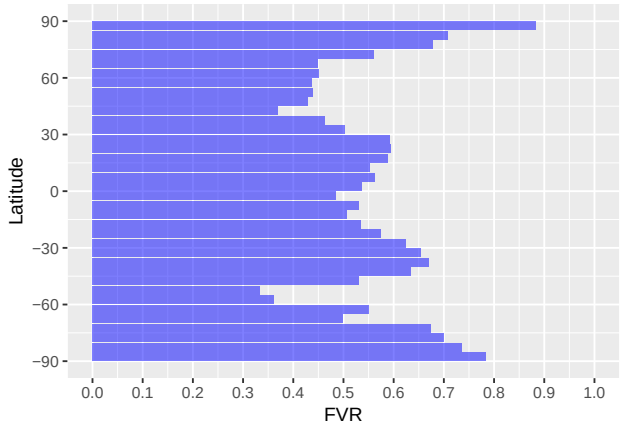
(b) 20CRv2, Annual Forcing Regressions



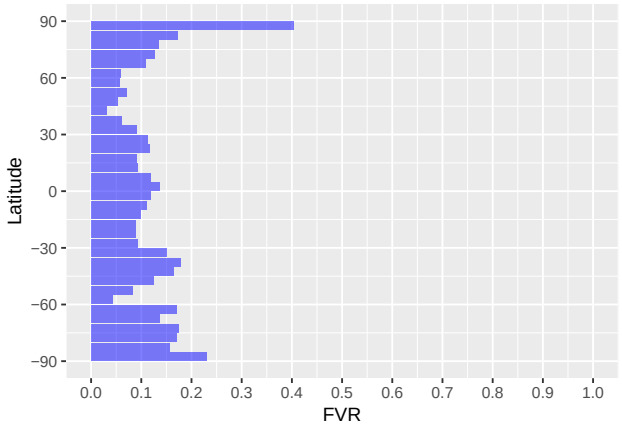
(c) 20CRv2, Climatological MEM Regressions



(d) 20CRv2, Annual MEM Regressions

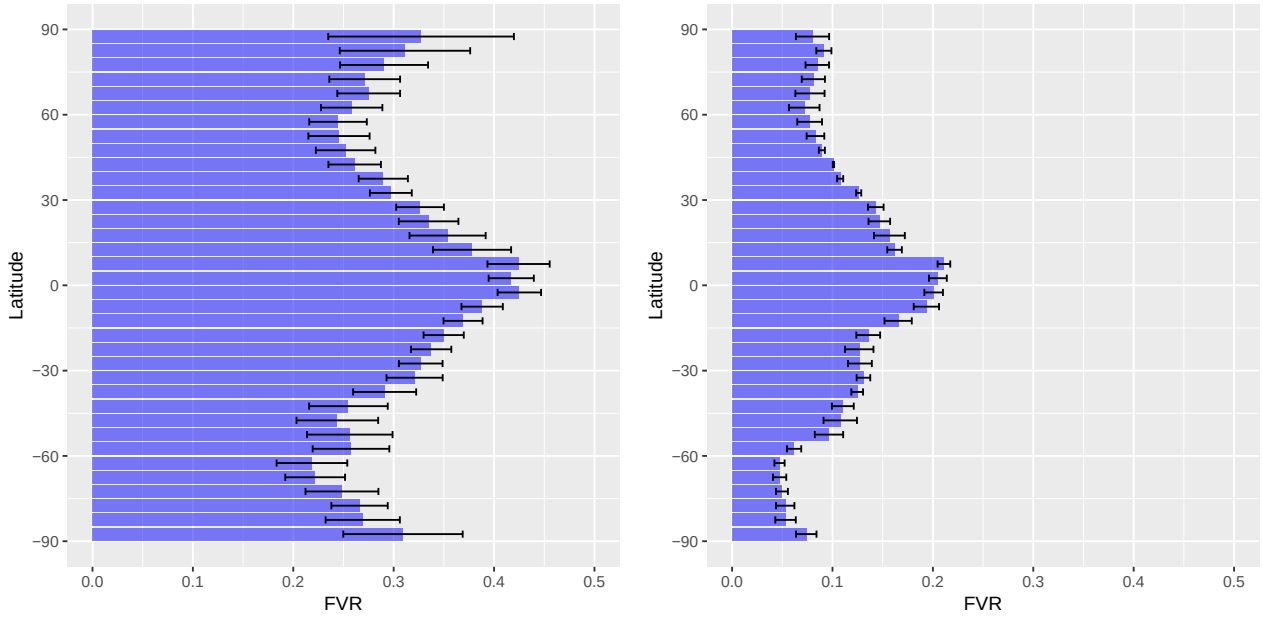


(e) 20CRv2, Climatological GM-MEM Regressions



(f) 20CRv2, Annual GM-MEM Regressions

Figure 4.20: *Zonal-Mean FVR for the 20CRv2 data set and the different regression methods and time scales.*

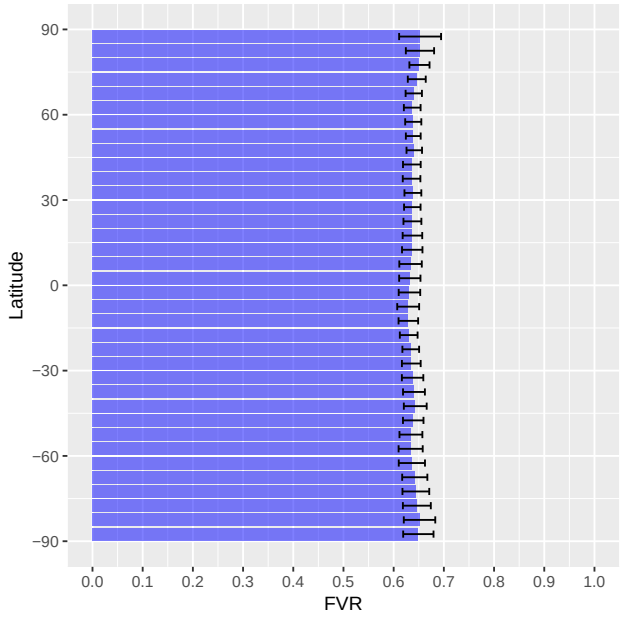


(a) CESM, Climatological GM-MEM Regressions *(b) HadCM3, Annual Forcing Regressions*

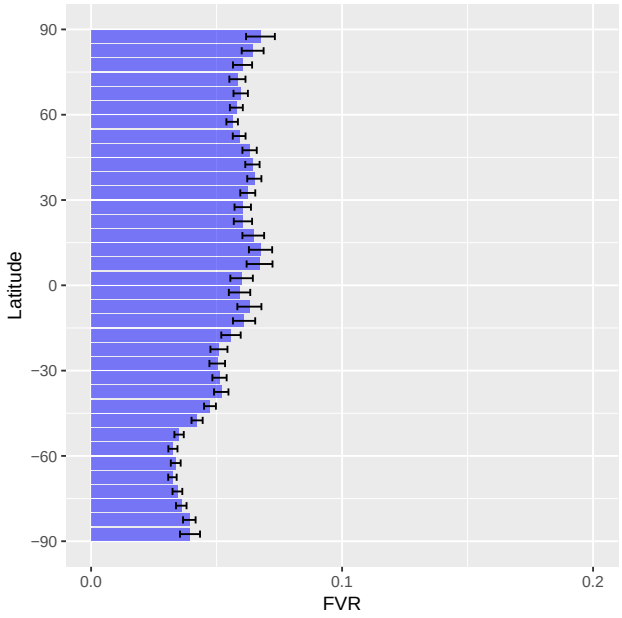
Figure 4.21: *Zonal-mean FVR for the two climate model data sets, for selected regression methods and time scales. Data shown are for the entire time of data availability of the two models, that is 851–2000 CE for CESM and 1401–2000 CE for HadCM3. The blue bars denote the zonal-mean ensemble-mean FVR, whiskers denote the ensemble mean plus/minus one standard deviation of ensemble members.*

4.5.3 Climate Field Reconstruction Data

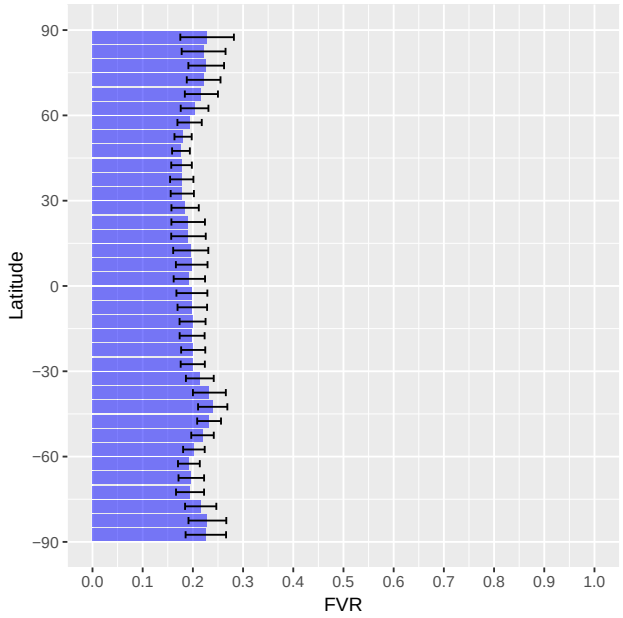
Figure 4.22 shows the zonal-mean FVR for selected CFR data sets, for climatological and annual forcing and MEM regressions. The results from GM-MEM regressions are not shown, as they are qualitatively very similar to the results shown here. All data sets agree in rather uniform distributions of zonal-mean FVR, with small ensemble spreads throughout. For the FVR data from annual forcing regressions, a tendency for higher FVR values in the northern hemisphere and the tropics can be observed. For the annual MEM regressions, very low FVR values can be observed, with small maxima in the Arctic and the southern hemispheric mid-latitudes. In both climatological FVR data shown here, the uncertainties from the ensemble spreads cover the range of variation of FVR with latitude. Therefore, they can be said to be independent of latitude. Similar to the climate model data, a check of the zonal-mean FVR for the last moving window (1851–2000) was performed to compare the CFR data to the instrumental and reanalysis data. The comparison revealed high consistency between instrumental and reanalysis and CFR data in all three regression methods and both annual and climatological time scales. Like for the model FVR data, the zonal mean has also been computed for the CFR data sets over the period 1401–2000. Results were similar to those for the entire period covered by the CFR data. The zonal distributions were very uniform, as depicted in figure 4.22.



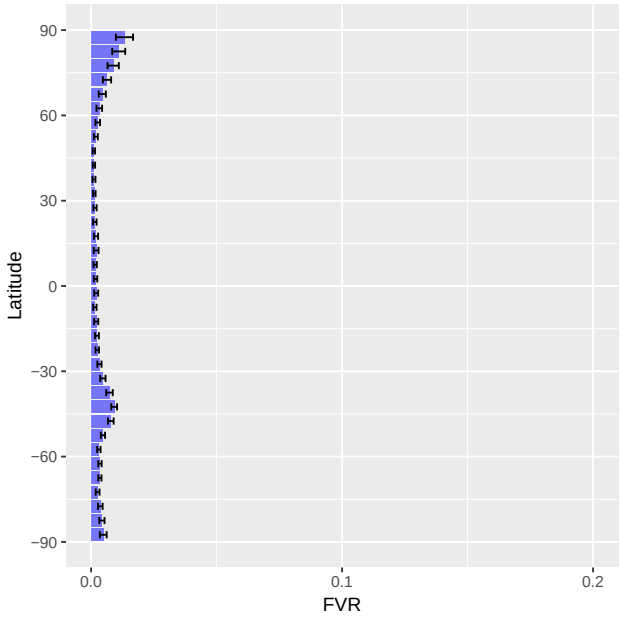
(a) PCR, Climatological Forcing Regressions



(b) Analogue, Annual Forcing Regressions



(c) CCA, Climatological MEM Regressions



(d) CPS, Annual MEM Regressions

Figure 4.22: *Zonal-Mean FVR from the CFR data sets, for selected regression methods and time scales. The blue bars denote the ensemble-mean, zonal-mean FVR, the whiskers denote the ensemble mean plus/minus one standard deviation of ensemble members. Note the different scale on subfigures (a) and (b) as compared to (c) and (d).*

4.5.4 Interpretation

The comparison of the three kinds of data sets (instrumental and reanalysis, climate model and CFR) revealed little agreement over the whole time of data availability. However, within the three groups, some consistent features could be observed. Also, this low agreement could be due to the different time spans that the data sets cover. Indeed, when comparing all the data sets over their common period (1851–2000), good agreement could be found for each regression method and time scale. Both the climate model and CFR data sets showed similarities to the instrumental and reanalysis data for the respective regression method. Moreover, for the FVR data from annual regressions, all three regression methods agreed in a local maximum of FVR in the Arctic. On the climatological time scale, agreement was less clear. Also for the period 1401–1851, little agreement could be found between climate model and CFR data sets. A possible reason for that could be that substantial amounts of spatial information is lost in the CFRs. Another reason could be a seasonal bias of the CFRs towards the northern hemispheric growing season, due to the numerous contributing dendrochronological records.

4.6 Spatial Structure of the FVR: Temporal Development

This section aims at giving a description of the temporal development of the spatial structure of FVR. This is a challenging enterprise, on the one hand due to the wealth of the data. On the other hand, it is difficult to depict the temporal development of 2D fields in 2D plots. For these reasons, PCA has been chosen as a method to reduce data amounts and to make visible the dominant underlying patterns of FVR and their temporal evolution.

The results of a PCA on FVR fields of two example data sets are shown, one model data set and one CFR data set. It is not the aim to give an exhaustive description of all the data generated in this thesis. Instead, representative examples are shown that visualize the overall properties of those data. In doing so, the FVR fields from the instrumental and reanalysis data sets have been omitted, since for them only a single map of FVR has been calculated for every regression method, without temporal development. The section is organized into one subsection that describes the PCA results of a FVR data set from a climate model and one that does the same for a CFR. A final subsection will contain the interpretation of the results described.

In this section and all subsequent ones, the PCA results shown were obtained from PCA on FVR fields using the covariance matrix of the data. On a sub-sample of the data, PCA calculations were also carried out on the correlation matrix of the data. Results were found to be qualitatively similar to the ones that used the covariance matrix.

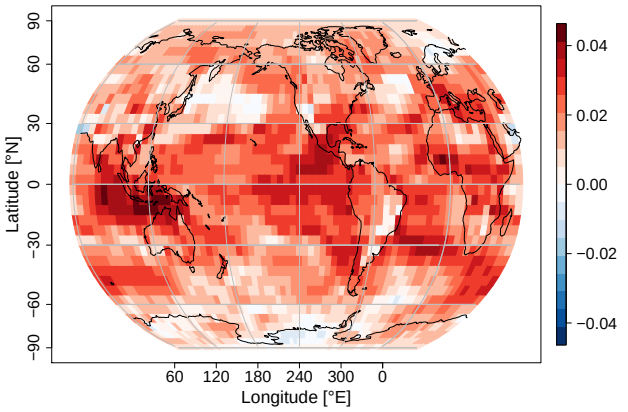
4.6.1 Climate Model Data

Figure 4.23 shows the outcomes of a PCA carried out on FVR fields generated from CESM run 1 by climatological MEM regressions. The loading field of PC1 (34% explained variance) is of uniform sign for virtually all of the grid cells. This implies that PC1 represents a globally simultaneous all ups-and-downs movement of FVR. Furthermore, only a few regions are neutral, namely parts of the Southern Ocean and Antarctica, and parts of the northern Atlantic and Pacific oceans.

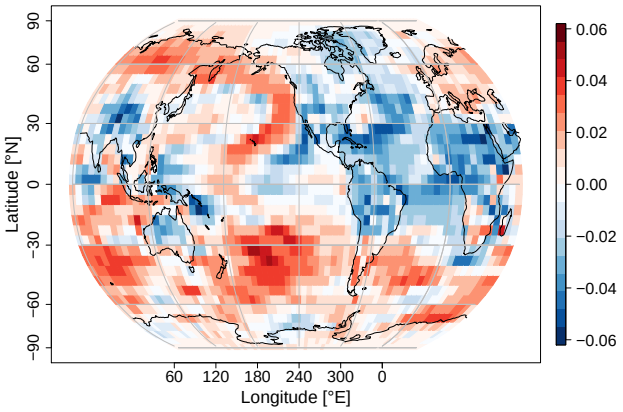
PC2 (11%), on the other hand, displays a loading field of both positive and negative loadings, with a complex spatial pattern. Most of Africa, the Atlantic and the Americas show negative loadings, while the southern Pacific ocean, the Indian ocean, Siberia and parts of the northern Pacific show positive loadings. The loading field patterns of PCs 3 (10%) and 4 (7%) are even more complex, those shall not be described in detail here.

The score time series plot in subfigure 4.23e reveals a complex development of the first 4 PCs in time. The two most dominant features of the PC1 score time series are a sharp rise in the 12^{th}, 13^{th} and 14^{th} centuries and a substantial increase of about the same magnitude from the 17^{th} to the 20^{th} centuries. The other PC score time series exhibit substantial variability as well.

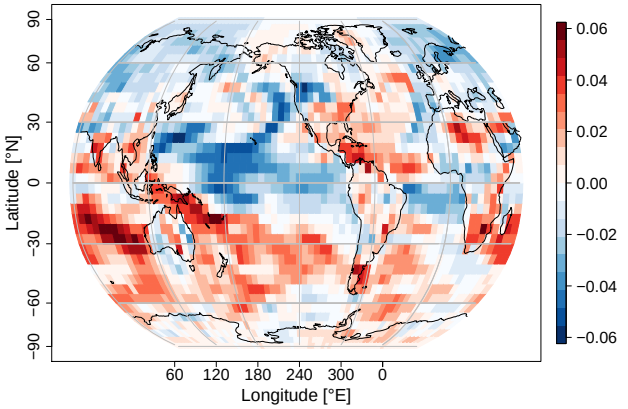
Subfigure 4.23f shows the scree plot of the PCA shown. PC1 contains about a third of the variance in the data set, while the other PCs contain substantially less.



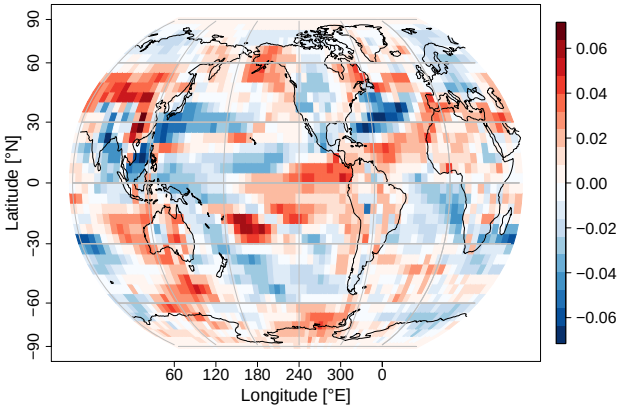
(a) PC1 Loading Field



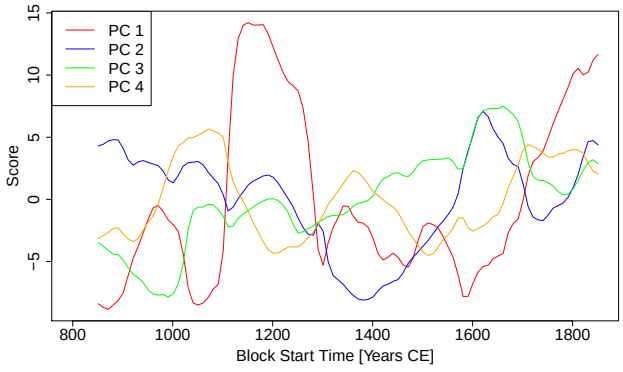
(b) PC2 Loading Field



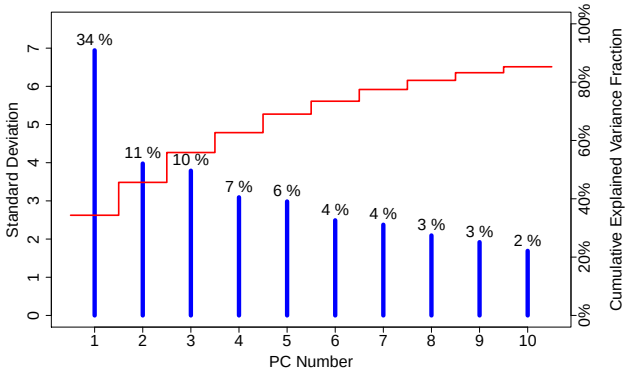
(c) PC3 Loading Field



(d) PC4 Loading Field



(e) PC 1-4 Score Time Series



(f) Scree Plot

Figure 4.23: *Output of a PCA of FVR fields from CESM run 1 (climatological MEM regressions). (a) to (d) loading fields of PCs 1-4, (e) score time series of PCs 1-4, (f) scree plot of the PCA. Blue bars show the standard deviation of the respective PC (left scale). The red line shows the cumulative explained variance fraction (right scale). Truncation after PC4 is arbitrary.*

4.6.2 Climate Field Reconstruction Data

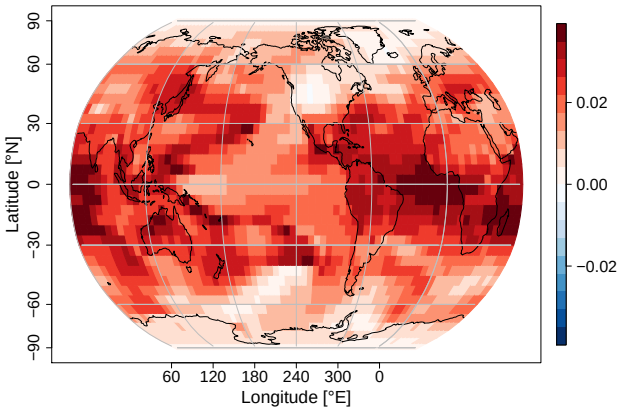
The outcomes of a PCA carried out on ensemble mean FVR fields from the DA CFR (annual forcing regressions) are shown in figure 4.24. Like in figure 4.23, the first PC loading field is of globally uniform sign, indicating an all ups-and-downs development of the FVR as the major mode of variability.

The loading field of PC2 is more complex, yet some patterns can be observed: In the central and eastern tropical Pacific ocean, there is a pronounced region of negative loadings. Wrapped around that, in the western subtropical Pacific of both hemispheres, there are regions with positive loadings. However, they are much more pronounced in the northern hemisphere and over Indonesia. Also in the North Atlantic, there is a pronounced pattern. Namely the north-western North Atlantic displays positive loadings, while the south-eastern and central North Atlantic exhibits negative loadings. Combining the observations from the Atlantic and the Pacific oceans, one can observe a wedge-shaped feature of negative loadings with its base on the African continent and its point in the western pacific ocean near Indonesia.

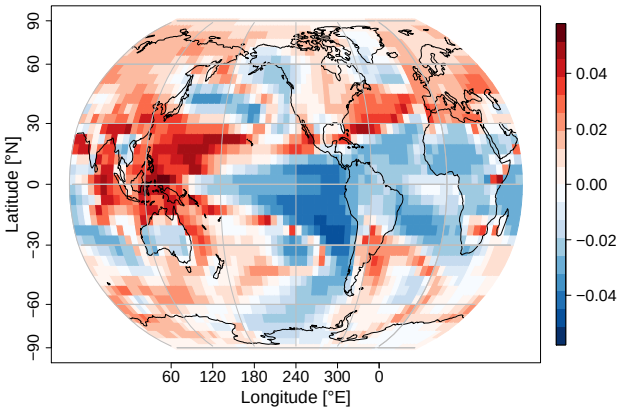
The loading field of PC3 seems to broadly reflect the wedge-shaped feature described before. However, here only the rough structure remains and the distribution of positive and negative loadings is more complex. A similar observation can be made in the loading field of PC4, with the addition that here, large areas have loadings close to zero.

The score time series of the PCs described before are shown in figure 4.24e. For PC1, the most dominant feature is clearly the sharp increase from the 17^{th} to the 20^{th} centuries. Also in the 13^{th}, 14^{th} and 15^{th} centuries, a significantly higher than average score can be observed. The score of PC2 drops at the same time, while otherwise closely following the development of PC1. For PCs 3 and 4, it is much harder to observe consistent features.

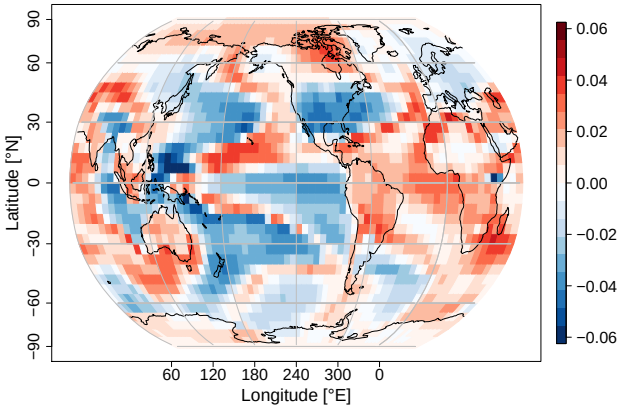
Finally, figure 4.24f displays the scree plot of the PCA shown. It is evident that PC1 contains almost half of the variance in the data set, while the higher-order PCs each contain much less.



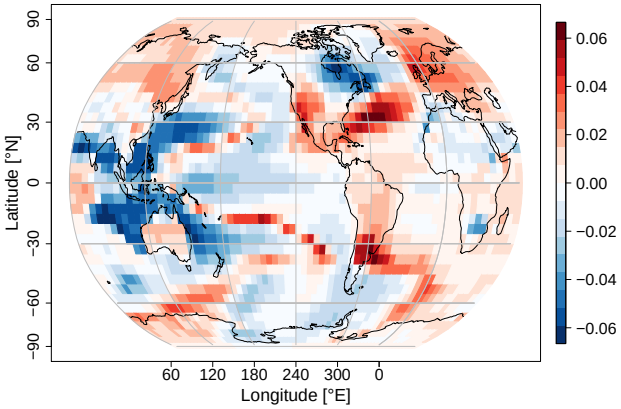
(a) PC1 Loading Field



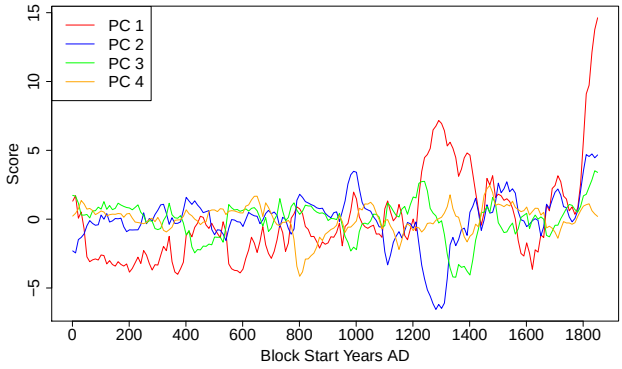
(b) PC2 Loading Field



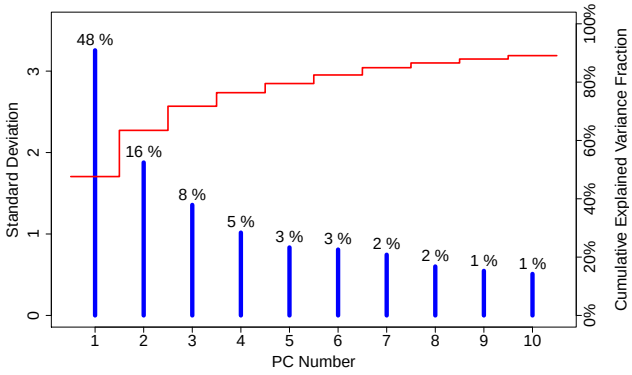
(c) PC3 Loading Field



(d) PC4 Loading Field



(e) PC 1-4 Score Time Series



(f) Scree Plot

Figure 4.24: *Output of a PCA of ensemble-mean FVR fields from the DA CFR (regression with forcing time series on the annual time scale). (a) to (d) loading fields of PCs 1-4, (e) score time series of PCs 1-4, (f) scree plot of the PCA. Blue bars show the standard deviation of the respective PC (left scale). The red line shows the cumulative explained variance fraction (right scale). Truncation after PC4 is arbitrary.*

4.6.3 Interpretation

In both data sets shown, the PCA revealed a loading field pattern of globally uniform sign in PC1. This means that the largest fraction of variability in FVR can be explained by a globally simultaneous all ups-and-downs behaviour of FVR. In addition, the score time series of the first PC showed a marked increase in the 19th and 20th centuries. This suggests a relationship of PC1 with GHG forcing. Moreover, since the PC1 loading field is of globally uniform sign, the response of surface temperature to forcing is globally uniform on a first order.

The strongly enhanced values of the PC1 score time series in the 12th, 13th and 14th centuries shown in figure 4.23e coincide precisely with the year 1257 entering or leaving the respective moving windows. The quality of this temporal match as well as the fact that this could be observed in both FVR data of the CESM (section 4.1.2) and the PCA results thereof lend enhanced confidence to the result that this feature is related to the eruption of Samalas volcano. However, it is questionable why such a pronounced response is only apparent for the Samalas eruption and not other major volcanic eruptions.

Furthermore, in the PCA results from the DA CFR data set, and more precisely in the loading field of PC2, two characteristic features have been discovered: The first are pronounced negative values of the loadings in the central and eastern tropical Pacific. This region suggests a relationship to ENSO. While not giving any indication that ENSO is forced or not, this instead means that whenever FVR is high in the ENSO region, it is low in the regions displaying positive loadings, and vice versa. The same is true for the region of strongly positive loadings in the western North Atlantic. Here, a relation to the Gulf Stream seems likely. The same considerations as for the ENSO region apply.

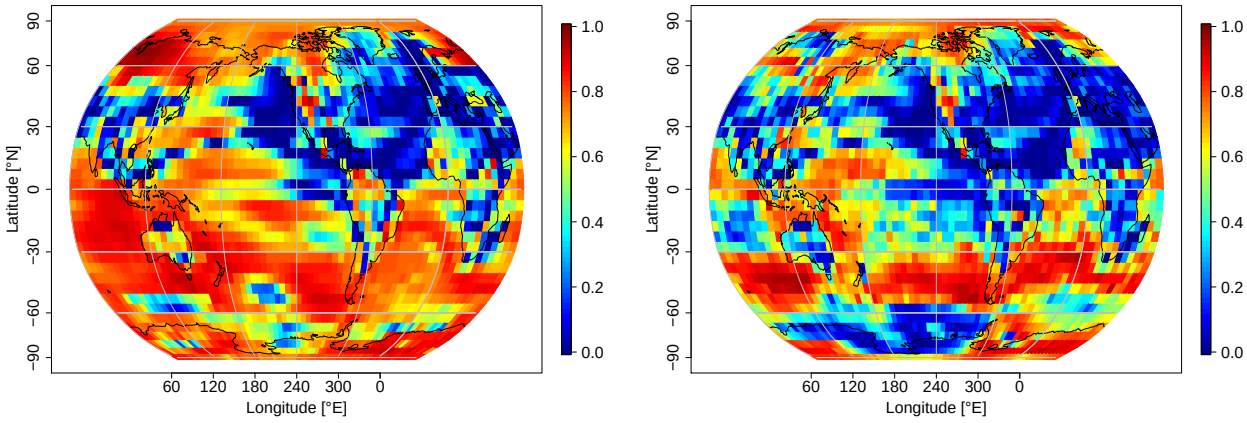
Overall, it can be stated that the temporal development of the spatial structure of FVR is very complex. Nevertheless, some interesting features could be extracted by means of a PCA.

4.7 Spatial Structure of the FVR: Similarities between Data Sets

This section will give examples of agreement of data sets, either in FVR directly or in the PCA results thereof. The first subsection (4.7.1) is dedicated to FVR data, while the second (4.7.2) covers data from PCA of FVR fields. As in section 4.6, the data shown are not intended to be comprehensive, but rather to give examples to illustrate some key properties. It needs to be stated also, that overall the differences in spatial structure of FVR between the data sets (see section 4.8) overweigh the similarities.

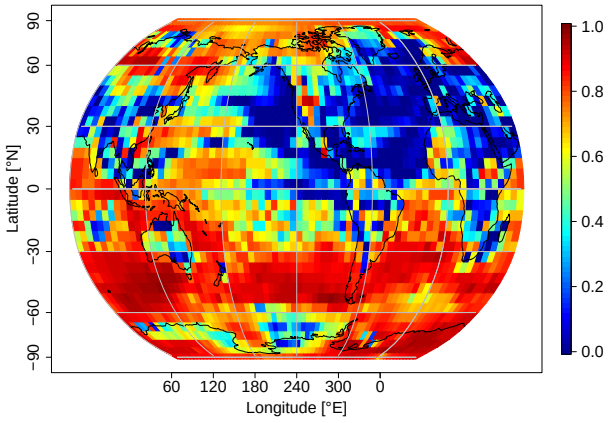
4.7.1 Forced Variance Ratio

Figure 4.25 shows the ensemble mean FVR maps of the six CFR data sets from climatological MEM regressions and for the last moving window (1851–2000). Some consistent patterns can be observed: for Analogue, CPS and DA, there are high values of FVR in the southern hemisphere, over the Indian Ocean, in Siberia and the Arctic. The other CFR data sets (CCA, GraphEM and PCR) agree in high values of FVR in the southern hemispheric mid-latitudes, the western tropical Pacific and the Arctic. Moreover, all data sets shown agree in low values of FVR the eastern North Pacific, parts of North America, the North Atlantic and Europe. Some subordinate features show disagreement between the data sets, such as the magnitude of FVR over the Indian ocean. The agreement described before could not be found in the FVR data from the climate models and the instrumental and reanalysis data sets.

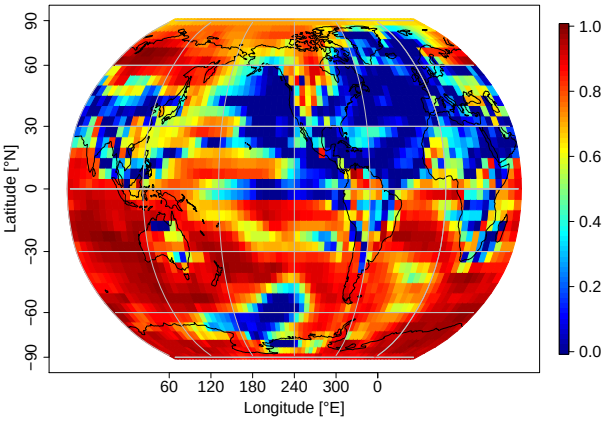


(a) Analogue

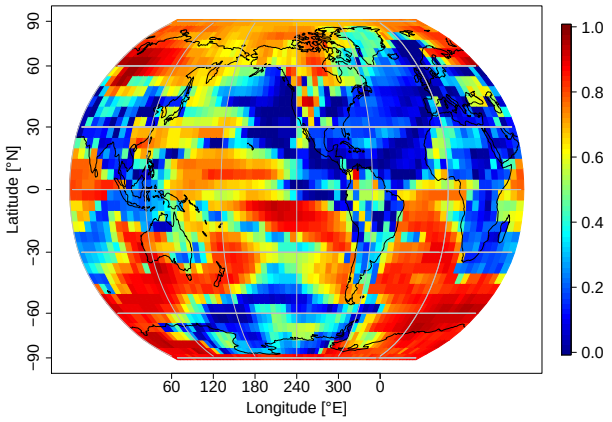
(b) CCA



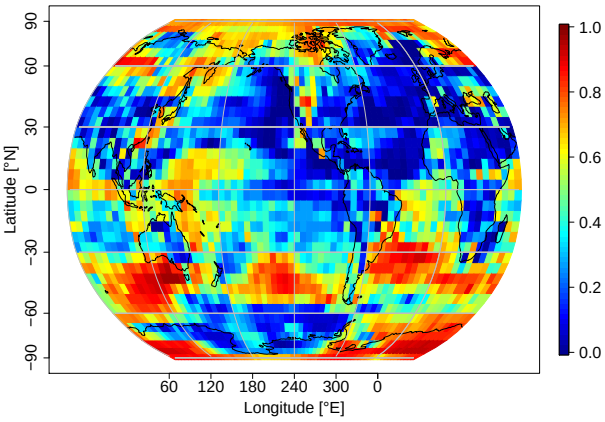
(c) CPS



(d) DA



(e) GraphEM



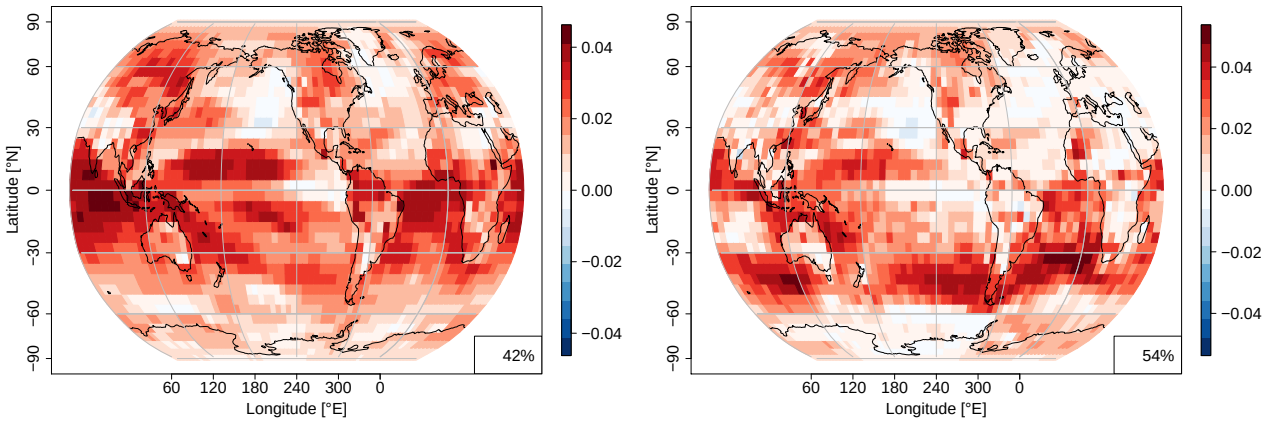
(f) PCR

Figure 4.25: *Ensemble-mean FVR from climatological MEM regressions for the six CFR data sets and the last moving window (1851–2000).*

4.7.2 Principal Component Analysis

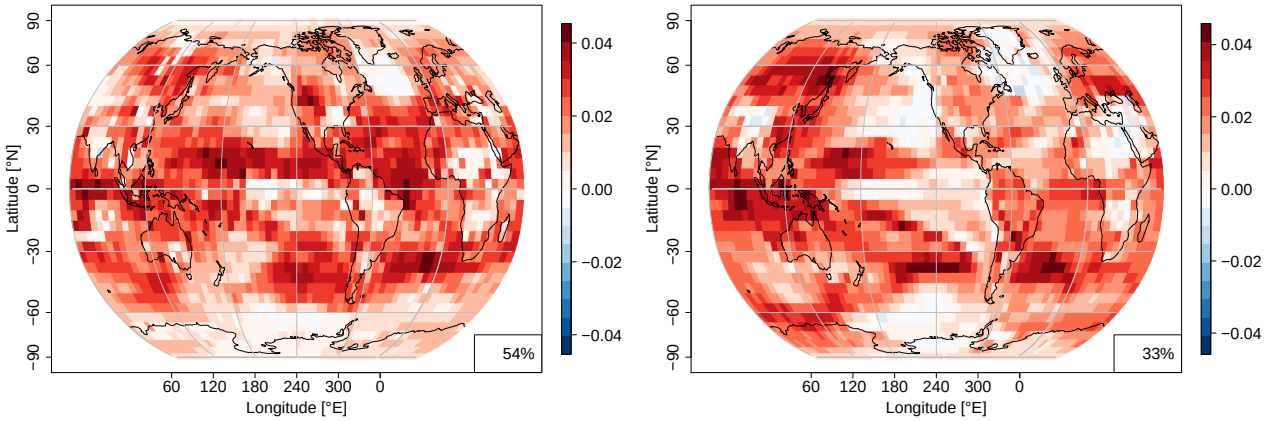
Figures 4.26 and 4.27 show the PC1 loading fields of a PCA carried out on FVR fields produced from climatological MEM regressions. In addition, figure 4.27 also shows the corresponding PC score time series. It is evident that for each data set shown here, the PC1 loading field is of globally uniform sign. This feature could be observed to be consistent in the FVR data from all three regression methods, both time scales (annual and climatological) and the ensemble mean of all data sets. It also holds true for the large majority of all ensemble members of the respective data sets (results not shown). The only notable exception thereof here is the HadCM3, whose PC1 loading field shows some slightly negative loadings over the eastern tropical Pacific, the south-eastern North Atlantic and over Africa.

The PC score time series depicted in figure 4.27c also show quite close agreement between the data sets. For example, simultaneous high values can be observed in all data sets between moving window start times 1301 CE to about 1581 CE. Thereafter, the values are substantially lower until about moving window start time 1771, and rise sharply after that. This sharp rise could, consistently with the results from the PC loading fields, also be found in the FVR data from all three regression methods, both time scales and most ensemble members of all data sets. The most apparent disagreement between data sets shown here are the high values of CESM between moving window start times 1101 and 1261, something not reflected in any of the other data sets. Finally, it can be noted that all the first PCs contain high fractions of the total variance in their respective data set, with the DA CFR being the lowest (33%).



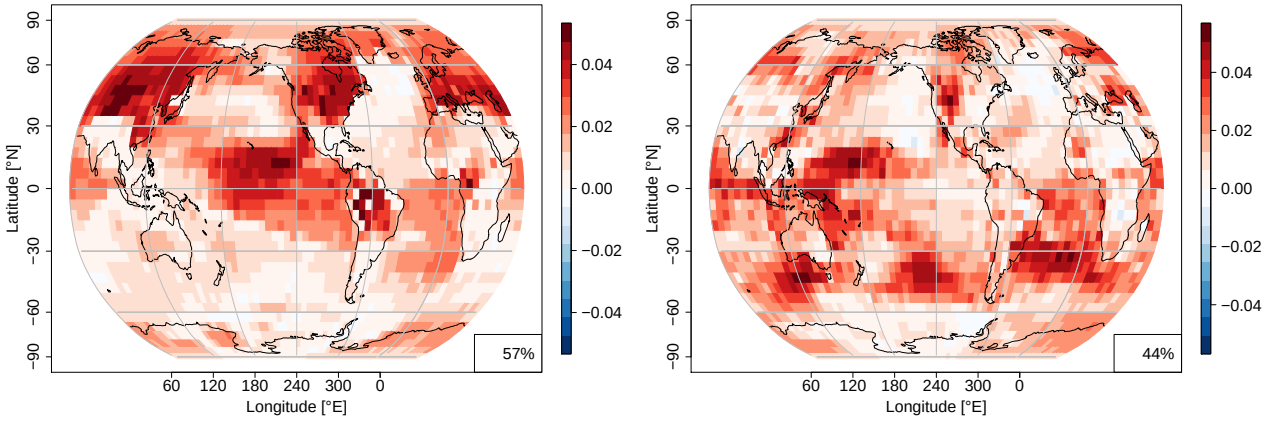
(a) Analogue

(b) CCA



(c) CPS

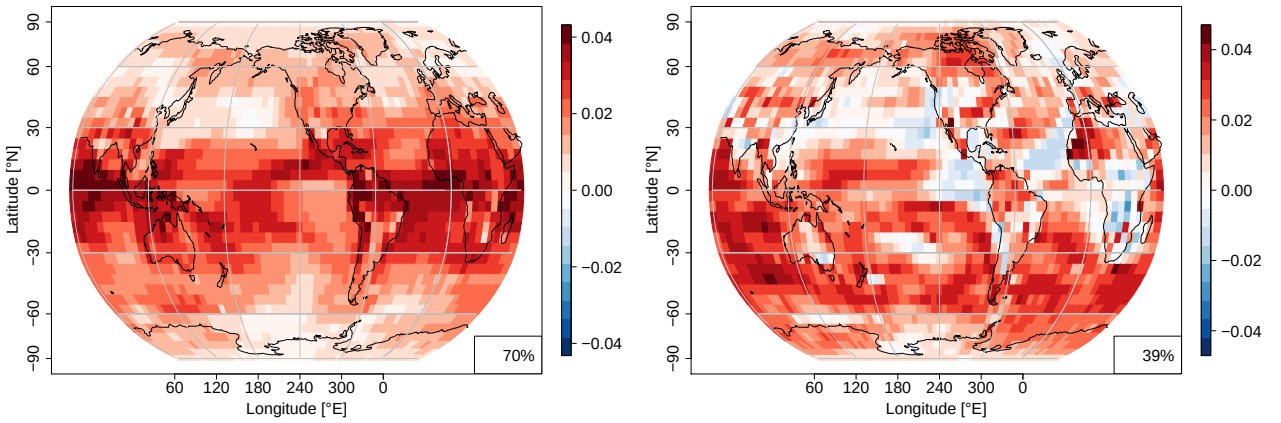
(d) DA



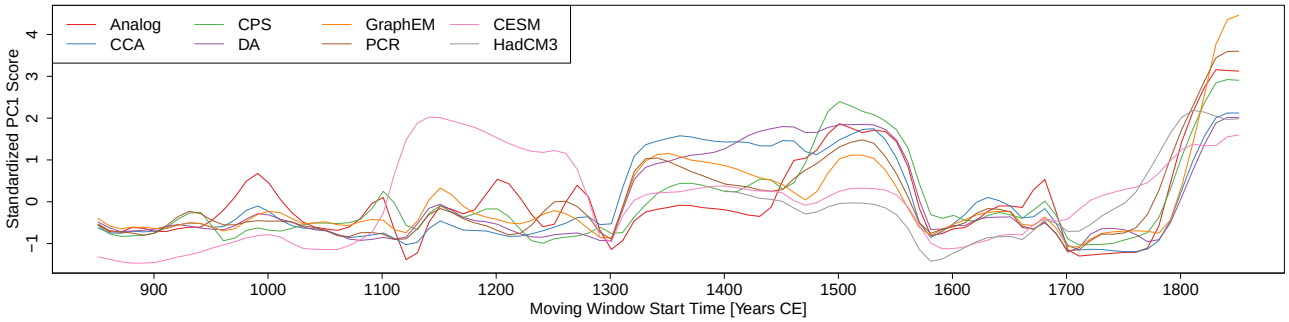
(e) GraphEM

(f) PCR

Figure 4.26: *(a-f) PC1 loading fields of a PCA carried out on ensemble-mean FVR data from climatological MEM regressions. Shown here are the PC1 loading fields from the six CFRs, including an indication of the fraction of explained variance of each PC1.*



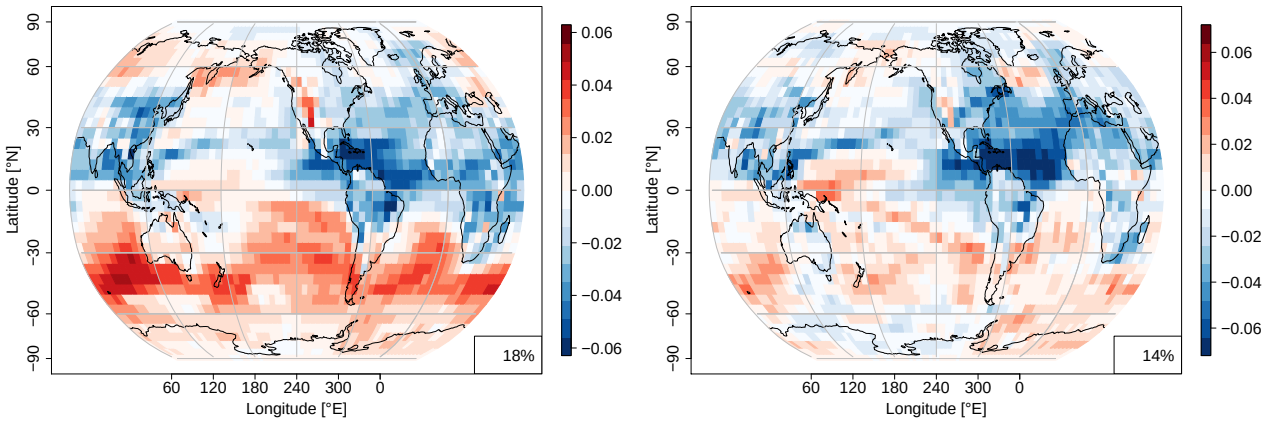
(a) CESM *(b) HadCM3*



(c) Standardized PC1 score time series of the shown data sets

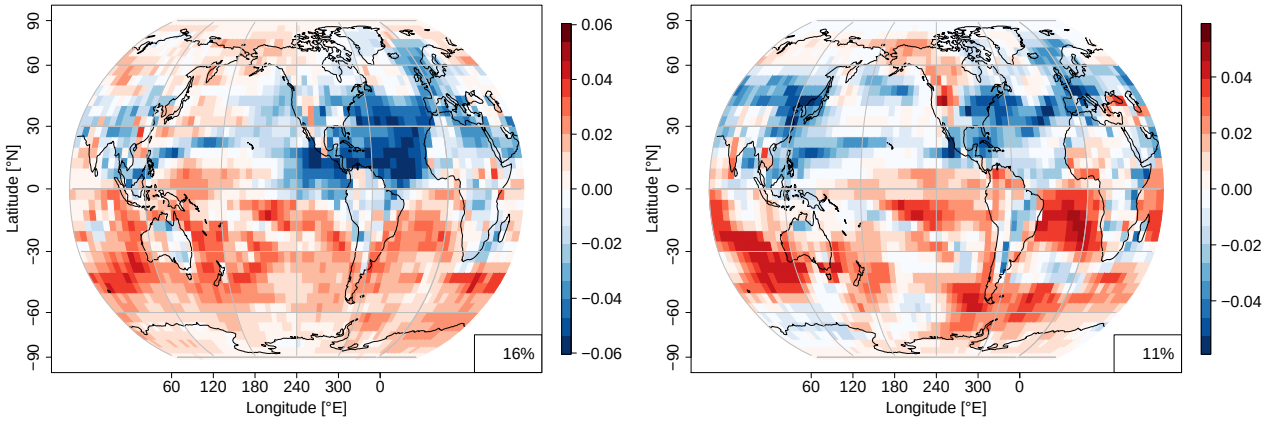
Figure 4.27: *(a-b) The same as in figure 4.26, but for the two climate model data sets, (c) standardized PC1 score time series of the data sets shown in figure 4.26 and figure 4.27. All data shown are from climatological MEM regressions. To enhance comparability, the sign of the first eigenvector of each data set has been chosen so that the mean of each loading field is positive. The sign of the PC1 score time series has also been switched accordingly. In addition, the score time series shown have been standardized to zero mean and unit standard deviation to enhance comparability. In the lower right-hand corner of all the loading field maps, the fraction of explained variance of the respective PC1 is given.*

Figures 4.28 and 4.29 are similar to figures 4.26 and 4.27, but here the loading fields and scores of the second PCs are shown. Consistently in all data sets shown, the loadings in the southern hemisphere are mostly positive. Conversely, there is a zone covering the eastern tropical Pacific, the Caribbean and the North Atlantic, sometimes Africa, where the loadings are predominantly negative. Broadly speaking, this implies a dipole between the northern and southern hemispheres. This dipole seems to be present in all the data sets shown here, with some deviations from the pattern. However, this feature could not be consistently found in the FVR data from the other regression methods and time scales. Also the PC score time series show some agreement, with the temporal development being mostly flat during most of the time. Exceptions thereof are a slight dip in PC2 score around moving window start time 1501 and a pronounced rise after moving window start time around 1781.



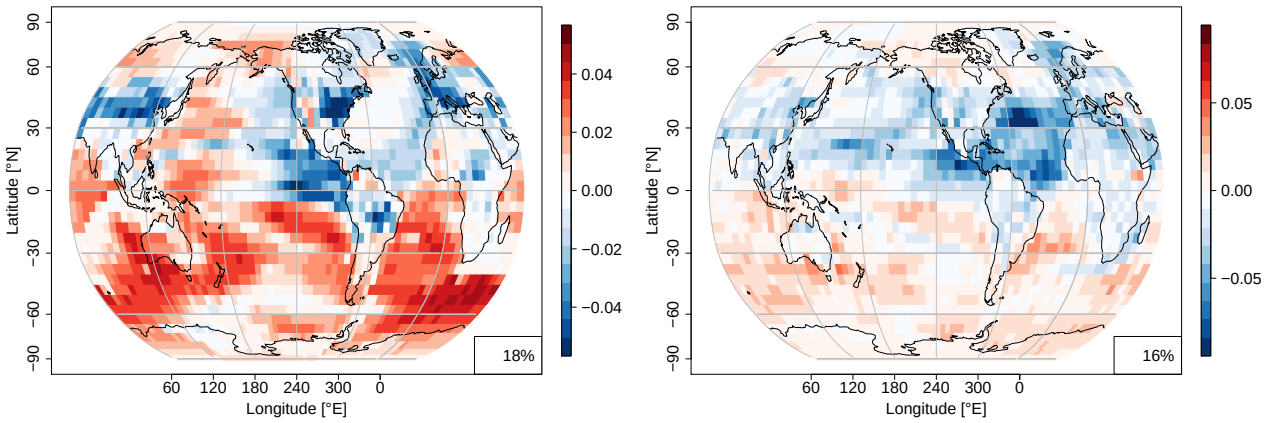
(a) Analogue

(b) CCA



(c) CPS

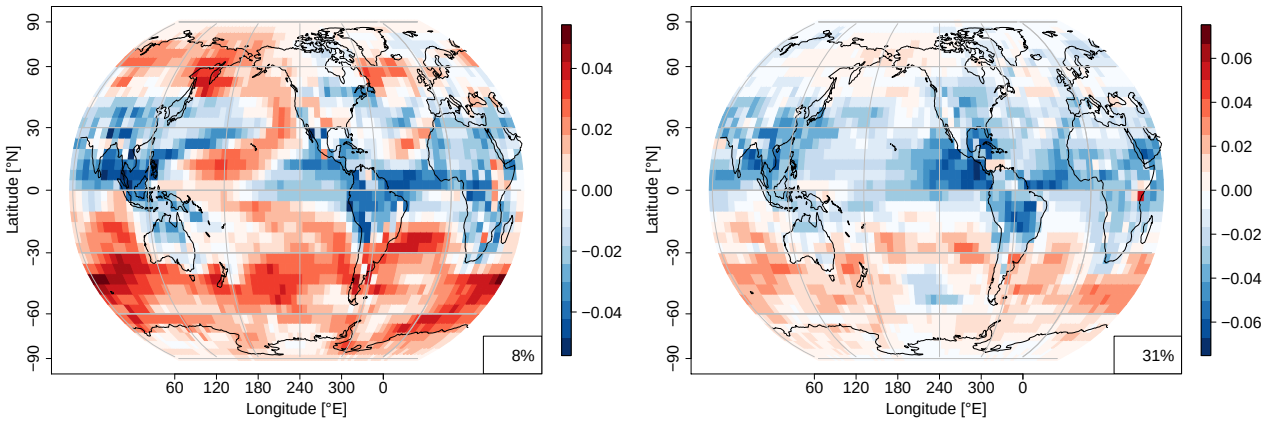
(d) DA



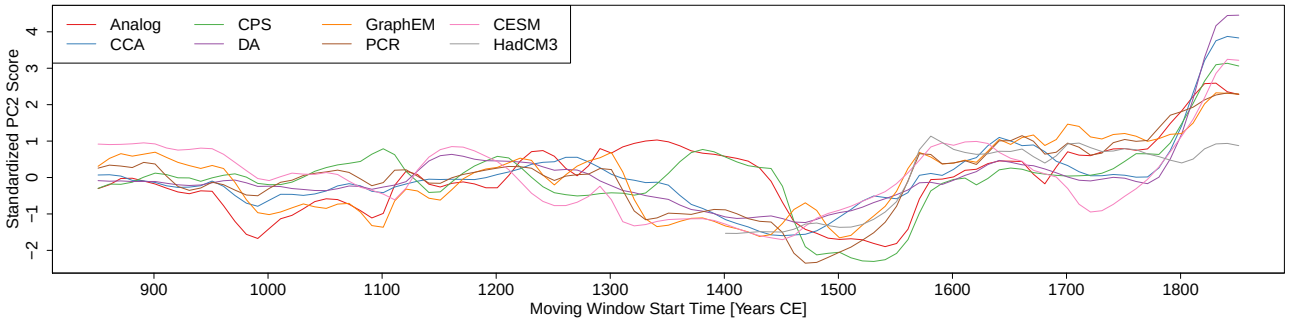
(e) GraphEM

(f) PCR

Figure 4.28: *(a-f) PC2 loading fields of a PCA carried out on ensemble-mean FVR data from climatological MEM regressions. Shown here are the PC2 loading fields from the six CFRs, including an indication of the fraction of explained variance for each PC2.*



(a) CESM *(b) HadCM3*



(c) Standardized PC2 score time series of the shown data sets

Figure 4.29: *(a-b) The same as in figure 4.28, but for the two model data sets, (c) standardized PC2 score time series of the data sets shown in figure 4.28 and figure 4.29. All data shown are from climatological MEM regressions. To enhance comparability, the sign of the first eigenvector of each data set has been chosen so that the last value of each non-standardized PC score time series is positive. The sign of the PC loading maps has also been switched accordingly. Additionally, the score time series shown have been standardized to zero mean and unit standard deviation to enhance comparability. In the lower right-hand corner of all the loading field maps, the fraction of explained variance of the respective PC2 is given.*

4.7.3 Interpretation

Some underlying commonality is suggested by the similarities presented in section 4.7.1, namely the low values of FVR in all the CFR data sets in the North Atlantic sector and adjacent areas. Bearing in mind that those FVR data were produced from climatological regressions with the MEM of the CESM, some link to the properties of CESM is likely. Moreover, the region indicated is the location of major industrial economic activity on the globe in the time span of interest (the last moving window, 1851 – 2000). A possible explanation for these observations is that the CESM underestimates 20$^{\text{th}}$ century warming in the northern hemisphere due to too strong aerosol forcing (Otto-Bliesner et al. 2016). Accordingly, a difference was introduced between the ensemble mean of CESM and the CFRs, implying lower agreement between the data sets mentioned and consequently, lower regression R^2 and FVR.

The PC1 loading field of globally uniform sign presented in figures 4.26 and 4.27 implies that FVR varies in a globally simultaneous all ups-and-downs pattern on a first order. Moreover, this feature was observed to be exceedingly stable with regard to time scale, regression method and data set, implying general validity of this result. The same is true for the properties of the PC1 score time series. Moreover, their sharp rise in the 18$^{\text{th}}$, 19$^{\text{th}}$ and 20$^{\text{th}}$ centuries likely reflects the increase of FVR in response to anthropogenic industrial GHG emissions.

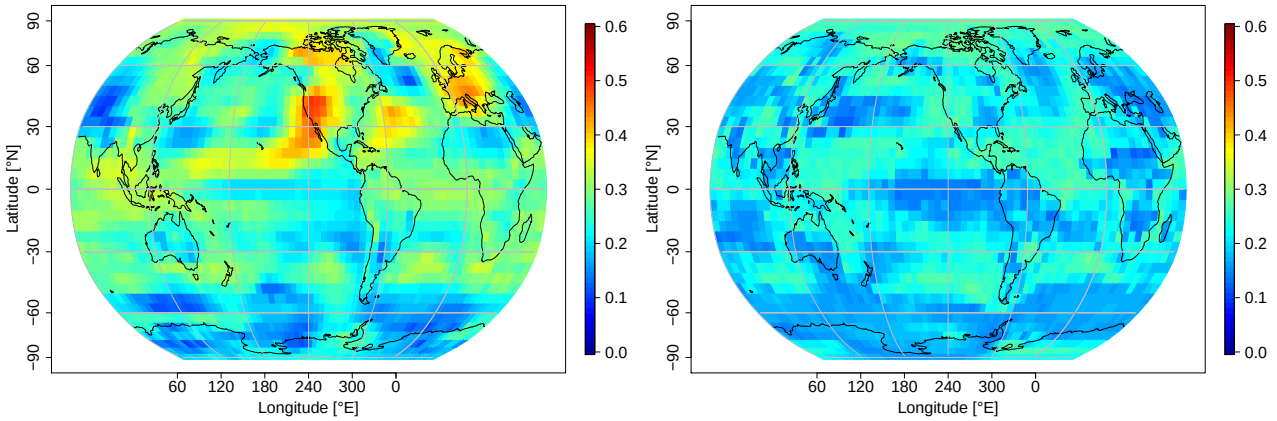
The dipole found between the two hemispheres in the loading fields of PC2 could not be observed consistently in all the FVR data. Therefore, more caution is appropriate with regard to this result. However, if GHG and solar forcing are regarded as globally uniform, a relation to volcanic forcing could be indicated here. An argument that speaks against this is the shape of the PC2 score time series, which seem not to show any relationship to volcanic forcing. Instead, as for PC1, they seem to have a relationship to the industrial rise in GHG. This would suggest that in the industrial period, FVR has increased in the southern hemisphere, while it decreased in the northern hemisphere. Again, a relationship to the underestimation of the anthropogenic warming in the CESM in the northern hemisphere is likely (Otto-Bliesner et al. 2016), bearing in mind that the FVR data underlying the PCA here come from regressions with the CESM MEM.

4.8 Spatial Structure of the FVR: Differences between Data Sets

This section will give examples of differences of the data sets in the spatial structure of FVR. To this end, maps of FVR are presented (subsection 4.8.1), as well as results from a PCA on FVR fields (subsection 4.8.2). Again, the data presented are not intended to be comprehensive, but rather to give illustrative examples of the key properties found in the entirety of the data produced in this thesis.

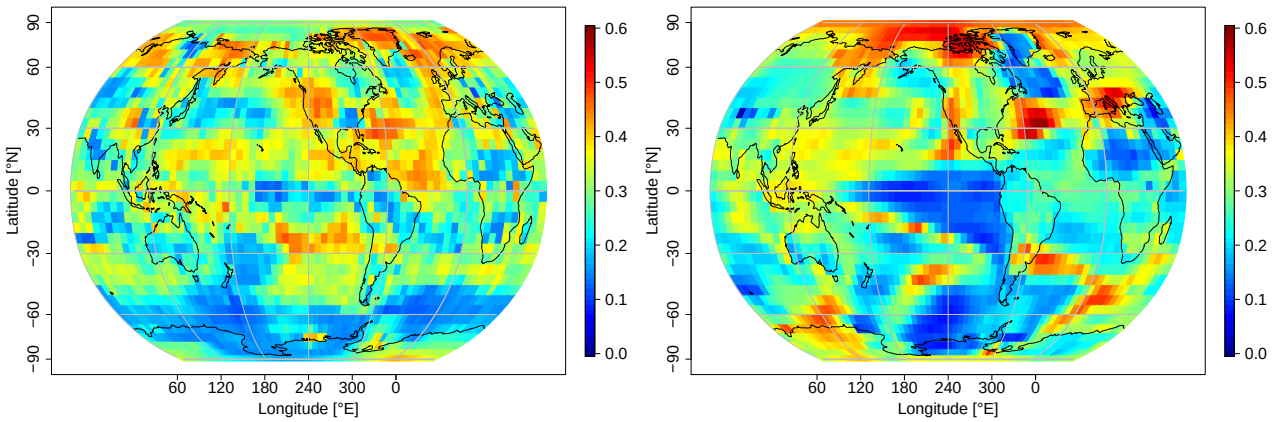
4.8.1 Forced Variance Ratio

Figures 4.30 and 4.31 show the ensemble-mean FVR of the six CFRs and the two climate models. More precisely, the mean of the moving windows with start times from 1401 to 1851 is shown and the data come from climatological GM-MEM regressions. The substantially different spatial structure of FVR among the data sets is evident. As already mentioned in the introduction of section 4.7, this is the prevalent case in most of the FVR data produced in this thesis. For the DA CFR, the wedge-shaped feature already described in subsection 4.6.2 can be observed again. The GraphEM CFR on the other hand displays a pattern reminiscent of the Pacific Decadal Oscillation (PDO), additionally with high values of FVR in the Arctic and North America as well as in Antarctica. The other four CFR data sets show much less clear patterns of FVR, and CPS in particular seems quite noisy. In the two model data sets (figure 4.31), there seems to be some emphasis on the tropics, with high FVR values there. An exception thereof is the tropical Pacific region in CESM. Additionally, in HadCM3, the high FVR values in the tropics seem to extend further north and south over the land masses than they do over the oceans. This is particularly evident in South America and Africa. It needs to be stated, however, that the spatial patterns of FVR from the CFRs likely have many spatial artifacts. These occur because of the spatial distribution of the proxy records, which leave large areas uncovered.



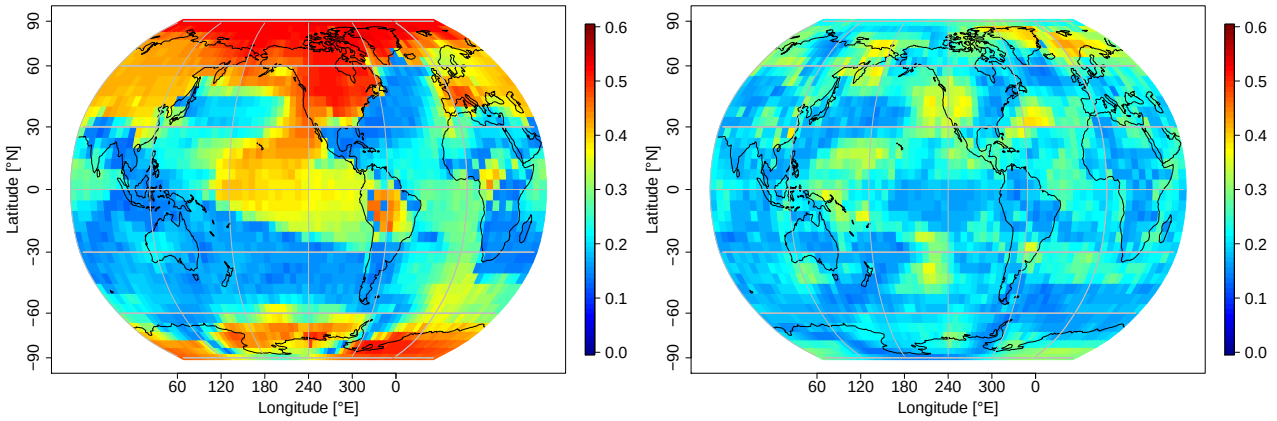
(a) Analogue

(b) CCA



(c) CPS

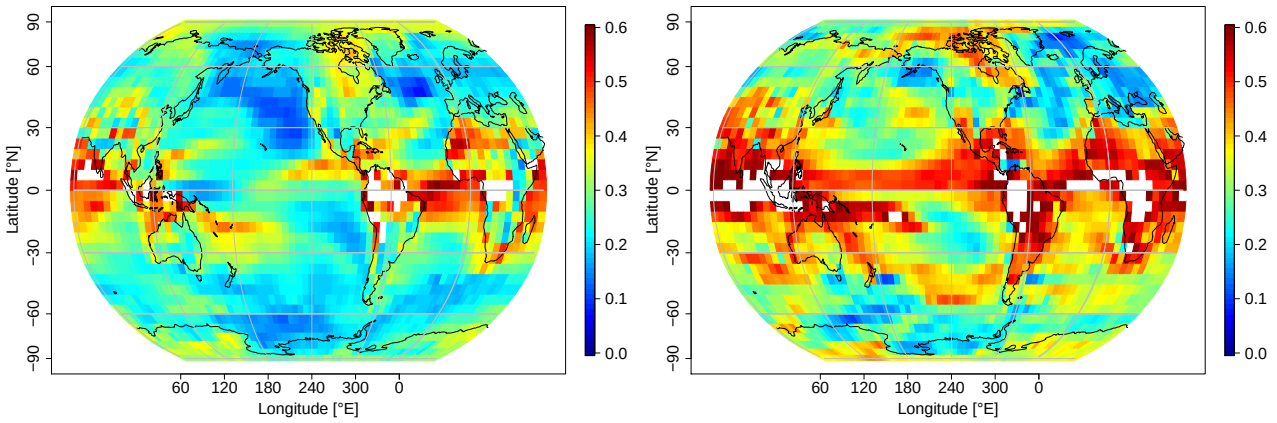
(d) DA



(e) GraphEM

(f) PCR

Figure 4.30: *Ensemble-mean FVR from climatological GM-MEM regressions of the six CFR data sets. Also, the maps shown display the mean of all moving windows with start times from 1401 to 1851.*

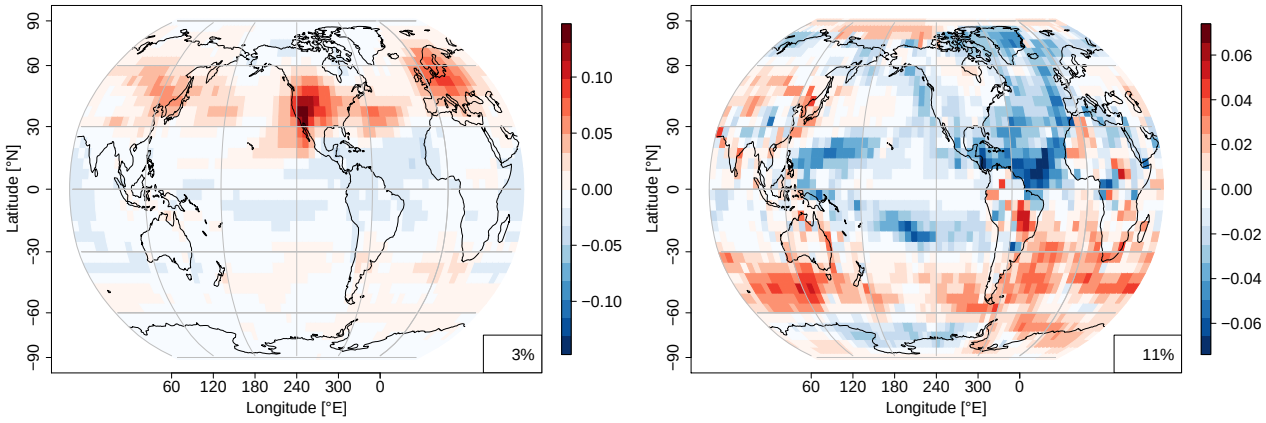


(a) CESM *(b) HadCM3*

Figure 4.31: *The same as in figure 4.30, but for the two climate model data sets.*

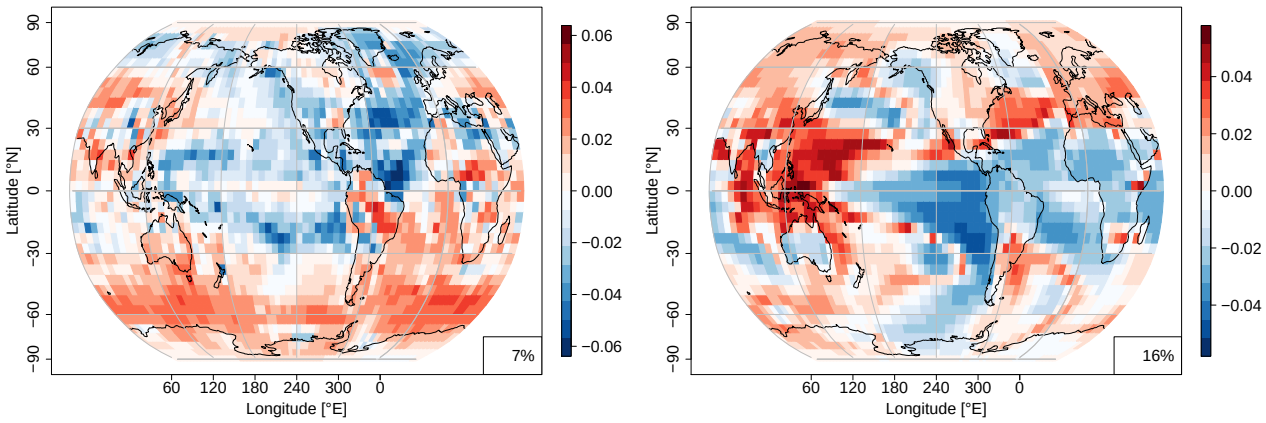
4.8.2 Principal Component Analysis

Figures 4.32 and 4.33 show the results of a PCA carried out on FVR data from annual forcing regressions. While figure 4.32 displays the PC2 loading fields for the six CFR data sets, figure 4.33 does so for the two climate model data sets. Figure 4.33c additionally gives the corresponding PC2 score time series. Also here, as in subsection 4.8.1, the differences between the data sets are evident. There seems to be a characteristic spatial pattern of PC2 loading for every data set. Also the PC2 score time series differ substantially with data set. However, this applies only to the second PC, since for the first, unambiguous similarities have been demonstrated (see section 4.7.2). Still, one similarity between the score time series can be observed in a substantial rise after about moving window start time 1801.



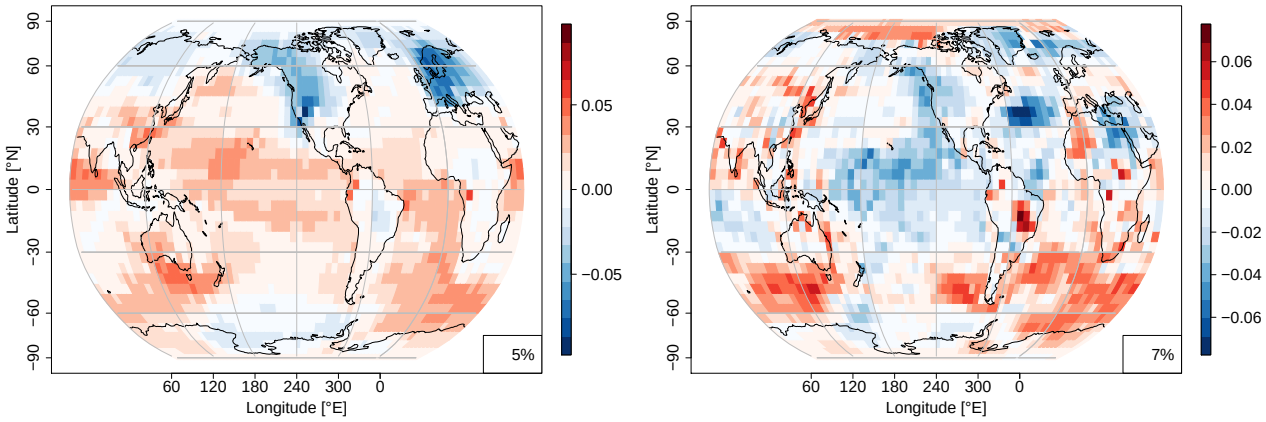
(a) Analogue

(b) CCA



(c) CPS

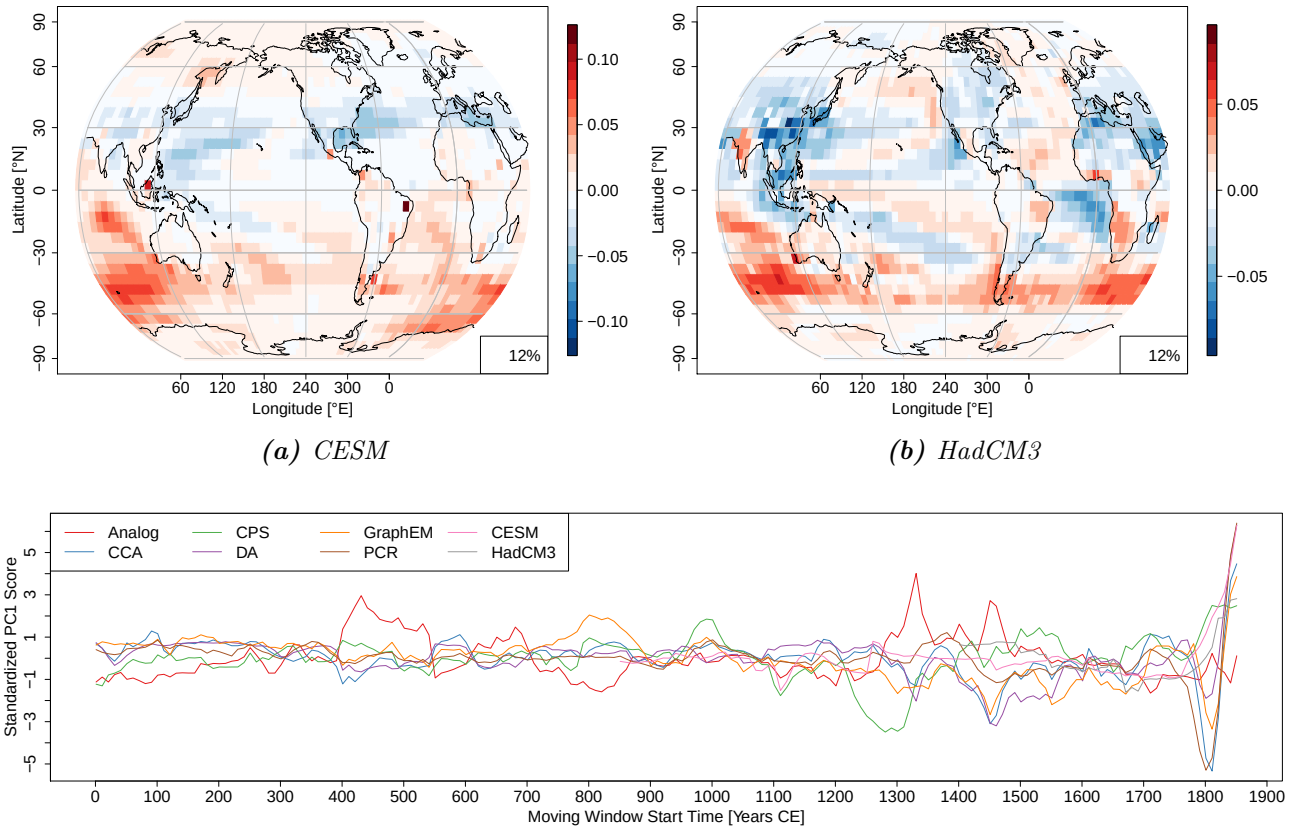
(d) DA



(e) GraphEM

(f) PCR

Figure 4.32: *Results of a PCA carried out on ensemble-mean FVR fields produced from annual forcing regressions. (a-f) PC2 loading fields of the six CFR data sets. The fraction of explained variance of each respective PC is given in the lower right-hand corner of each map.*



(a) CESM

(b) HadCM3

(c) Standardized PC2 score time series of the shown data sets

Figure 4.33: *(a-b) The same as in figure 4.32, but here for the two climate model data sets. (c) Standardized PC2 score time series belonging to the loading fields shown in figures 4.32 and 4.33. The sign of the eigenvectors of the PC2 has been chosen so that the last value of the non-standardized PC2 score time series is positive, to enhance comparability. The signs of the PC loading maps has also been switched accordingly. In addition, the score time series have been standardized to zero mean and unit standard deviation.*

4.8.3 Interpretation

The FVR maps and PCA results shown in this section suggest substantial differences between the data sets. In fact, this is the case for most FVR data produced in this thesis. However, each data set seems to have its own characteristic spatial spatial structure of FVR, respectively the PCA loading maps thereof. For the climate model data sets, figure 4.31 gave some indication that FVR might be higher in the tropics, consistent with the results from section 4.5.2 and those of Lyu et al. 2015. This is in keeping with basic climatological reasoning, the climate being much more stable in the tropics, with persistent hot temperatures and high precipitation. Consequently, the effect of a forcing influence should be much better detectable in the tropics as opposed to the extratropics, where a signal from forcing is likely to get undetectable in the presence of substantial noise. Thus, one would expect the FVR to be much higher in the tropics. Furthermore, the fractions of explained variance are similar between PC2 and the higher-order PCs (all low). Therefore, the distinction between PCs 2 and 3 is likely not significant for many data sets. This explains that the spatiotemporal patterns as compared between the data sets are relatively random.

4.9 Spatial Structure of the FVR: Similarities between Regression Methods

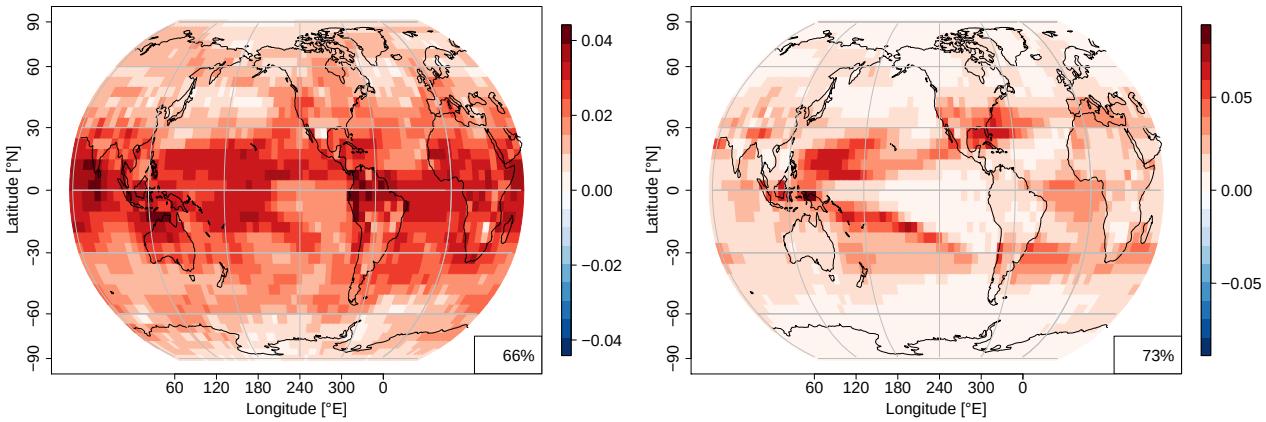
This section will give examples of similarities that have been found between the spatial structure of FVR from the six different regression methods (forcing, MEM and GM-MEM, each on the annual and on the climatological time scale). Here, only FVR data from climate models and from CFR are examined, the ones from instrumental and reanalysis data sets shall be considered in section 4.10.1. Section 4.9.2 will give an example of similarity between regression methods in climate model FVR data, while section 4.9.3 will do so for FVR data from the GraphEM CFR. Finally, section 4.9.4 will give an interpretation of the data shown.

4.9.1 Instrumental and Reanalysis Data

For the spatial structure of FVR from the instrumental and reanalysis data, little similarities could be found. For this reason, and to unite the corresponding plots in the same place for overview, these data are shown in subsection 4.10.1.

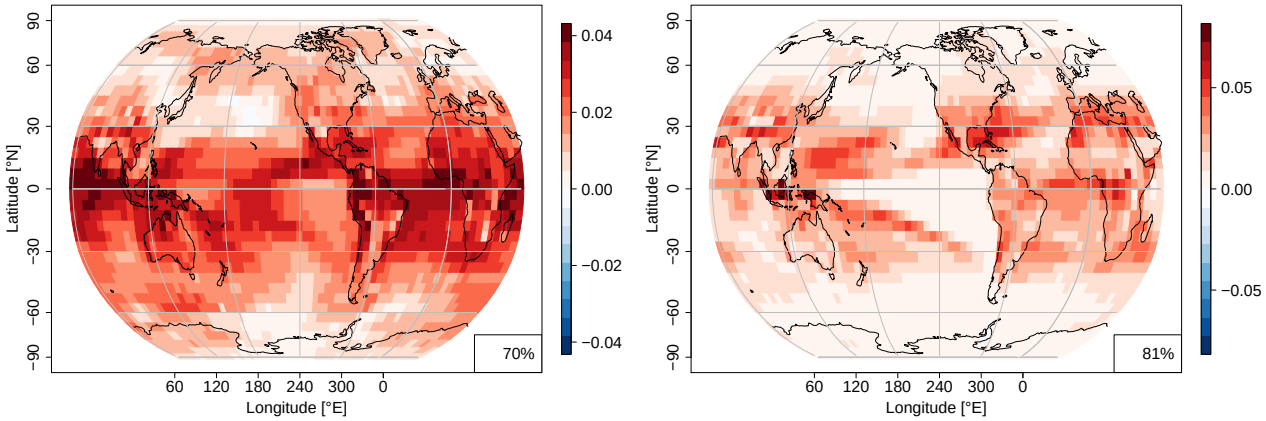
4.9.2 Climate Model Data

Figure 4.34 shows the PC1 loading fields from a PCA carried out on ensemble-mean FVR fields from the CESM, for all six regression methods. It can be clearly seen that the loading fields are of globally uniform sign in all instances. In addition, the loadings from annual regressions seem to be much more neutral on average, while those from climatological regressions are more strongly positive. However, the scales of the maps are different and absolute loading values might not be comparable. Therefore, it is more appropriate to speak of more evenly distributed loadings in the climatological regressions, where a larger number of grid points shows high loadings. Conversely, on the annual time scale, a smaller number of grid points displays high loadings, with the large majority being neutral. For the loading fields from climatological regressions, a consistently stronger emphasis on the tropics and subtropics can be observed. The mid-latitudes and polar regions, on the other hand, are more neutral. For the loading fields from annual regressions, also the tropics seem to exhibit more strongly positive loadings, but here with the exception of the central and eastern tropical Pacific. The extratropics here are markedly neutral.



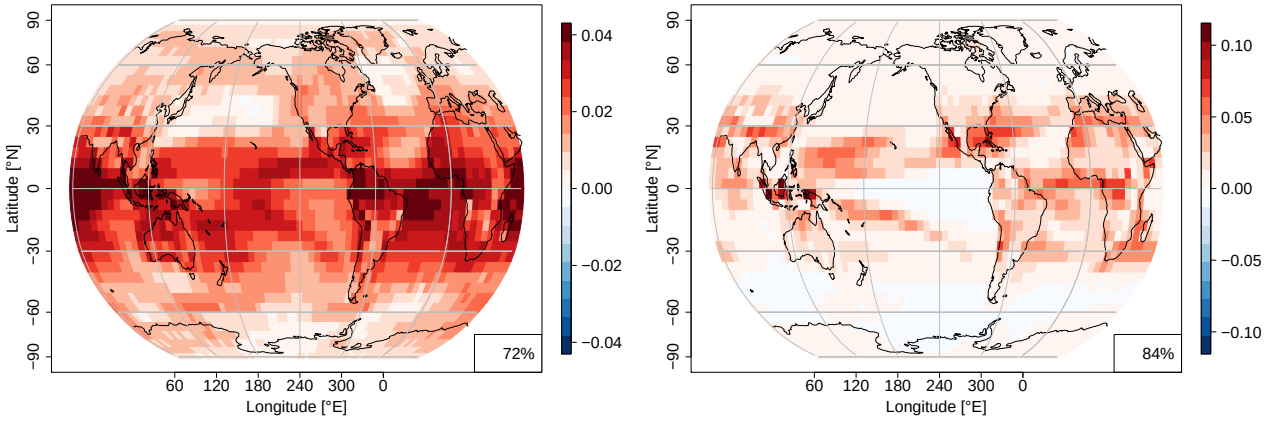
(a) Climatological Forcing Regressions

(b) Annual Forcing Regressions



(c) Climatological MEM Regressions

(d) Annual MEM Regressions



(e) Climatological GM-MEM Regressions

(f) Annual GM-MEM Regressions

Figure 4.34: *PC1 loading fields from PCAs carried out on ensemble-mean FVR fields from the CESM and all regression methods. In the lower right corner of each map, the fraction of variance explained by the respective PC1 is given. The sign of the respective PC1s has been chosen so that the mean of each loading field is positive.*

4.9.3 Climate Field Reconstruction Data

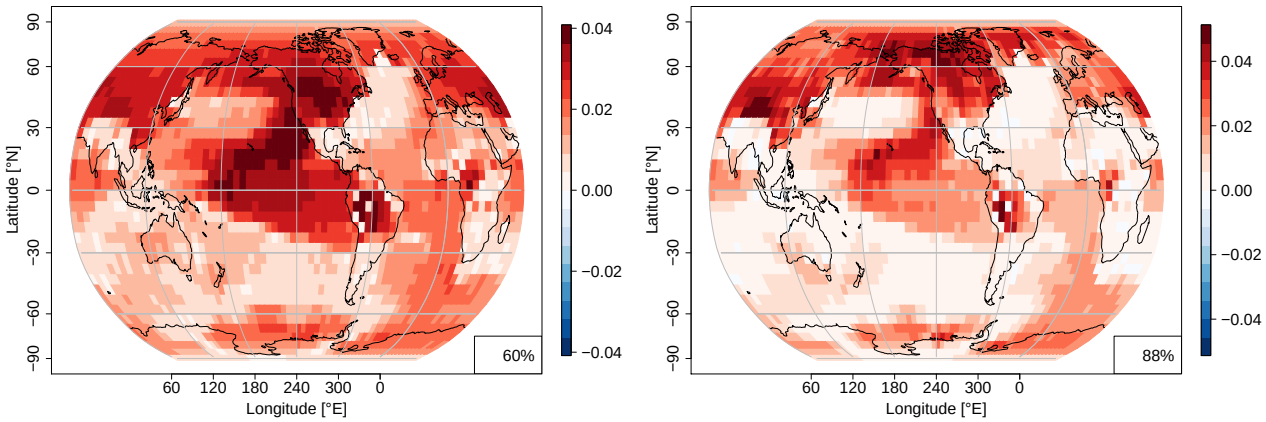
Figure 4.35 shows PC1 loading fields generated from a PCA on ensemble-mean FVR data sets from the GraphEM CFR, for all regression methods. A very consistent pattern can be seen in all sub-figures (regression methods), namely strongly positive loadings in the regions of the central and eastern Pacific, northern South America, North America, Europe and Asia. The loadings in other regions are less strongly positive, such as in the tropical and southern Atlantic ocean, and in Antarctica, or close to zero in the remaining regions. Only the loading field from annual MEM regressions differs substantially from this pattern, with most loadings being close to zero. In the loading field from the annual GM-MEM regressions, the pattern is somewhat less pronounced, yet still there. All the PC1s contain a large fraction of the variability in their respective FVR data set, with annual MEM regressions being the lowest (57%). It needs to be stated however, that this pattern seems to be characteristic only for the GraphEM CFR, and could not be found in any of the other data sets. The pattern can also clearly be seen in FVR directly in figure 4.30e.

4.9.4 Interpretation

The higher loading values in the tropics and subtropics that could be observed for the PCA from CESM suggest that on a first order, this model responds to forcing primarily in those regions. Moreover, this property is consistent with regression method and time scale and also, a large fraction of the variance in each FVR data set is accounted for by this pattern. An exception from this pattern is the markedly neutral region in the central and eastern tropical Pacific in the loading fields from the annual regressions. This pattern suggests a connection with ENSO. The region being neutral in loading implies that FVR remains stable here, while at the same time, FVR increases or decreases in the rest of the tropics. This can be interpreted as ENSO being independent of the forced response of the whole tropical region, in keeping with its properties as a mode of internal variability.

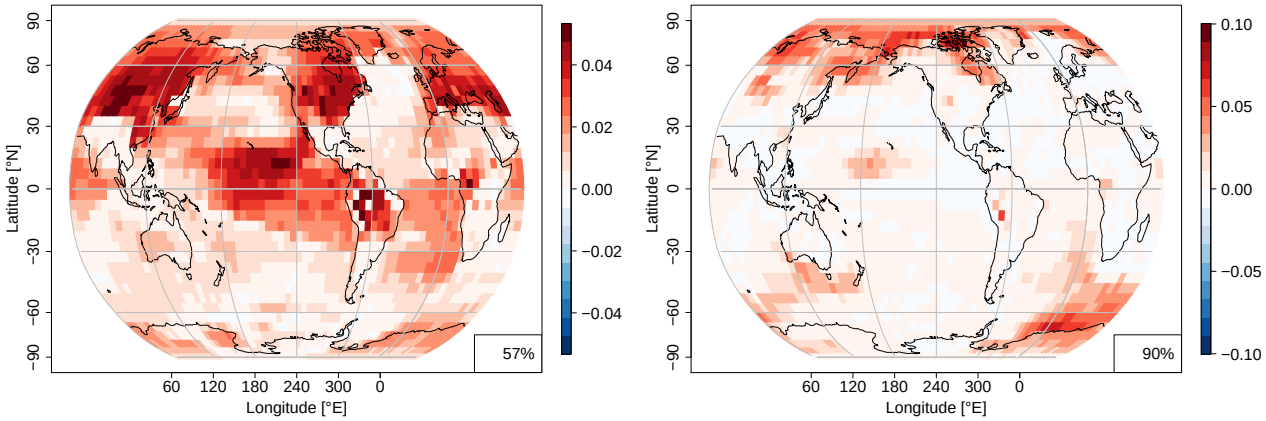
The pattern discovered in figure 4.35 is very consistent with regression method. However, it seems to be limited to the GraphEM CFR. It is reminiscent of the spatial pattern of the PDO. On a first order (fractions of explained variance are at least 60% in all instances), this pattern implies that FVR is either high or low in the PDO region, while it remains the same in the other regions not emphasized by the pattern. This corresponds to the surface temperature in the region being strongly forced in some times, and weakly forced in others. Therefore, it cannot be said that PDO is forced or unforced in general. On the contrary, it seems that PDO correlates with forcing in some times, while being uncorrelated in others. However, the pattern described is less prevalent in the PC loading maps from FVR on the annual time scale. This makes sense bearing in mind that PDO varies on the decadal time scale, so that its footprint should be more pronounced in the FVR data from climatological regressions.

The similarities presented in this chapter were mostly constrained to the results of PCA of FVR fields and not the FVR fields themselves. It could be that the filtering introduced by the truncation after PC1 reduced some of the differences between the regression methods. Also, the results shown in section 4.9 were constrained to PC1 loading fields, thus only describing one component (but the most important) of the FVR fields. The similarities shown in this chapter do not generalize to the majority of the FVR data produced in this thesis. For most of the data, the differences between the regression methods were dominant (see the next section).



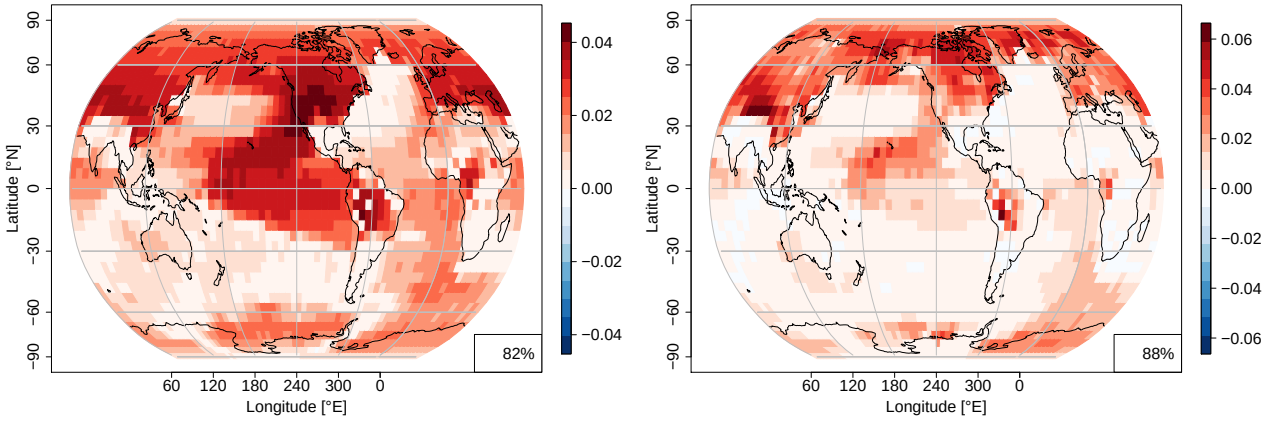
(a) Climatological Forcing Regressions

(b) Annual Forcing Regressions



(c) Climatological MEM Regressions

(d) Annual MEM Regressions



(e) Climatological GM-MEM Regressions

(f) Annual GM-MEM Regressions

Figure 4.35: *PC1 loading fields from PCAs carried out on ensemble-mean FVR fields from the GraphEM CFR and all regression methods. In the lower right corner of each map, the fraction of variance explained by the respective PC1 is given. The sign of the respective PC1s has been chosen so that the mean of each loading field is positive.*

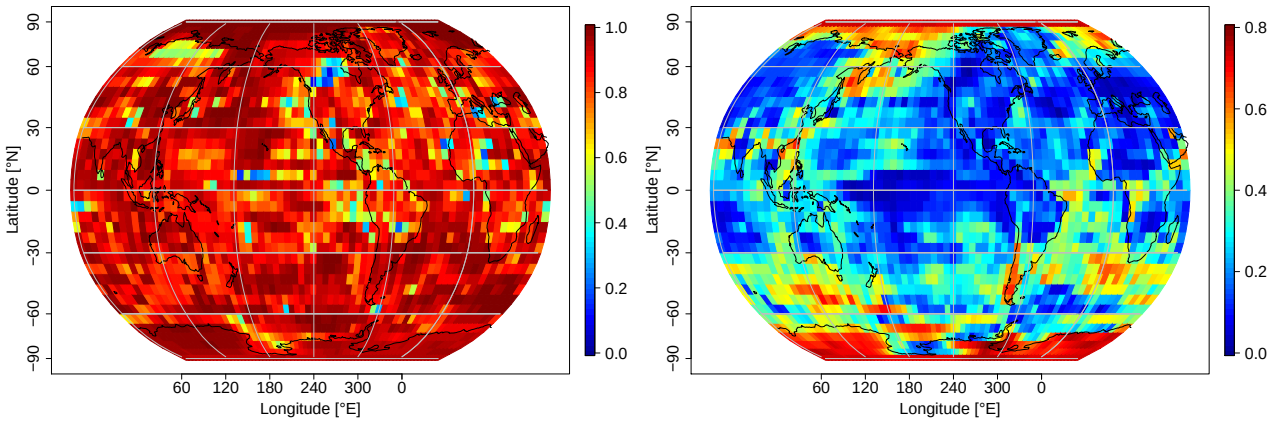
4.10 Spatial Structure of the FVR: Differences between Regression Methods

This section will give examples of disagreement between FVR data from the different regression methods. To this end, FVR maps from the last moving window (1851–2000) are shown for instrumental and reanalysis, climate model and CFR data sets. Section 4.10.1 will present all the FVR data from instrumental and reanalysis data sets produced in this thesis. Section 4.10.2 will give examples of FVR maps from the last moving window from CESM for all the regression methods. Third, section 4.10.3 will present a comparison between the the FVR fields from MEM and GM-MEM regressions for three selected CFR data sets. Last, section 4.10.4 will give some interpretations of the data shown.

4.10.1 Instrumental and Reanalysis Data

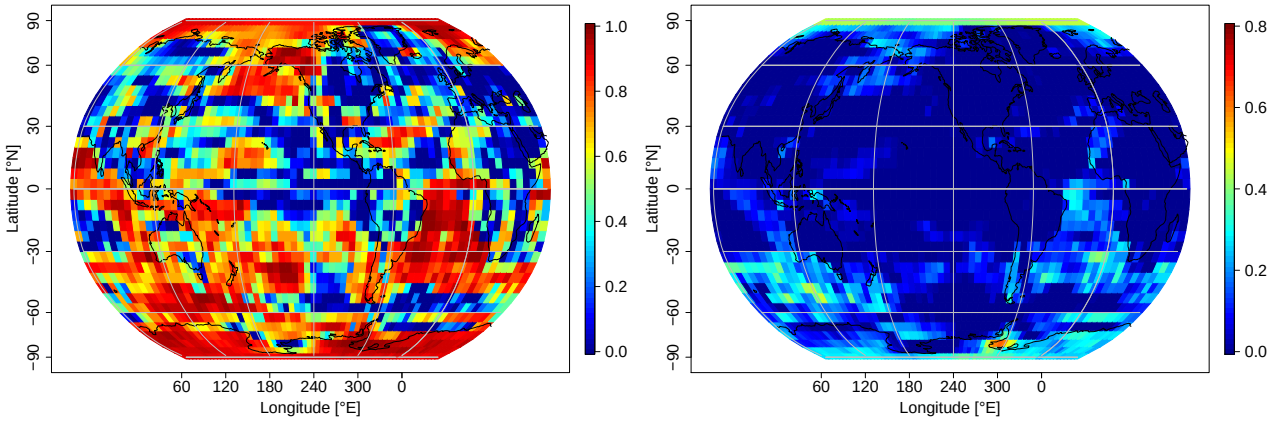
Figure 4.36 shows the FVR maps of the 20CRv2 data set, from all regression methods. Substantial differences can be observed between the different regression methods. In particular, the FVR data from climatological forcing regressions are very high and near-globally uniform. The FVR map from annual forcing regressions displays a pronounced pattern with high FVR at the two poles and predominantly low values elsewhere. The maps from climatological MEM and GM-MEM regressions on the other hand, show considerable variability in the magnitude of FVR between grid points, being relatively noisy. Further, the FVR maps from annual MEM and GM-MEM regressions show rather low and uniform distributions. Similarities exist mostly between the maps from climatological MEM and GM-MEM regressions, as well as between those from annual MEM and GM-MEM regressions.

Figure 4.37 shows the same as figure 4.36, but for the FVR data from HadCRUT4. Also here, the FVR field from climatological forcing regressions features high values in the vast majority of the grid cells. On the other hand, the FVR values from annual forcing regressions are substantially lower, with some spatial structure in them. However, it is hard to detect a clear pattern in this structure. The FVR maps from climatological MEM and GM-MEM regressions look similar to the corresponding ones from 20CRv2 (figure 4.36), but they seem more strongly patterned and less noisy. The maps from annual MEM and GM-MEM regressions are again very similar to the corresponding ones from 20CRv2 (figure 4.36), with very low overall FVR and no clear spatial structure. Therefore, also here the six regression methods produced somewhat inconsistent outcomes in terms of spatial structure and overall magnitude of FVR.



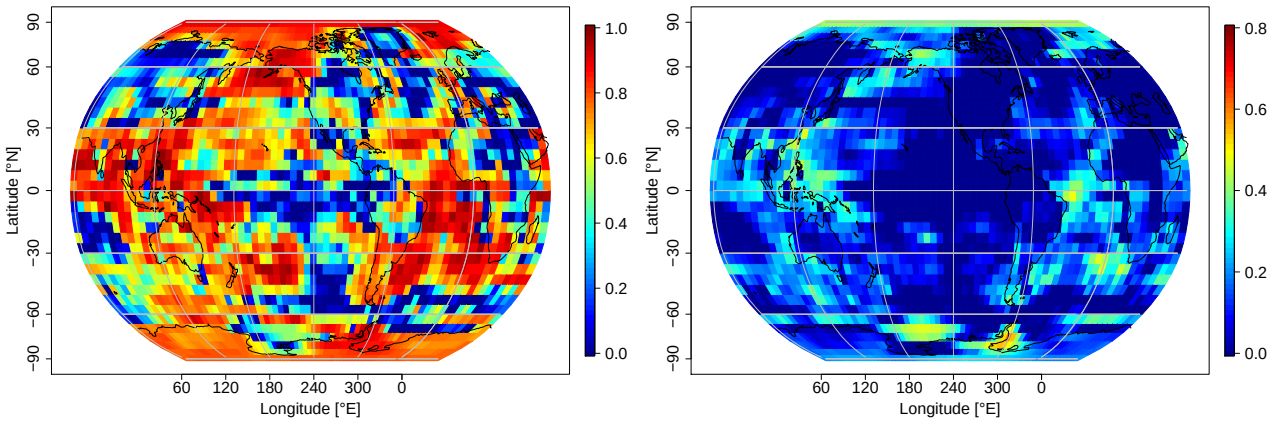
(a) Climatological Forcing Regressions

(b) Annual Forcing Regressions



(c) Climatological MEM Regressions

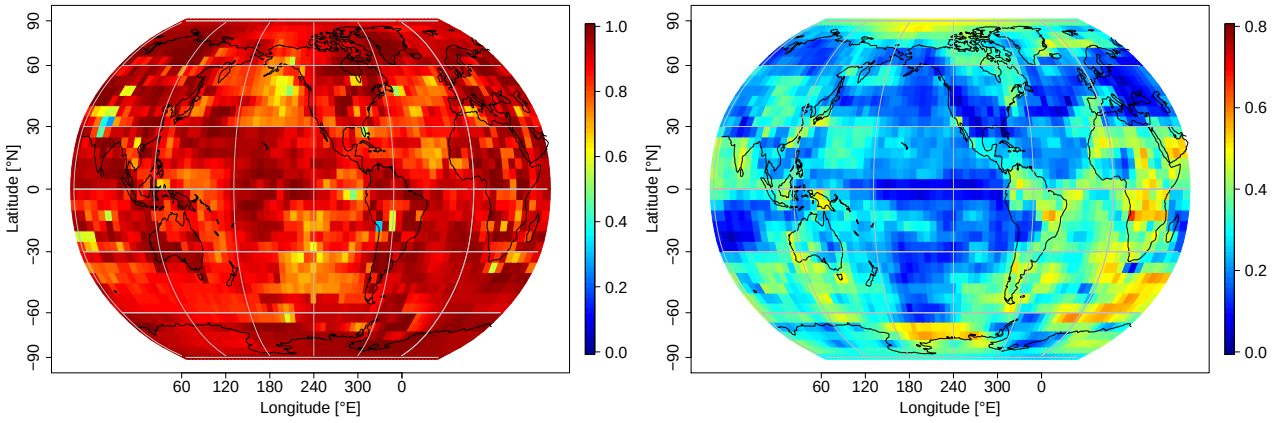
(d) Annual MEM Regressions



(e) Climatological GM-MEM Regressions

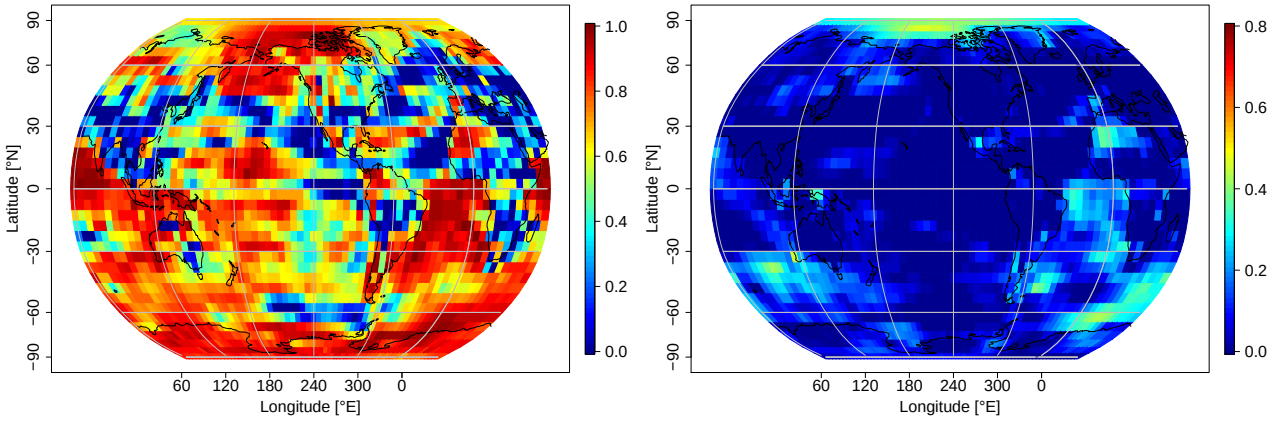
(f) Annual GM-MEM Regressions

Figure 4.36: *(a-f) FVR maps from 20CRv2, produced from the different regression methods. The data encompass the last moving window, 1851-2000, the only one available for the instrumental and reanalysis data. Note the different scale in subfigures (b), (d) and (f) as compared to (a), (c) and (e).*



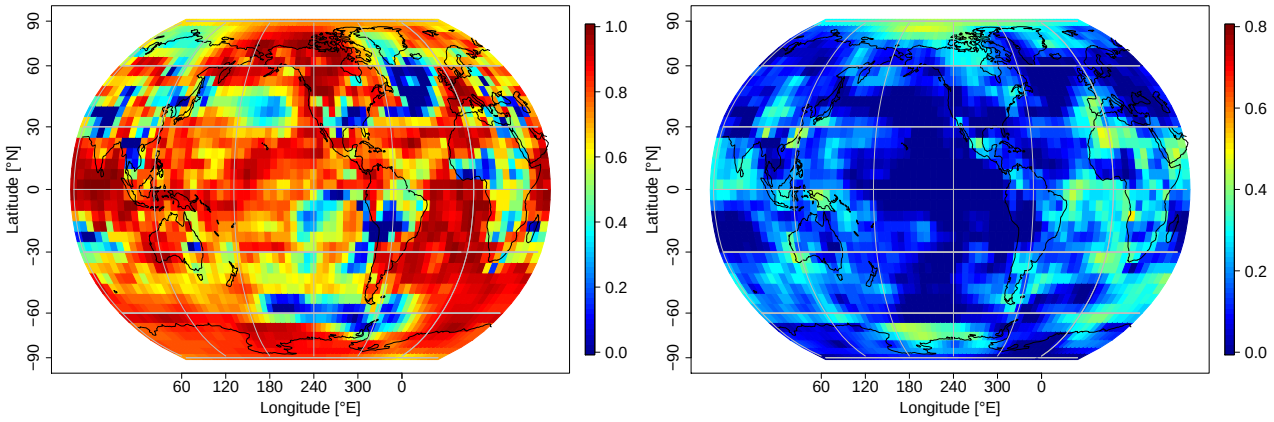
(a) Climatological Forcing Regressions

(b) Annual Forcing Regressions



(c) Climatological MEM Regressions

(d) Annual MEM Regressions



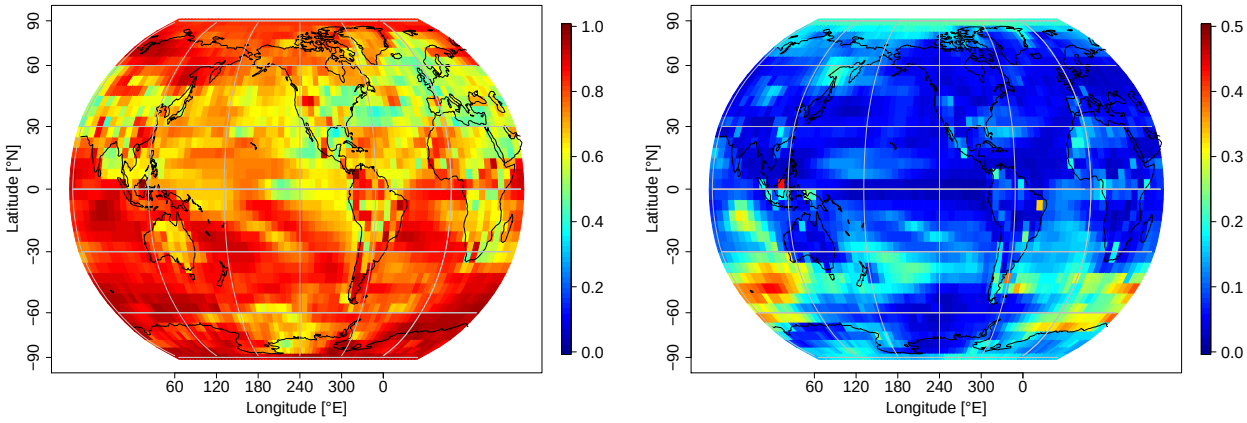
(e) Climatological GM-MEM Regressions

(f) Annual GM-MEM Regressions

Figure 4.37: *(a-f) FVR maps from HadCRUT4, produced from the different regression methods. The data encompass the last moving window, 1851-2000, the only one available for the instrumental and reanalysis data. Note the different scale in subfigures (b), (d) and (f) as compared to (a), (c) and (e).*

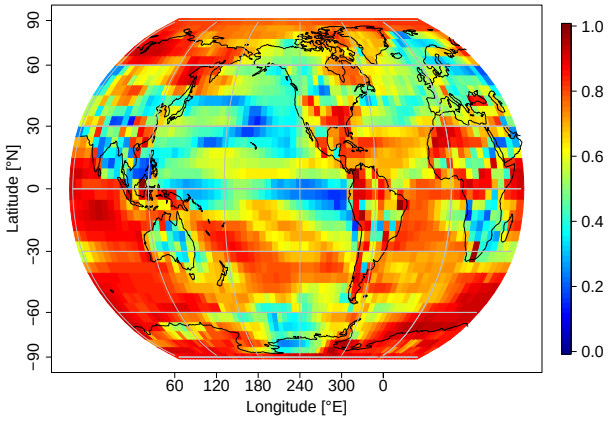
4.10.2 Climate Model Data

Figure 4.38 shows the ensemble-mean FVR from the CESM, for all the six regression methods and the last moving window. Also here, substantial differences between the regression methods can be found. The major difference consists in the magnitude of the overall FVR being much lower for the annual regressions, as already shown in section 4.4.2. In this regard, note also the different scales in the subfigures for the annual regressions as compared to the climatological regressions. Beyond that, the FVR for the climatological forcing regressions seems to be rather globally uniform, with somewhat lower values in the North Atlantic sector, parts of North America, Europe, the Middle East and Africa. For the climatological MEM regressions, there are more extensive areas of lower FVR, namely in the eastern tropical Pacific, in the North Pacific and in the Pacific sector of the Southern Ocean. The FVR map from the climatological GM-MEM regressions shows areas of low FVR mainly in the northern hemispheric mid-latitudes, with the southern hemisphere and the Arctic being quite strongly forced. For the three annual regressions, FVR is quite low for most of the globe, with the exception of the Indian ocean sector of the Southern Ocean in the annual forcing and MEM regressions.

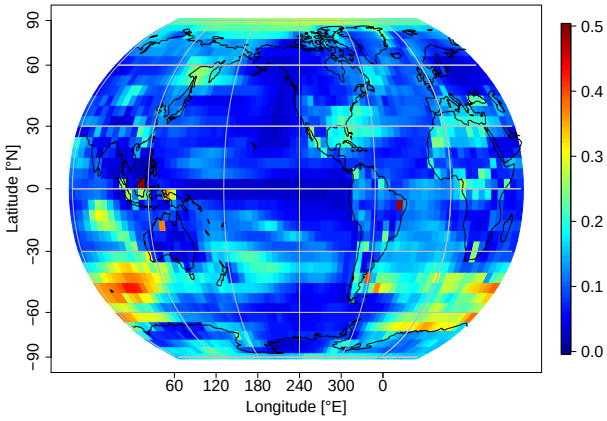


(a) Climatological Forcing Regressions

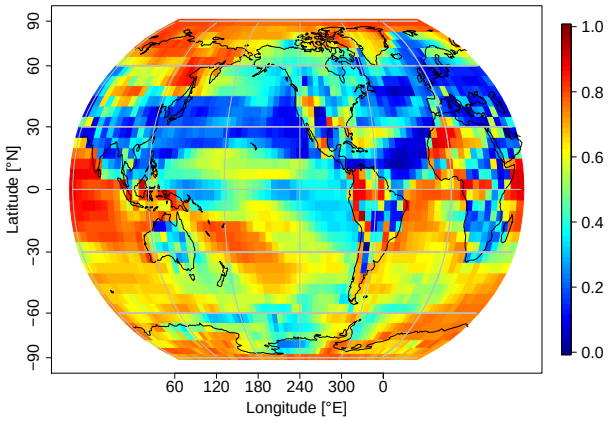
(b) Annual Forcing Regressions



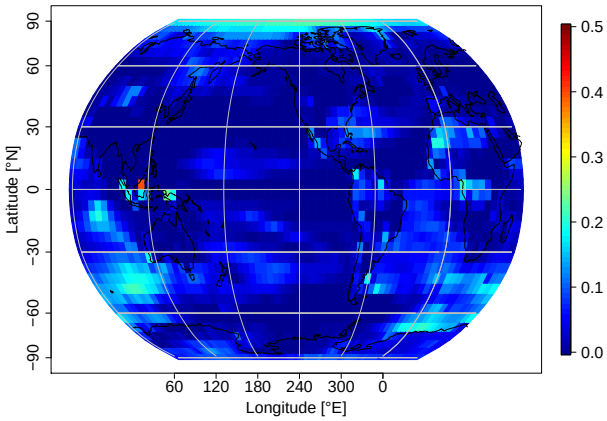
(c) Climatological MEM Regressions



(d) Annual MEM Regressions



(e) Climatological GM-MEM Regressions



(f) Annual GM-MEM Regressions

Figure 4.38: *Ensemble-mean FVR maps from the last moving window (1851-2000) of the CESM, from all six regression methods. Note the different scale on the maps from annual regressions ((b), (d) and (f)) as compared to the ones from climatological regressions ((a), (c) and (e)).*

4.10.3 Climate Field Reconstruction Data

Figure 4.39 shows ensemble-mean FVR fields from the Analogue, CPS and DA CFRs, for the last moving window (1851-2000). In contrast to figures 4.36, 4.37 and 4.38, however, here only the results from MEM and GM-MEM regressions are shown for the three CFRs. It is clearly visible that the FVR fields from GM-MEM regressions are much more spatially homogeneous than the ones from MEM regressions. The FVR fields from GM-MEM regressions have more values close to the average in general, while the ones from MEM regressions are very high in some regions and very low in others. Since the figures 4.39b, 4.39d and 4.39f are the same as the corresponding ones in figure 4.25, the considerations about their spatial structure already made in section 4.7.1 apply here (low forcing of the North Atlantic sector as opposed to other regions).

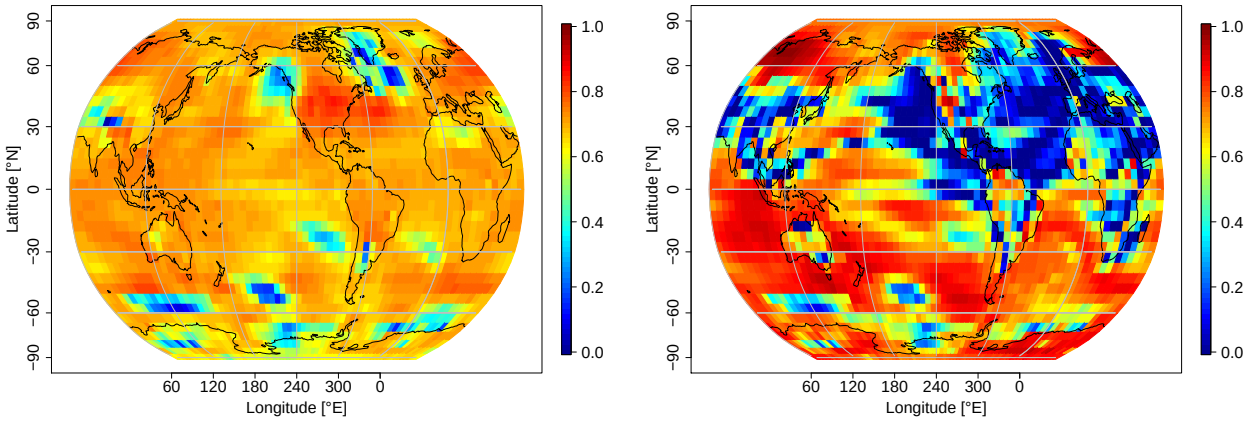
4.10.4 Interpretation

It appears that for all of instrumental and reanalysis, climate model and CFR data, the spatial structure of FVR strongly depends on the regression method chosen. This result is valid beyond the examples shown here, and was true for most of the FVR data produced in this thesis. Notable exceptions thereof were presented in section 4.9.

Nevertheless, many of the differences presented can be explained. For example, aerosol effects can explain the low FVR values in the northern hemisphere of the instrumental and reanalysis data sets (climatological MEM regressions of 20CRv2 and HadCRUT4). Correction for this effect would yield FVR maps more similar to the climatological forcing regressions, with globally consistent and high FVR. A similar effect could be the cause for lower northern hemisphere FVR in the climatological GM-MEM regressions of CESM (figure 4.38e). Additionally, some low-forced regions could be due to dynamical properties of the climate system, for example upwelling ocean currents that do not capture the forced response yet (for example the Humboldt current region in figure 4.36e). Also the differences between annual and climatological regressions can be explained. There, the considerations made in 4.4.4 apply (weather-related noise on the annual time scale, respectively enhanced R^2 due to autocorrelation on the climatological time scale).

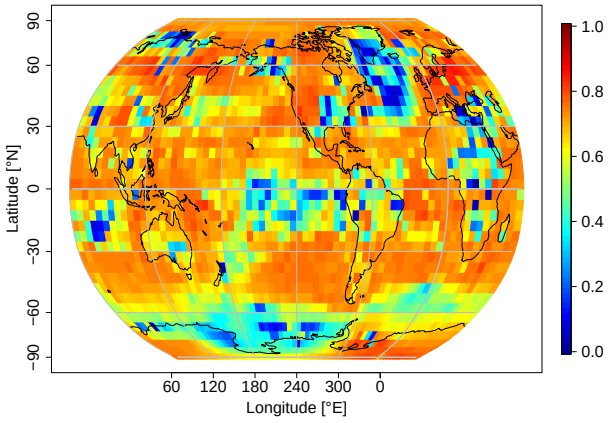
The comparison of MEM and GM-MEM FVR maps shown in figure 4.39 clearly hints at the difference between the two regression methods. The obvious interpretation here is as follows. In GM-MEM, the temperature time series from the three CFRs was regressed against the GM-MEM of the CESM, the same time series for every grid point. This implied much greater spatial uniformity than for MEM regressions. There, the temperature time series were regressed against the MEM of CESM at each grid point individually, allowing for the possibility of higher agreement in some areas and lower agreement in others. This then manifests itself in the much higher heterogeneity in FVR fields for MEM as compared to GM-MEM.

Therefore, even though substantial differences could be shown between the different regression methods, many of those differences can be explained.

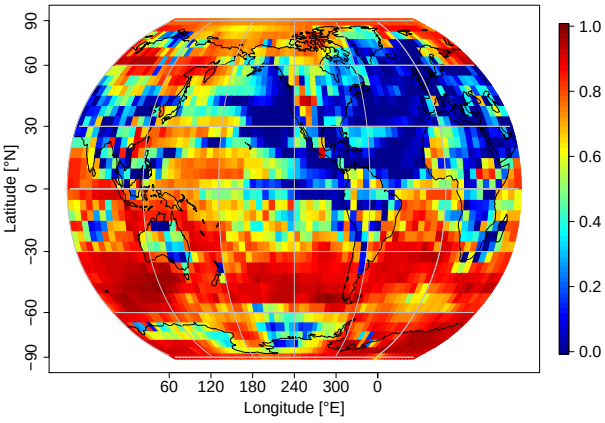


(a) Analogue GM-MEM

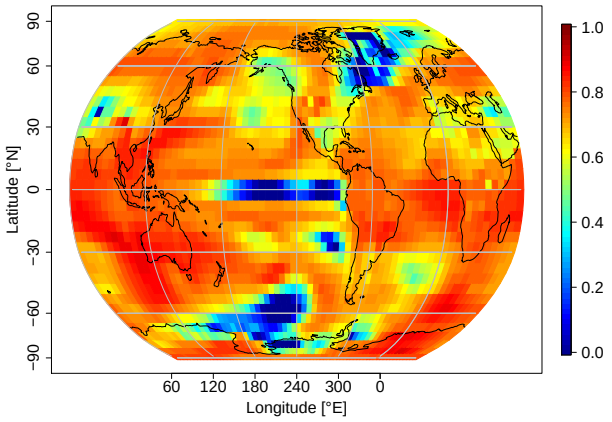
(b) Analogue MEM



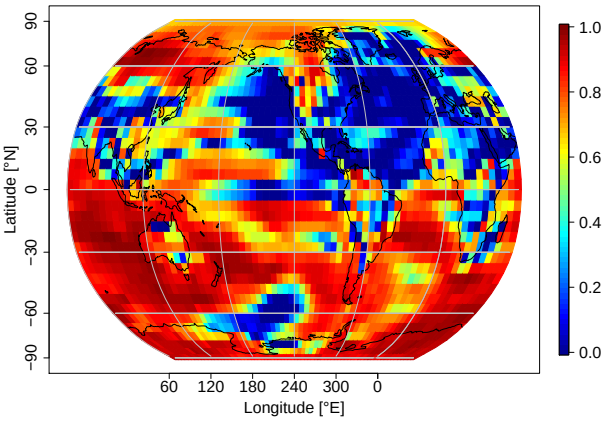
(c) CPS GM-MEM



(d) CPS MEM



(e) DA GM-MEM



(f) DA MEM

Figure 4.39: *Comparison of MEM and GM-MEM regression FVR data for selected CFR data sets for the last moving window (1851-2000).*

4.11 Validation of the Linear Regression Calculations

4.11.1 Cross-Validation

Also the cross-validation of the regression calculations described in section 3.2.4 confirmed the results generated before. That is, both the original calculations of FVR and the cross-validations yielded the same values of R^2 (differences $\leqslant 10^{-4}$) in a very large majority of the data points sampled.

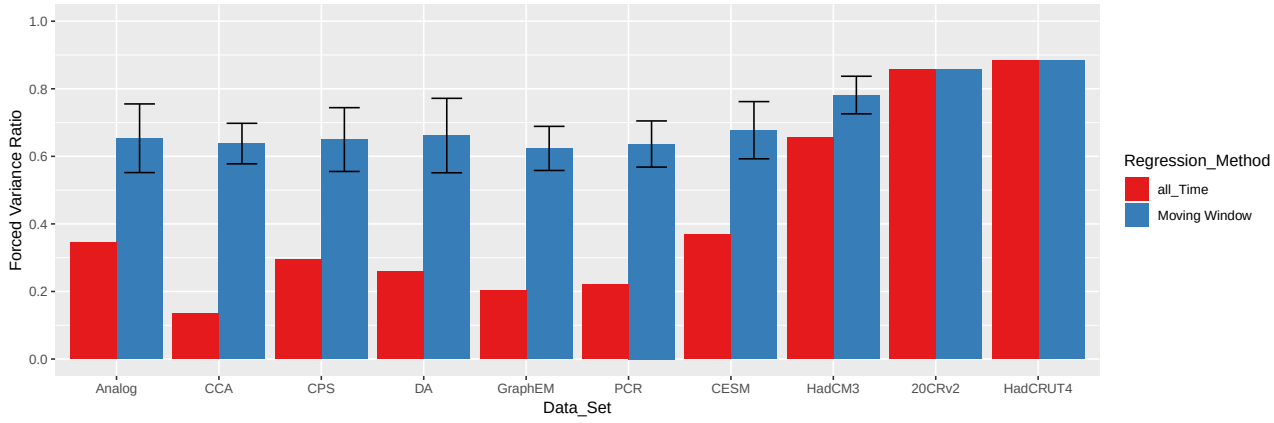
4.11.2 All-Time Regressions

The results of the climatological regressions carried out over the whole time of data availability are visualized in figure 4.40 and compared to the ones from the climatological moving-window regressions. It is evident that the differences between regressions with moving windows and regressions over the whole time are most pronounced for the CFR data sets, less so for CESM and even less for HadCM3. For the instrumental and reanalysis data sets, the differences are zero by definition, since the whole time of data availability equals the time covered by the last moving window. Therefore, there seems to be a dependence between the total length of time covered by the data set and the difference in GM-EM FVR between the all-time regressions and the moving window regressions.

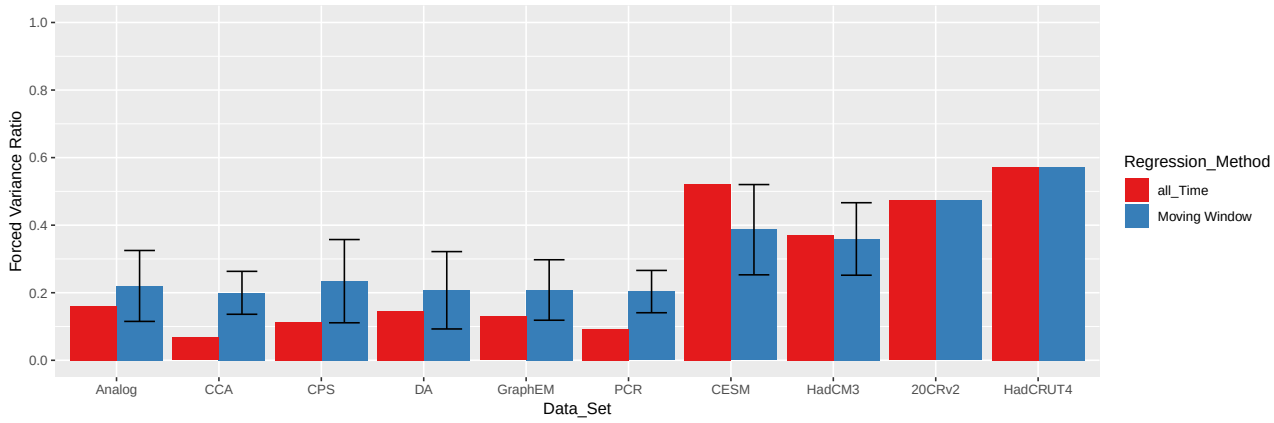
When the individual data sets are considered, the difference between all-time regressions and moving window regressions seems most pronounced for the CCA CFR. The other CFRs also exhibit substantial differences, though less than CCA. The case of the CESM is most interesting, with a substantially lower GM-EM FVR value in forcing regressions, even a higher one in MEM regressions and a similar one in GM-MEM regressions. HadCM3 on the other hand seems to be quite constant, with relatively small differences throughout.

Additionally, the absolute differences in FVR seem to be highest in the forcing regressions, and lower in the MEM and GM-MEM regressions. For the differences relative to the respective value from the moving window regressions, the same appears to be true, if less pronounced. Also, for the forcing regressions, the all-time regression FVR values are clearly further outside the range of the standard deviations of the moving window regressions. For the MEM regressions, more GM-EM FVR values fall within this range, and even more for GM-MEM.

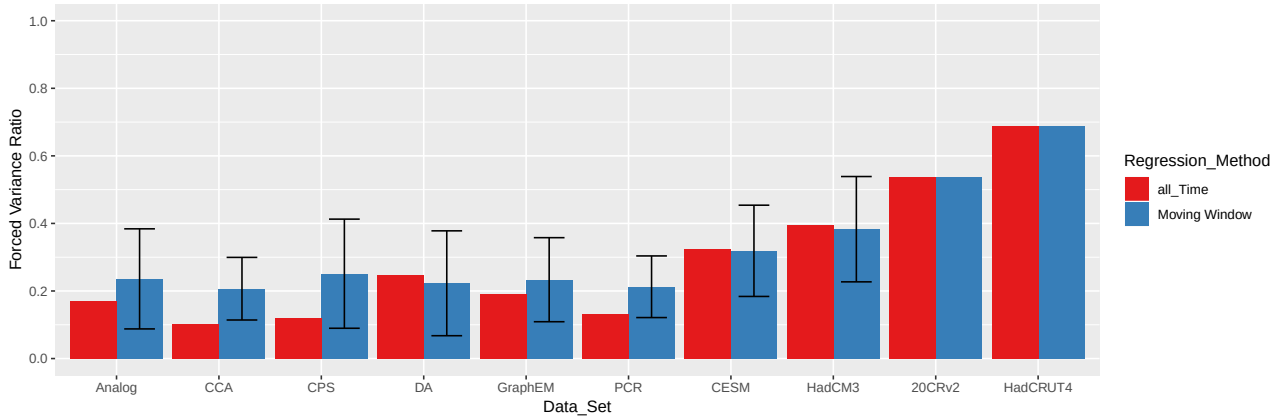
The all-time regressions have been carried out on the annual time scale as well. There, the results are more consistent between moving window and all-time regressions. The GM FVR values from the all-time regressions mostly fall within the range of the moving-window regressions (mean plus/minus one standard deviation of moving windows, results not shown).



(a) Data from forcing regressions



(b) Data from MEM regressions



(c) Data from GM-MEM regressions

Figure 4.40: *GM-EM FVR of climatological regressions covering the whole period of data availability (red) as compared to the moving window approach (blue). The blue bars indicate the mean over all moving windows. Error bars indicate this mean plus/minus one standard deviation of moving windows. (a) Data from forcing regressions, (b) data from MEM regressions, (c) data from GM-MEM regressions.*

4.11.3 Tests for Multicollinearity in Forcing Regressions

Table 4.1 shows the outcomes of the assessment for multicollinearity in the forcing predictors described in subsection 3.2.4. The numbers give the percentages of the samples in which multicollinearity was detected.

For the climatological forcing regressions, multicollinearity was detected in several of the indicators and in substantial amounts of the samples. It appears that the climatological forcing regressions were more prone to multicollinearity than the annual ones. Still, also for the annual regressions, multicollinearity is detected in a substantial fraction of the samples. This is mostly true for the condition number test and Farrar's Chi-Squared, to a lesser extent for Theil's method. For the climatological regressions, the same indicators exhibit large fractions of multicollinearity detection, with the addition of the Red indicator.

	Annual [%]	Climatological [%]
Determinant |X'X|	0	0
Farrar Chi−Square	80.1	100
Red Indicator	8.2	48.5
Sum of Lambda Inverse	0	19.1
Theil's Method	35.2	41.6
Condition Number	100	100

Table 4.1: *Outcomes of the assessment for multicollinearity described in 3.2.4. The numbers given are the percentages of samples for which multicollinearity was detected. The middle column gives the results for the annual forcing regressions, the column on the right for the climatological forcing regressions.*

4.11.4 Interpretation

The cross-validation confirmed the results of the linear regressions calculated before. Therefore, there did not seem to be errors in the calculations themselves and they can be regarded as correct, apart from conceptual errors.

The all-time regressions indicated substantially higher FVR than the regressions with moving windows (figure 4.40). This is indicative of overestimation of FVR in the moving window approach. Since the reduced degrees of freedom and autocorrelation are less of a concern in the all-time regressions (their time series are much longer and therefore have enhanced degrees of freedom), their results likely are more confidential. On the other hand, the results from the climatological moving-window regressions need to be interpreted with care. In them, the general magnitude of FVR most likely is overestimated.

Multicollinearity could be detected in significant fractions of the samples from forcing regressions (both climatological and annual). This suggests reduced confidence also in their FVR values. The problem is most pronounced in the climatological forcing regressions, less so in the annual ones.

Based on the results from this section, an attempt can be made to determine the most realistic values of FVR. The good agreement between the all-time and moving window regressions for the two

model data sets (climatological MEM and GM-MEM regressions, figure 4.40) suggests high confidence in this result. Therefore, FVR can be robustly estimated to be about 0.3–0.4 over the last millennium on the climatological time scale. The estimates from the all-time regressions of the CFR data sets with values around 0.1-0.2 can be taken as a lower boundary. On the other hand, the estimates from the moving-window forcing regressions (figure 4.40a, FVR values of around 0.7) can be taken as a clear upper boundary, respectively overestimation.

Similar considerations can be made at the annual time scale. Also there, results from the forcing regressions need to be interpreted with care, since they suffer from multicollinearity. However, autocorrelation in regression residuals is not a problem here. Therefore, from figures 4.13 and 4.14, and also from figures 4.17 and 4.19, FVR can be robustly estimated to be rather low (<0.05) on the annual time scale over the period 1401–2000.

5 Discussion

5.1 Global-Mean FVR

In section 4.1, it was demonstrated that FVR exhibits substantial variability over the common era. This result defies the notion of a time-invariant FVR, in accordance with the findings of Lyu et al. 2015. Moreover, several periods of low and high GM FVR could be identified as consistent across data sets and regression methods on the climatological time scale. The agreement lends enhanced confidence to this result. As was suggested in section 4.1.2, the period of enhanced FVR from the 14$^{\text{th}}$ to 16$^{\text{th}}$ centuries could be due to enhanced volcanic forcing. On the one hand, the enhanced FVR cannot have been caused by higher GHG concentrations, as they are still at low pre-industrial values at this time. Further, solar forcing likely is not the cause either, since it was demonstrated not to have impacted climate strongly over the last millennium (Schurer et al. 2013b), respectively not significantly over the common era (PAGES 2k Consortium 2019). However, volcanic forcing was demonstrated to have impacted climate at this time (PAGES 2k Consortium 2019), leading to the onset of the little ice age (Miller et al. 2012). The rise of FVR starting from the 19$^{\text{th}}$ century most likely is caused by anthropogenic GHG emissions. In addition, Brönnimann et al. 2019 could show that the onset of this high-forced period was caused by volcanic eruptions.

It was demonstrated before that climate model simulations might not represent the dynamical response to volcanic eruptions well, and in particular overestimate the cooling of the tropical troposphere (Driscoll et al. 2012). Gent et al. 2011 showed that Community Climate System Model version 4 (CCSM4) as well overestimates the cooling in response to volcanic eruptions. Finally, Otto-Bliesner et al. 2016 demonstrated the same effect for CESM as compared to northern hemispheric surface temperature reconstructions. This effect can explain the strong response of CESM in response to the eruption of the Samalas volcano and the corresponding high FVR values, as compared to the CFR data. However, it remains questionable why only the Samalas eruption, and no other large volcanic eruptions, led to such a strong response.

In section 4.3, it was shown that the CFR and climate model data sets exhibit significant differences in terms of GM FVR for most combinations of data sets. However, the differences likely are significant due to the very large sample sizes and the data sets broadly agree on the overall magnitude of GM FVR. This agreement of CFR and climate model data sets is striking since the data come from fundamentally different sources for the two groups of data sets. Also, the agreement lends enhanced confidence to this result. Some agreement can even be shown for the temporal development of the GM FVR (see above and section 4.1). The agreement between CFR data and climate model data was shown before for GM surface temperature (PAGES 2k Consortium 2019), now this result is extended to GM FVR.

In section 4.4, significant differences between the regression methods in terms of the magnitude of GM FVR were demonstrated. In most cases, the FVR values from MEM and GM-MEM regressions were

much lower than those from forcing regressions. These lower FVR values could be due to poor agreement of the forcing of CESM, and consequently its MEM, with the real forcing. This is corroborated by the fact that the FVR values from CESM are generally higher than those of the other data sets. This is an indication that FVR values from MEM and GM-MEM regressions might be slightly too low. However, the forcing regressions have been shown to suffer from multicollinearity (section 4.11.3) and their FVR values were most strongly reduced when subjected to regressions with the entire time series (section 4.11.2). Therefore, the FVR values from MEM and GM-MEM regressions likely are closer to the truth and the ones from forcing regressions likely are too high. The disagreement with MEM and GM-MEM regressions, the stronger reduction in all-time regressions and multicollinearity all speak for reduced confidence in the results from forcing regressions. This is most true for the climatological forcing regressions, less so for the annual ones.

Even though significant differences have been detected between the GM FVR data for MEM and GM-MEM regressions, they broadly agree in the overall magnitude of FVR in most cases. An exception thereof are the model FVR data on the annual time scale, where CESM seems to agree better with its own grid-point wise MEM than with its GM-MEM. With this exception explained, the MEM and GM-MEM regressions can be said to broadly agree.

The constraints of autocorrelation and reduced degrees of freedom, discussed in section 5.3, pose some limits on the validity of the estimates of GM FVR from climatological MEM and GM-MEM regressions. Also, their FVR values might be slightly too low due to disagreement between the forcing of the CESM and the real forcing, as discussed above. However, the constraints of autocorrelation and reduced degrees of freedom were addressed by the regressions with the entire time series. Their results can therefore be taken as the most robust estimate of GM FVR on the climatological time scale. This most robust estimate was determined to be about 0.3–0.4 over the last millennium. For the annual time scale, a rather low estimate of FVR < 0.05 was considered most realistic over the period 1401–2000.

5.2 Spatial Structure of the FVR

The relative importance of internal variability is higher at smaller spatial scales (Hawkins & Sutton 2009). Therefore, lower values of FVR can be expected at those scales. This has strong implications for the results of this thesis, since all the regression calculations were carried out on grid cells of 5° longitude by 5° latitude spatial resolution. At higher spatial resolution, FVR would likely be found lower than in the present analysis. Conversely, at lower spatial resolution, FVR would likely be found higher.

The regular longitude/latitude spacing of the grid in the present analysis implies that the grid cells at high latitudes are smaller in terms of area than those at low latitudes. As discussed above, this smaller spatial scale implies lower values of FVR. It would be possible to correct for this effect. For example, one could divide the FVR at each grid cell by the area of the grid cell, assuming a linear relationship between spatial scale and FVR. This would enhance the FVR at high latitudes and lower FVR at low latitudes, in relative terms. This dependence between FVR and latitude needs to be kept in mind for the following parts of the discussion, where the spatial structure of FVR is described.

The results from the zonal-mean FVR data revealed consistent features within the three groups of

instrumental and reanalysis, climate model and CFR data sets. However, across the three groups, little common features could be found. A likely reason for that is that the respective data sets span different amounts of time. This is most pronounced for the instrumental and reanalysis data sets, which only cover the period from 1851–2000. Therefore, to enhance comparability, zonal means were investigated over the same time periods.

However, also over the period 1401–2000, no agreement in the zonal-mean FVR could be found between climate model and CFR data sets. The climate models mostly showed distributions with a global maximum in the tropics, local minima in the mid-latitudes of both hemispheres and local maxima in the high latitudes. The maximum in the tropics is in keeping with the results of Lyu et al. 2015 and basic climatological reasoning, as climate there is much more persistent and therefore a signal from forcing should be more readily apparent (see section 4.8.3). The local maxima in the high latitudes are in contradiction to the aforementioned effect of the smaller spatial scale on FVR. It therefore seems likely that the maxima in the high latitudes would be even more pronounced if one were to correct for this effect. However, it needs to be noted that the uncertainty is relatively high in the FVR values of the models at high latitudes, for the period 1401–2000. Over this period, the CFR data sets showed rather uniform distributions of FVR with latitude. Correction for the effect of spatial scale would likely lead to enhanced values of FVR at high latitudes, as for the climate model data.

When the data sets were compared over their common period 1851–2000, good agreement could be found on the annual time scale. In particular, all data sets agreed in a local maximum of FVR in the Arctic. Similar to the model data sets in the period 1401–2000, correction for the effect of the smaller spatial scale would enhance this feature. These maximum values of FVR in the Arctic are in agreement with the strong warming in this region that could be observed in the 20th century and attributed to anthropogenic GHG forcing (IPCC 2013; Gillett et al. 2008; Wang et al. 2007). In contradiction to that, Mahlstein et al. 2011 and Hawkins et al. 2014 could show that GHG warming emerges earlier in low latitudes, a property that is traditionally associated with regions of high FVR (Lyu et al. 2015). However, for the industrial period, no data set or regression method displayed higher than average values of FVR in the tropics in the present analysis, in contradiction to the results of Lyu et al. 2015. Therefore, the relationship between FVR and emergence of the anthropogenic GHG warming signal might be more complex than one would intuitively think. However, it can be stated also that the pattern with higher FVR in the tropics could be found in the FVR data from the two climate models, when the time of analysis is expanded beyond the industrial period (figures 4.21 and 4.31).

Section 4.6 demonstrated the spatiotemporal evolution of FVR using a PCA of two example data sets. It could be shown that FVR exhibits a complex spatiotemporal development and is neither uniform in space nor constant in time. This result holds true over all climate model and CFR data sets and can therefore be considered robust.

From visual examination of FVR maps (results not shown), it was apparent that the patterns of FVR evolve slowly enough to be resolved on a 10-year time step (the leap time of the moving windows used here). However, it could be that this relatively slow evolution is due to the properties of the underlying moving windows approach. Namely, with moving windows of 150 years length and 10 years leap time, the amount of data entering and leaving the moving window is relatively small as compared to the amount of data that stays within the window. Thus, it is likely that the R^2 value of one moving window is similar to the R^2 value of the previous moving window (autocorrelation). Therefore, it is

unclear whether the speed of this evolution is an actual property of the climate system or whether it is an artifact of the analysis used here.

Some similarities that have been found between the data sets were presented in section 4.7. The most consistent among these is the PC1 loading field of globally uniform sign, which could be found in all climate model and CFR data sets, in all three regression methods and both time scales. Also, the development of the PC1 score time series was very similar among the data sets. Therefore, this result seems to possess general validity. An analogue result was obtained by Wills et al. 2018, who found a spatially uniform first-order response of Pacific sea surface temperatures to anthropogenic GHG forcing. It therefore seems likely that forcing affects surface temperatures in a globally uniform manner, on a first order. Similarities could also be found in the spatial distribution of FVR for the six CFR methods in the time from 1851–2000 and the PC2 loading fields and score time series of the FVR data from climatological MEM regressions.

However, beyond the similarities mentioned before, little agreement in spatial structure could be found among the data sets. Examples of disagreement between data sets were presented in section 4.8. It needs to be stated that those dissimilarities outweighed the similarities overall. One reason for that could be high uncertainties in regions with sparse data in the CFRs. Another reason could be uncertainties in simulating internal variability and regional responses to external forcing in the climate models.

The global uniformity in the PC1 loading fields and its consistency with data set described before could be observed for all regression methods. The conclusion that FVR varies in a globally uniform manner on a first order therefore seems robust. For forcing and GM-MEM regressions, this uniformity seems plausible intuitively, as the forcing predictor is the same for every grid cell. For the MEM regressions, this seems less likely, because the forcing predictor is different at every grid cell. Moreover, for all regression methods, the response variable (surface temperature) is different at every grid cell, so that it seems even more surprising to see a first-order response of globally uniform sign.

Furthermore, for the predictors in the forcing regressions, it seems reasonable to assume that the GHG forcing is globally uniform, since the most important GHG are well mixed in the atmosphere. For the solar forcing, global uniformity cannot be assumed due to latitudinal gradients in solar irradiation, even when averaged over the course of a year. However, it needs to be said that solar variability was shown not to strongly affect climate in the common era, as mentioned above (Schurer et al. 2013b; PAGES 2k Consortium 2019).

For the volcanic forcing, the distribution of the sulphate aerosols in the stratosphere depends on the location of the volcanic eruption as well as on the atmospheric circulation in the time following the eruption. In fact, the response of surface temperature to volcanic eruptions could be shown to be spatially heterogeneous (Kelly et al. 1996; Eliseev & Mokhov 2008). It was argued that these spatial differences are "reduced by the seasonally opposed influences of the radiative response (summer continental cooling) and stratospherically forced dynamical response (winter continental warming)" and that therefore "the long-term annual average impact of volcanoes is relatively homogeneous spatially" (Shindell et al. 2003). However, also in the cross-correlation analysis performed in this study (as a precursor to the forcing regression calculations), surface temperature was seen to respond spatially heterogeneous, both positively and negatively, to volcanic eruptions (results not shown). However, this cross-correlation analysis was not tested statistically and therefore its results need to be interpreted

with care. Therefore, it may remain open to debate whether the surface temperature response to volcanic forcing is spatially heterogeneous or not.

Overall, the behavior of FVR is therefore likely to be spatially heterogeneous, due to spatially explicit predictors and predictands, spatial gradients in solar irradiation and potentially spatially heterogeneous responses to volcanic eruptions. However, this spatial heterogeneity seems to be contained in the higher-order PCs, with the globally uniform signal contained in the first PC.

When comparing the FVR data from MEM and GM-MEM regressions, the much higher spatial variability in the MEM regressions is apparent. This was shown in section 4.10.3. In general, the FVR fields from MEM regressions were somewhat more noisy, whereas the ones from GM-MEM regressions were more smooth. This difference can easily be explained by the spatial uniformity of the GM-MEM predictor, as opposed to the spatially explicit form of the MEM predictor(s). The MEM predictor(s) can be said to have scored more "lucky hits" with high agreement between predictor and predicted variable and also more grid points with low agreement, resulting in higher variability. The R^2 values from GM-MEM regressions, on the other hand, scattered more closely around the mean, resulting in more smooth spatial patterns.

The FVR maps shown in section 4.10 displayed both regions of high and low FVR in the instrumental period, on the climatological time scale. However, for the regions of low FVR, explanations could be provided. Therefore, corrections for those effects (incorporation of aerosol forcing in the methodology, for instance) would likely yield a globally rather uniform field of high FVR on the climatological time scale.

Therefore, concerning the issue of the global warming hiatus raised in the introduction, it can be said that surface temperature is clearly forced during the period 1851–2000, at least on the climatological time scale (see the results in chapter 4.10). This is true for most of the earth's surface, and for regions displaying low FVR, explanations could be given. Therefore, temporary hiatus periods are likely due to internal variability signals superimposed on a background warming signal. This result is in agreement with previous research (e.g. Trenberth 2015).

Similar considerations can be made for the issue of detection and attribution of climate change. That is, detection and attribution of climate change should be possible in most regions, at least on the climatological time scale. However, detection of regional climate change has proven to be a significant challenge (Stott et al. 2010; IPCC 2013). This could be due to the fact that detection and attribution studies are carried out using annual-mean (or even more highly resolved) data. Analyses on filtered data are likely limited by the same constraints of autocorrelation and reduced degrees of freedom as have been encountered in this thesis, so that analyses on the climatological time scale are methodologically problematic.

Since FVR was found to be high for large portions of the earth's surface during the instrumental period, uncertainty of future climate projections can be said to strongly depend on the GHG emission scenario. This is in keeping with previous research (Hawkins & Sutton 2009; Hawkins & Sutton 2011).

Considering the challenge of finding optimal locations for paleoclimate reconstructions (Bradley 1996), no definite outcomes can be presented. FVR was found to vary strongly with space and time over the CE, and results were found to depend on data set and method of analysis. Therefore, further research is needed to address this challenge. Improved methodology (see suggestions in section 6.2) could possibly shed more light on this issue.

5.3 Limitations

The most important constraint to the validity of the results presented in this thesis is the limited validity of the linear regressions models used to produce the FVR data. In particular, for a linear regression model to be valid, its residuals need to be independent (i.e. not autocorrelated) and homoscedastic, among other conditions (Sheather 2009). However, particularly the residuals from climatological regressions exhibited substantial autocorrelation. This autocorrelation was introduced by the filtering of the dependent and independent variables before the regression. One consequence of autocorrelated residuals is overestimation of R^2 (e.g. Mukherjee et al. 1998). Therefore, caution is appropriate when interpreting the FVR data from climatological regressions. Indeed, when performing the regressions again over the whole time of data availability to account for reduced degrees of freedom, the FVR values were substantially lower (see section 4.11.2). Therefore, the FVR values presented in 4.11.2 can be regarded as more realistic than the ones from the climatological moving window regressions. To account for the overestimation of R^2 in the climatological regressions, the pursuit of the following approach was considered: First, calculate the correlation of the two time series in question, then adjust this correlation for lag-1 autocorrelation (Wilks 2011) and square the correlation afterwards to get R^2. However, this approach was not further pursued since it only corrects for lag-1 autocorrelation when the time series in question are correlated also at higher lags and correction for autocorrelation at higher lags is not straightforward. For the annual regressions, the impacts of autocorrelation should be much smaller, although also on this time scale, there is some autocorrelation in the data used here (results not shown). In addition, we would naturally expect the residuals to be autocorrelated, since they represent internal variability, which is usually autocorrelated.

A further word of caution is appropriate with regard to the overall magnitude of the FVR values presented in this thesis. Namely, errors in predictor variables of linear regression models lead to attenuation of the R^2 value (Fuller 1987). Thus, apart from other corrections, the FVR values presented here likely underestimate the true FVR value. An exception are the MEM and GM-MEM regressions of the CESM. There, the errors in the forcing of the model can only be regarded as errors with respect to the true, real forcing. But in the model itself, the forcing prescribed acts as the true forcing, with no error. Therefore, when CESM is regressed onto its own MEM or GM-MEM, the attenuation described before should not apply.

Following from the previous remark, an interesting issue arises: It can be assumed that the quality of the forcing predictors (and therefore also the the CESM MEM and GM-MEM) improves over time both in terms of dating uncertainty and magnitude of the forcing. This corresponds to a reduction of the noise in the predictor over time. Therefore, the value of R^2 may increase over time and exhibit significant trends, even though there is no true increase in FVR. Thus, the trends shown in section 4.2 might partly, or entirely, be spurious. This applies to the forcing regressions, and to the MEM and GM-MEM regressions with the instrumental and reanalysis data, with HadCM3 and with the CFRs. For CESM, the exception discussed above applies. Since the trends are still present in the data when the high-forced industrial period is omitted (see section 4.2.3) and a natural increase of FVR over time seems unlikely, the trends likely have been caused by the increasing data quality in the predictor data sets.

For the regressions of the CFRs and the instrumental and reanalysis data sets onto the CESM MEM

and GM-MEM, the following considerations apply: First, the assumption is made that the forcing of the model (Schmidt et al. 2012; Gao et al. 2008; Vieira & Solanki 2010) is a good approximation of the true forcing. A second assumption is that the response of the CESM to its forcing in terms of response time and amplitude is similar to that of the real climate system. For volcanic forcing, this might be questionable, as was mentioned above. These limitations can explain the enhanced FVR values of CESM as compared to the CFRs in MEM and GM-MEM regressions (demonstrated in section 4.3). Interestingly, comparing the instrumental and reanalysis data sets to CESM, the difference seems to be present only for the MEM regressions, but not for GM-MEM regressions (figures 4.1b and 4.1c). For HadCM3, similar considerations apply, though here the comparison is not with the real forcing or its response, but with the those of the different model.

For the forcing and MEM regressions, the assumption was made that the forcing is globally uniform. In reality, this might not hold true, and studies have shown that the response of surface temperature to forcing is indeed spatially heterogeneous (Kelly et al. 1996; Eliseev & Mokhov 2008, see the discussion above).

Also, in this thesis the assumption was made that internal variability can be regarded as independent from external forcing. However, this assumption might not hold true, as studies have shown for GHG forcing of the North Atlantic Oscillation (NAO) (Ulbrich & Christoph 1999; Osborn 2004), GHG forcing of ENSO (Timmermann 1999; Timmermann 2001; Cai et al. 2015) and volcanic forcing of ENSO (McGregor & Timmermann 2011).

It needs also be stated that the variance fraction of internal variability cannot simply calculated by subtracting the regression R^2 values from 1. The reason for that is that the residuals from the regression (representing internal variability) might also contain a noise component. Therefore, the fraction of variance accounted for by internal variability might be smaller than 1–R^2.

The forcing cross-correlations described in section 3.2.1 were computed over the whole time of data availability at once and not over the moving windows like the forcing regressions. Therefore, a constant response of the climate system to forcing was assumed.

A final limitation concerns potential seasonal biases in the CFR data sets. Because the CFR data sets strongly rely on dendrochronological proxy records in the northern hemisphere (see figure 2.2), they are potentially biased to represent the northern hemisphere growing season. This could in part explain differences between the data sets.

6 Conclusions and Outlook

6.1 Conclusions

To wrap up, a summary of the results from this thesis is given in the following, especially with regard to the research questions and hypotheses. Recall that research questions QA1-QA3 were concerned with the GM FVR, and research questions QB1-QB3 with the spatial structure of the FVR. Furthermore, questions QA1 and QB1 were concerned with the temporal stability of FVR, questions QA2 and QB2 with its consistency across data sets and questions QA3 and QB3 with its consistency across regression methods.

Research question QA1 was concerned with the temporal stability of the GM FVR. It could be shown that FVR exhibits substantial variability in time and that this result defies the notion of a time-invariant GM FVR. One aspect of this variability, the strong rise of GM FVR in the 19^{th} and 20^{th} centuries indicate that present climate is strongly forced, most likely due to the influence of anthropogenic GHG. This strong forcing implies that hiatus periods are likely due to the superposition of internal variability signals on the strong background warming from GHG forcing. The current strong forcing, which is likely to still increase due to further anthropogenic GHG emissions, also implies that the relative importance of internal variability is likely to decrease in the future, which should reduce uncertainty in climate scenario projections.

Further, mostly upward trends in GM FVR could be detected. However, these trends likely are an artifact of improving data quality over time. Hypothesis HA1 can therefore be falsified with regard to the variability of GM FVR, but retained with regard to the trend aspect.

Research question QA2 was dedicated to the consistency of the GM FVR across the different data sets. For most combinations of data sets, significant differences could be detected. Therefore, hypothesis HA2 needs to be rejected strictly speaking. However, in most cases the overall magnitude of GM FVR falls within the same ranges for the different data sets. This broad agreement, especially between climate model and CFR data sets, can be considered a major result from this thesis.

The consistency of GM FVR between the different regression methods, addressed by research question QA3 and hypothesis HA3, needed to be rejected. Significant differences were detected between the different regression methods in all cases. However, for most of these inconsistencies, explanations could be provided.

Research question QB1 was concerned with the temporal stability of the spatial structure of the FVR. The hypothesized (HB1) stability could clearly be rejected. FVR exhibits a complex spatiotemporal evolution over the CE. This was shown on the basis of two example data sets, but was found consistent over all data sets and regression methods in this thesis. This finding can be considered another main conclusions of this thesis. It implies that there likely are no regions where FVR is persistently high or low. However, this statement should be subjected to closer scrutiny in further investigations.

Further, research question QB2 called for an investigation of the consistency of the spatial patterns of FVR with data set. In the results of PCAs of FVR fields, an exceedingly consistent pattern with globally uniform PC1 loading fields could be identified. This result indicates that forcing, on a first order, influences climate in a globally uniform way. The corresponding rise of the PC1 score time series in the industrial period indicates that this first-order globally uniform response also holds true for present climate. However, in the FVR fields themselves, the differences between the data sets were prevalent, so that the hypothesized consistency (HB2) needed to be rejected. Therefore, further research is needed (see suggestions in 6.2) to robustly identify locations of high, respectively low, FVR. Such further investigations may then also shed more light on the question of optimal locations for paleoclimate proxies. Similar considerations apply to the question of detection and attribution of regional climate change. However, it is questionable whether such locations of persistently high or low FVR can be found, in face of the variability of the spatial pattern of FVR (see the considerations made above).

Further, research question QB3 was dedicated to the consistency of the spatial structure of FVR with regression method. Some commonalities could be found, but the differences between the regression methods outweighed the commonalities overall. Therefore, hypothesis HB3 needs to be rejected in favor of inconsistency. However, for many of the differences between regression methods, explanations could be provided.

Finally, a main and rather solid conclusion from this thesis is the estimation of most robust values of GM FVR. These were found to be around 0.3–0.4 over the last millennium on the climatological time scale and small (FVR < 0.05) over the period 1401–2000 on the annual time scale.

6.2 Outlook

Further research could extend the work from this thesis in several regards. One suggestion would be to improve the regression methodology. For the regressions on the annual time scale, AR(1) processes could be taken into account using generalized least squares regression. However, the problem remains that the autocorrelation structure of the time series are different depending on the grid point and data set. Therefore, automated model selection would be appropriate to fit optimal models to each subset of the data. However, computing effort poses a limit on this solution, as automated regression model selection is computationally much more expensive than ordinary least-squares regression. Furthermore, for the regressions on the climatological time scale, a solution that does not suffer from the inflation of R^2 due to autocorrelation and reduced degrees of freedom would be a major improvement.

Another possible improvement to the regression methodology is to separate the regression calculations from the calculations of FVR. In such an approach, regression would be used to compute time series of internal variability (residuals) and forced variability (fitted values). This step could be carried out at once for the entire period of data availability for each data set. Subsequently, running variances could be computed for both the internal and forced variability time series, and their corresponding ratio defined as the FVR. This approach should be more flexible and also less computationally expensive than the approach pursued in this thesis.

A further possible extension of the present work is to split the FVR computed onto the different predictors. This approach could additionally be extended to include time series of modes of internal

variability (such as ENSO, NAO and SAM on the annual time scale and PDO, AMO on the multi-decadal time scale) as additional predictors (Wang et al. 2018). In this manner, individual fields of explained variance fraction could be computed for every predictor. Such an approach was initially planned as an extension to the present thesis, but needed to be skipped because the analysis presented here was found to be extensive enough for the given time frame.

As an alternative to the basic linear regression approach used here, regression-based detection and attribution (D&A) using total least squares (Allen & Stott 2003; Schurer et al. 2013a) could be an option in future projects. This methodology allows for an explicit quantification of uncertainty and may therefore lead to better results.

Another, rather obvious extension of the present thesis is to investigate FVR for different climatic variables. For example, precipitation or sea level pressure could be examined, but in principle, any climatic variable for which sufficient data are available could be used.

The effects of temporal scale are of general interest to the topic of FVR. In this thesis, it was decided to limit the analysis to the annual and climatological time scales, mostly due to temporal constraints of the project. Further research could focus on determining the FVR on different time scales. However, such an enterprise should pay close attention to the effects of filtering or other methodologies, to avoid the problems of autocorrelation and reduced degrees of freedom encountered in this thesis.

Similar to the consideration of different temporal scales, further research could examine the effects of different spatial scales. In this thesis, the analysis was limited to consider only the spatial scale with a regular resolution of 5° longitude by 5° latitude, due to temporal constraints and the highest resolution allowed for by the data sets present. Future work could focus on FVR at different spatial scales, examining variables in the GM, at hemispheric scales, and on smaller scales. This would shed more light on the relationship between spatial scale and FVR, and perhaps even allow for a quantification of this relationship. It is hypothesized that FVR increases with the spatial scale considered, so that the highest values of FVR should be found on the global-mean scale. At large spatial scales, such an analysis would have the additional advantage of being much less computationally and data intensive.

Bibliography

Allen, M. R. & P. A. Stott (2003). "Estimating signal amplitudes in optimal fingerprinting, part I: Theory". In: *Climate Dynamics* 21.5-6, pp. 477–491.

Barkhordarian, A., H. von Storch & E. Zorita (2012a). "Anthropogenic forcing is a plausible explanation for the observed surface specific humidity trends over the Mediterranean area". In: *Geophysical Research Letters* 39.19, pp. 1–5.

Barkhordarian, A., H. von Storch, E. Zorita & J. J. Gómez-Navarro (2016). "An attempt to deconstruct recent climate change in the Baltic Sea basin". In: *Journal of Geophysical Research* 121.22, pp. 13,207–13,217.

Barkhordarian, A. (2012). "Investigating the Influence of Anthropogenic Forcing on Observed Mean and Extreme Sea Level Pressure Trends over the Mediterranean Region". In: *The Scientific World Journal* 2012, pp. 1–16.

Barkhordarian, A., J. Bhend & H. von Storch (2012b). "Consistency of observed near surface temperature trends with climate change projections over the Mediterranean region". In: *Climate Dynamics* 38.9-10, pp. 1695–1702.

Barkhordarian, A., H. von Storch & J. Bhend (2013). "The expectation of future precipitation change over the Mediterranean region is different from what we observe". In: *Climate Dynamics* 40.1-2, pp. 225–244.

Barkhordarian, A., H. von Storch, E. Zorita, P. C. Loikith & C. R. Mechoso (2018). "Observed warming over northern South America has an anthropogenic origin". In: *Climate Dynamics* 51.5, pp. 1–14.

Bell, T. L. (1982). "Optimal Weighting of Data to Detect Climatic Change: Application to the Carbon Dioxide Problem". In: *Journal of Geophysical Research* 87.C13, pp. 11161–11170.

— (1986). "Theory of Optimal Weighting of Data to Detect Climatic Change". In: *Journal of the Atmospheric Sciences* 43.16, pp. 1694–1710.

Bjerknes, J. (1966). "A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature". In: *Tellus* 18.4, pp. 820–829.

Bradley, R. S. (1996). "Are there optimum sites for global paleotemperature reconstruction?" In: *NATO ASI Series* 141, pp. 603 –624.

Brönnimann, S. (2015). *Climatic Changes Since 1700*. Ed. by M. Beniston. Vol. 55. Cham: Springer, p. 360.

Brönnimann, S., J. Franke, S. U. Nussbaumer, H. J. Zumbühl, D. Steiner, M. Trachsel, G. C. Hegerl, A. Schurer, M. Worni, A. Malik, J. Flückiger & C. C. Raible (2019). "Last phase of the Little Ice Age forced by volcanic eruptions". In: *Nature Geoscience* 12.8, pp. 650–656.

Butterworth, S. (1930). "On the Theory of Filter Amplifiers". In: *Experimental Wireless and the Wireless Engineer*. Vol. 7, pp. 536–541.

Cai, W., A. Santoso, G. Wang, S. W. Yeh, S. I. An, K. M. Cobb, M. Collins, E. Guilyardi, F. F. Jin, J. S. Kug, M. Lengaigne, M. J. Mcphaden, K. Takahashi, A. Timmermann, G. Vecchi, M. Watanabe & L. Wu (2015). "ENSO and greenhouse warming". In: *Nature Climate Change* 5.9, pp. 849–859.

Compo, G. P., J. S. Whitaker, P. D. Sardeshmukh, N. Matsui, R. J. Allan, X. Yin, B. E. Gleason, R. S. Vose, G. Rutledge, P. Bessemoulin, S. Brönnimann, M. Brunet, R. I. Crouthamel, A. N. Grant, P. Y. Groisman, P. D. Jones, M. C. Kruk, A. C. Kruger, G. J. Marshall, M. Maugeri, H. Y. Mok, O. Nordli, T. F. Ross, R. M. Trigo, X. L. Wang, S. D. Woodruff & S. J. Worley (2010). *International Surface Pressure Databank (ISPDv2). Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. Dataset.*

Compo, G. P., J. S. Whitaker, P. D. Sardeshmukh, N Matsui, R. J. Allan, X Yin, E Gleason, R. S. Vose, G Rutledge, P Bessemoulin, S Brönnimann, M Brunet, R. I. Crouthamel, A. N. Grant, P. Y. Groisman, P. D. Jones, M. C. Kruk, A. C. Kruger, G. J. Marshall, M Maugeri, H. Y. Mok, Ø Nordli, T. F. Ross, R. M. Trigo, X. L. Wang, S. D. Woodruff & S. J. Worley (2011). "The Twentieth Century Reanalysis Project". In: *Quarterly Journal of the Royal Meteorological Society* 137, pp. 1–28.

Cowtan, K. & R. G. Way (2014). "Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends". In: *Quarterly Journal of the Royal Meteorological Society* 140, pp. 1935–1944.

Cram, T. A., G. P. Compo, X. Yin, R. J. Allan, C. McColl, R. S. Vose, J. S. Whitaker, N. Matsui, L. Ashcroft, R. Auchmann, P. Bessemoulin, T. Brandsma, P. Brohan, M. Brunet, J. Comeaux, R. Crouthamel, B. E. Gleason, P. Y. Groisman, H. Hersbach, P. D. Jones, T. Jónsson, S. Jourdain, G. Kelly, K. R. Knapp, A. Kruger, H. Kubota, G. Lentini, A. Lorrey, N. Lott, S. J. Lubker, J. Luterbacher, G. J. Marshall, M. Maugeri, C. J. Mock, H. Y. Mok, Ø. Nordli, M. J. Rodwell, T. F. Ross, D. Schuster, L. Srnec, M. A. Valente, Z. Vizi, X. L. Wang, N. Westcott, J. S. Woollen & S. J. Worley (2015). "The International Surface Pressure Databank version 2". In: *Geoscience Data Journal* 2.1, pp. 31–46.

Cressie, N. (1990). "The Origins of Kriging". In: *Mathematical Geology* 22.3, pp. 239–252.

Crowley, T. J. & M. B. Unterman (2013). "Technical details concerning development of a 1200 yr proxy index for global volcanism". In: *Earth System Science Data* 5.1, pp. 187–197.

Crowley, T. J., S. P. Obrochta & J. Liu (2014). "Recent global temperature "plateau" in the context of a new proxy reconstruction". In: *Earth's Future* 2, pp. 281–294.

Dai, A., J. C. Fyfe, S. P. Xie & X. Dai (2015). "Decadal modulation of global surface temperature by internal climate variability". In: *Nature Climate Change* 5, pp. 555–560.

Davis, R. E. (1978). "Predictability of Sea Level Pressure Anomalies over the North Pacific Ocean". In: *Journal of Physical Oceanography* 8, pp. 233 –246.

Deser, C., A. Phillips, V. Bourdette & H. Teng (2012). "Uncertainty in climate change projections: The role of internal variability". In: *Climate Dynamics* 38.3-4, pp. 527–546.

Driscoll, S., A. Bozzo, L. J. Gray, A. Robock & G. Stenchikov (2012). "Coupled Model Intercomparison Project 5 (CMIP5) simulations of climate following volcanic eruptions". In: *Journal of Geophysical Research Atmospheres* 117.17.

Easterling, D. R. & M. F. Wehner (2009). "Is the climate warming or cooling?" In: *Geophysical Research Letters* 36.8, pp. 4–6.

Eliseev, A. V. & I. I. Mokhov (2008). "Influence of volcanic activity on climate change in the past several centuries: Assessments with a climate model of intermediate complexity". In: *Izvestiya - Atmospheric and Ocean Physics* 44.6, pp. 671–683.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer & K. E. Taylor (2016). "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization". In: *Geoscientific Model Development* 9.5, pp. 1937–1958.

Faden, M. (2018). "The Stationarity of the Atlantic Multi-Decadal Oscillation and its Teleconnections over the Last Millennium". Master Thesis. University of Bern.

Frankcombe, L. M., M. H. England, M. E. Mann & B. A. Steinman (2015). "Separating internal variability from the externally forced climate response". In: *Journal of Climate* 28.20, pp. 8184–8202.

Fuller, W. A. (1987). *Measurement Error Models*. Ames, Iowa: John Wiley & Sons, pp. 1–440.

Gao, C., A. Robock & C. Ammann (2008). "Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models". In: *Journal of Geophysical Research* 113.23, pp. 1–15.

Gent, P. R., G. Danabasoglu, L. J. Donner, M. M. Holland, E. C. Hunke, S. R. Jayne, D. M. Lawrence, R. B. Neale, P. J. Rasch, M. Vertenstein, P. H. Worley, Z. L. Yang & M. Zhang (2011). "The community climate system model version 4". In: *Journal of Climate* 24.19, pp. 4973–4991.

Gillett, N. P., D. A. Stone, P. A. Stott, T. Nozawa, A. Y. Karpechko, G. C. Hegerl, M. F. Wehner & P. D. Jones (2008). "Attribution of polar warming to human influence". In: *Nature Geoscience* 1.11, pp. 750–754.

Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell & R. A. Wood (2000). "The simulation fo SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments". In: *Climate Dynamics* 16, pp. 147–168.

Gujarati, D. N. & D. C. Porter (2010). *Essentials of Econometrics*. Fourth Edi. New York: McGraw-Hill Irwin, pp. 1–554.

Harzallah, A. & R. Sadourny (1995). "Internal versus SST-forced atmospheric variability as simulated by an atmospheric general circulation model". In: *Journal of Climate* 8.3, pp. 474–495.

Hasselmann, K. (1979). "On the signal-to-noise problem in atmospheric response studies". In: *Meteorology over the tropical oceans*. Ed. by D. Shaw. London: Royal Meteorological Society, pp. 251–259.

— (1993). "Optimal fingerprints for the detection of time- dependent climate change". In: *Journal of Climate* 6.10, pp. 1957–1971.

Hawkins, E. & R. Sutton (2009). "The potential to narrow uncertainty in regional climate predictions". In: *Bulletin of the American Meteorological Society* 90.8, pp. 1095–1107.

— (2011). "The potential to narrow uncertainty in projections of regional precipitation change". In: *Climate Dynamics* 37.1, pp. 407–418.

Hawkins, E., B. Anderson, N. Diffenbaugh, I. Mahlstein, R. Betts, G. Hegerl, M. Joshi, R. Knutti, D. McNeall, S. Solomon, R. Sutton, J. Syktus & G. Vecchi (2014). "Uncertainties in the timing of unprecedented climates". In: *Nature* 511.7507, E3–E6.

Hegerl, G. & F. Zwiers (2011). "Use of models in detection and attribution of climate change". In: *Wiley Interdisciplinary Reviews: Climate Change* 2.4, pp. 570–591.

Hegerl, G. C. & G. R. North (1997). "Comparison of statistically optimal approaches to detecting anthropogenic climate change". In: *Journal of Climate* 10.5, pp. 1125–1133.

Hu, Z. Z., A. Kumar, B. Jha & B. Huang (2012). "An analysis of forced and internal variability in a warmer climate in CCSM3". In: *Journal of Climate* 25.7, pp. 2356–2373.

Hurrell, J. W., M. M. Holland, P. R. Gent, S. Ghan, J. E. Kay, P. J. Kushner, J. F. Lamarque, W. G. Large, D. Lawrence, K. Lindsay, W. H. Lipscomb, M. C. Long, N. Mahowald, D. R. Marsh, R. B. Neale, P. Rasch, S. Vavrus, M. Vertenstein, D. Bader, W. D. Collins, J. J. Hack, J. Kiehl & S. Marshall (2013). "The community earth system model: A framework for collaborative research". In: *Bulletin of the American Meteorological Society* 94.9, pp. 1339–1360.

Hyndman, R. J. & G. Athanasopoulos (2018). *Forecasting: principles and practice*. 2nd Editio. Melbourne: OTexts.

Hyndman, R. J. & Y. Khandakar (2008). "Automatic Time Series Forecasting: The forecast Package for R". In: *Journal of Statistical Software* 27.3, pp. 1–22.

Imdadullah, M., M. Aslam & S. Altaf (2016). "mctest: An R Package for Detection of Collinearity among Regressors". In: *The R Journal* 8.2, p. 495.

IPCC (2013). *Climate Change 2013: The Physical Science Basis*. Ed. by T. F. Stocker & Q. Dahe. New York: Cambridge University Press, pp. 1 –1535.

Jones, P. D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon & C. P. Morice (2012). "Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010". In: *Journal of Geophysical Research Atmospheres* 117.5.

Jungclaus, J. H., E. Bard, M. Baroni, P. Braconnot, J. Cao, L. P. Chini, T. Egorova, M. Evans, J. Fidel González-Rouco, H. Goosse, G. C. Hurtt, F. Joos, J. O. Kaplan, M. Khodri, K. Klein Goldewijk, N. Krivova, A. N. Legrande, S. J. Lorenz, J. Luterbacher, W. Man, A. C. Maycock, M. Meinshausen, A. Moberg, R. Muscheler, C. Nehrbass-Ahles, B. I. Otto-Bliesner, S. J. Phipps, J. Pongratz, E. Rozanov, G. A. Schmidt, H. Schmidt, W. Schmutz, A. Schurer, A. I. Shapiro, M. Sigl, J. E. Smerdon, S. K. Solanki, C. Timmreck, M. Toohey, I. G. Usoskin, S. Wagner, C. J. Wu, K. Leng Yeo, D. Zanchettin, Q. Zhang & E. Zorita (2017). "The PMIP4 contribution to CMIP6 - Part 3: The last millennium, scientific objective, and experimental design for the PMIP4 past1000 simulations". In: *Geoscientific Model Development* 10.11, pp. 4005–4033.

Kanamitsu, M., J. Alpert, K. Campana, P. Caplan, D. Deaven, M. Iredell, B. Katz, H.-L. Pan, J. Sela & G. White (1991). "Recent Changes Implemented into the Global Forecast System at NMC". In: *Weather and Forecasting* 6.3, pp. 425–435.

Kelly, P. M., P. D. Jones & J. Pengqun (1996). "The spatial response of the climate system to explosive volcanic eruptions". In: *International Journal of Climatology* 16.5, pp. 537–550.

Kendall, M. G. (1938). "A New Measure of Rank Correlation". In: *Biometrika* 30.1/2, pp. 81–93.

— (1975). *Rank Correlation Methods*. 4th Editio. London: Charles Griffin, pp. 1–202.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker & M. Saunby (2011a). "Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties". In: *Journal of Geophysical Research Atmospheres* 116.14, pp. 1–13.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker & M. Saunby (2011b). "Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization". In: *Journal of Geophysical Research Atmospheres* 116.14, pp. 1–22.

Kosaka, Y. & S.-P. Xie (2013). "Recent global-warming hiatus tied to equatorial Pacific surface cooling". In: *Nature* 501, pp. 403–407.

— (2016). "The tropical Pacific as a key pacemaker of the variable rates of global warming". In: *Nature Geoscience* 9.9, pp. 669–673.

Lavigne, F., J.-P. Degeai, J.-C. Komorowski, S. Guillet, V. Robert, P. Lahitte, C. Oppenheimer, M. Stoffel, C. M. Vidal, Surono, I. Pratomo, P. Wassmer, I. Hajdas, D. S. Hadmoko & E. de Belizal (2013). "Source of the great A.D. 1257 mystery eruption unveiled, Samalas volcano, Rinjani Volcanic Complex, Indonesia". In: *Proceedings of the National Academy of Sciences* 110.42, pp. 16742–16747.

Lean, J. L. (2018). "Observation-based detection and attribution of 21st century climate change". In: *Wiley Interdisciplinary Reviews: Climate Change* 9.2, pp. 1–20.

Lorenz, E. N. (1956). *Empirical Orthogonal Functions and Statistical Weather Prediction*. Tech. rep. Cambridge, Massachusetts: Massachusetts Institute of Technology, Department of Meteorology, p. 52.

Lyu, K., X. Zhang, J. A. Church & J. Hu (2015). "Quantifying internally generated and externally forced climate signals at regional scales in CMIP5 models". In: *Geophysical Research Letters* 42.21, pp. 9394–9403.

Mahlstein, I., R. Knutti, S. Solomon & R. W. Portmann (2011). "Early onset of significant local warming in low latitude countries". In: *Environmental Research Letters* 6.3, pp. 1–6.

Mann, H. B. (1945). "Nonparametric Tests Against Trend". In: *Econometrica* 13.3, pp. 245–259.

Matsumura, S., G. Huang, S.-P. Xie & K. Yamazaki (2010). "SST-Forced and Internal Variability of the Atmosphere in an Ensemble GCM Simulation". In: *Journal of the Meteorological Society of Japan* 88.1, pp. 43 –62.

McGregor, S. & A. Timmermann (2011). "The effect of explosive tropical volcanism on ENSO". In: *Journal of Climate* 24.8, pp. 2178–2191.

McPhaden, M. J., S. E. Zebiak & M. H. Glantz (2006). "ENSO as an Integrating Concept in Earth Science". In: *Science* 314, pp. 1740–1745.

Meinshausen, M., E. Vogel, A. Nauels, K. Lorbacher, N. Meinshausen, D. M. Etheridge, P. J. Fraser, S. A. Montzka, P. J. Rayner, C. M. Trudinger, P. B. Krummel, U. Beyerle, J. G. Canadell, J. S. Daniel, I. G. Enting, R. M. Law, C. R. Lunder, S. O'Doherty, R. G. Prinn, S. Reimann, M. Rubino, G. J. Velders, M. K. Vollmer, R. H. Wang & R. Weiss (2017). "Historical greenhouse gas concentrations for climate modelling (CMIP6)". In: *Geoscientific Model Development* 10.5, pp. 2057–2116.

Miller, G. H., Á. Geirsdóttir, Y. Zhong, D. J. Larsen, B. L. Otto-Bliesner, M. M. Holland, D. A. Bailey, K. A. Refsnider, S. J. Lehman, J. R. Southon, C. Anderson, H. Björnsson & T. Thordarson (2012). "Abrupt onset of the Little Ice Age triggered by volcanism and sustained by sea-ice/ocean feedbacks". In: *Geophysical Research Letters* 39.2, pp. 1–5.

Moorthi, S., H.-L. Pan & P. Caplan (2001). *Changes to the 2001 NCEP Operational MRF/AVN Global Analysis/Forecast System*. Tech. rep., p. 14.

Morice, C. P., J. J. Kennedy, N. A. Rayner & P. D. Jones (2012). "Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set". In: *Journal of Geophysical Research Atmospheres* 117.8, pp. 1–22.

Mukherjee, C., H. White & M. Wuyts (1998). *Econometrics and Data Analysis for Developing Countries*. 1st Editio. Abingdon: Routledge, pp. 1–495.

Neukom, R., N. Steiger, J. J. Gómez-navarro, J. Wang & J. Werner (2019). "No evidence for globally coherent warm and cold periods over the pre-industrial Common Era". In: *Nature* 571, pp. 550 –554.

North, G. R. & K. Y. Kim (1995). "Detection of forced climate signals. Part II: simulation results". In: *Journal of Climate* 8.3, pp. 409–417.

North, G. R., K.-Y. Kim, S. S. Shen & J. W. Hardin (1995). "Detection of forced climate signals. Part I: filter theory". In: *Journal of Climate* 8.3, pp. 401–408.

Obukhov, A. (1947). "Statistically homogeneous fields on a sphere". In: *Uspethi Mathematicheskikh Nauk* 2, pp. 196 –198.

Osborn, T. J. (2004). "Simulating the winter North Atlantic Oscillation: The roles of internal variability and greenhouse gas forcing". In: *Climate Dynamics* 22.6-7, pp. 605–623.

Otto-Bliesner, B. L., E. C. Brady, J. Fasullo, A. Jahn, L. Landrum, S. Stevenson, N. Rosenbloom, A. Mai & G. Strand (2016). "Climate Variability and Change Since 850 CE: an Ensemble Approach with the Community Earth System Model". In: *Bulletin of the American Meteorological Society* 97.5, pp. 735–754.

PAGES 2k Consortium (2017). "A global multiproxy database for temperature reconstructions of the Common Era". In: *Scientific Data* 4, pp. 1–33.

— (2019). "Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era". In: *Nature Geoscience* 12.8, pp. 643–649.

Pearson, K. (1895). "Note on Regression and Inheritance in the Case of Two Parents". In: *Proceedings of the Royal Society of London* 58, pp. 240–242.

Pope, V. D., M. L. Gallani, P. R. Rowntree & R. A. Stratton (2000). "The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3". In: *Climate Dynamics* 16.2-3, pp. 123–146.

Rowell, D. P., C. K. Folland, K. Maskell & M. N. Ward (1995). "Variability of summer rainfall over tropical north Africa (1906–92): Observations and modelling". In: *Quarterly Journal of the Royal Meteorological Society* 121.523, pp. 669–704.

Saha, S., S. Nadiga, C. Thiaw, J. Wang, W. Wang, Q. Zhang, H. M. Van den Dool, H. L. Pan, S. Moorthi, D. Behringer, D. Stokes, M. Peña, S. Lord, G. White, W. Ebisuzaki, P. Peng & P. Xie (2006). "The NCEP Climate Forecast System". In: *Journal of Climate* 19.15, pp. 3483–3517.

Satterthwaite, F. E. (1946). "An Approximate Distribution of Estimates of Variance Components". In: *Biometrics Bulletin* 2.6, pp. 110–114.

Schmidt, G. A., J. H. Jungclaus, C. M. Ammann, E Bard, P Braconnot, T. J. Crowley, G Delaygue & F Joos (2011). "Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0)". In: *Geoscientific Model Development* 4.1, pp. 33–45.

Schmidt, G. A., J. H. Jungclaus, C. M. Ammann, E. Bard, P. Braconnot, T. J. Crowley, G. Delaygue, F. Joos, N. A. Krivova, R. Muscheler, B. L. Otto-Bliesner, J. Pongratz, D. T. Shindell, S. K.

Solanki, F. Steinhilber & L. E. A. Vieira (2012). “Climate forcing reconstructions for use in PMIP simulations of the Last Millennium (v1.1)”. In: *Geoscientific Model Development* 5.1, pp. 185–191.

Schulzweida, U. (2018). *Climate Data Operator (CDO) User Guide (Version 1.9.5)*. Hamburg.

Schurer, A. P., G. C. Hegerl, M. E. Mann, S. F. Tett & S. J. Phipps (2013a). “Separating forced from chaotic climate variability over the past millennium”. In: *Journal of Climate* 26.18, pp. 6954–6973.

Schurer, A. P., S. F. B. Tett & G. C. Hegerl (2013b). “Small influence of solar variability on climate over the past millennium”. In: *Nature Geoscience* 7.2, pp. 104–108.

Sheather, S. J. (2009). *A Modern Approach to Regression with R*. Ed. by G. Casella, S. Fienberg & I. Olkin. New York: Springer, pp. 1–392.

Shindell, D. T., G. A. Schmidt, R. L. Miller & M. E. Mann (2003). “Volcanic and solar forcing of climate change during the preindustrial era”. In: *Journal of Climate* 16.24, pp. 4094–4107.

Sigl, M., M. Winstrup, J. R. McConnell, K. C. Welten, G. Plunkett, F. Ludlow, U. Büntgen, M. Caffee, N. Chellman, D. Dahl-Jensen, H. Fischer, S. Kipfstuhl, C. Kostick, O. J. Maselli, F. Mekhaldi, R. Mulvaney, R. Muscheler, D. R. Pasteris, J. R. Pilcher, M. Salzer, S. Schüpbach, J. P. Steffensen, B. M. Vinther & T. E. Woodruff (2015). “Timing and climate forcing of volcanic eruptions for the past 2,500 years”. In: *Nature* 523.7562, pp. 543–549.

Steinhilber, F., J. Beer & C. Fröhlich (2009). “Total solar irradiance during the Holocene”. In: *Geophysical Research Letters* 36.19, pp. 1–5.

Steinman, B. A., M. E. Mann & S. K. Miller (2015). “Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures”. In: *Science* 347.6225, pp. 988–991.

Storch, H. von & G. Hannoschöck (1984). “Comments on ”empirical orthogonal function analysis of wind vectors over the tropical pacific region””. In: *Bulletin of the American Meteorological Society* 65, p. 162.

Stott, P. A. & S. F. Tett (1998). “Scale-dependent detection of climate change”. In: *Journal of Climate* 11.12, pp. 3282–3294.

Stott, P. A., N. P. Gillett, G. C. Hegerl, D. J. Karoly, D. A. Stone, X. Zhang & F. Zwiers (2010). “Detection and attribution of climate change: A regional perspective”. In: *Wiley Interdisciplinary Reviews: Climate Change* 1.2, pp. 192–211.

Tett, S. F., R. Betts, T. J. Crowley, J. Gregory, T. C. Johns, A. Jones, T. J. Osborn, E. Öström, D. L. Roberts & M. J. Woodage (2007). “The impact of natural and anthropogenic forcings on climate and hydrology since 1550”. In: *Climate Dynamics* 28.1, pp. 3–34.

Thompson, D. W., J. M. Wallace, P. D. Jones & J. J. Kennedy (2009). “Identifying signatures of natural climate variability in time series of global-mean surface temperature: Methodology and insights”. In: *Journal of Climate* 22.22, pp. 6120–6141.

Timmermann, A. (1999). “Detecting the Nonstationary Response of ENSO to Greenhouse Warming”. In: *Journal of the Atmospheric Sciences* 56.14, pp. 2313–2325.

— (2001). “Changes of ENSO stability due to greenhouse warming”. In: *Geophysical Research Letters* 28.10, pp. 2061–2064.

Ting, M., Y. Kushnir, R. Seager & C. Li (2009). “Forced and internal twentieth-century SST trends in the North Atlantic”. In: *Journal of Climate* 22.6, pp. 1469–1481.

Toohey, M. & M. Sigl (2017). “Volcanic stratospheric sulfur injections and aerosol optical depth from 500BCE to 1900CE”. In: *Earth System Science Data* 9.2, pp. 809–831.

Trenberth, K. E. (2015). “Has there been a hiatus?” In: *Science* 349.6249, pp. 691–692.

Ulbrich, U. & M. Christoph (1999). “A shift of the NAO and increasing storm track activity over Europe due to anthropogenic greenhouse gas forcing”. In: *Climate Dynamics* 15.7, pp. 551–559.

Vieira, L. E. A. & S. Solanki (2010). “Evolution of the solar magnetic flux on time scales of years to millenia”. In: *Astronomy & Astrophysics* 100, pp. 1–13.

Vieira, L. E., S. K. Solanki, N. A. Krivova & I. Usoskin (2011). “Evolution of the solar irradiance during the Holocene”. In: *Astronomy and Astrophysics* 531, pp. 1–20.

Wang, J., B. Yang, F. C. Ljungqvist, J. Luterbacher, T. J. Osborn, K. R. Briffa & E. Zorita (2017). “Internal and external forcing of multidecadal Atlantic climate variability over the past 1,200 years”. In: *Nature Geoscience* 10.7, pp. 512–517.

Wang, J., B. Yang, T. J. Osborn, F. C. Ljungqvist, H. Zhang & J. Luterbacher (2018). “Causes of East Asian Temperature Multidecadal Variability Since 850 CE”. In: *Geophysical Research Letters* 45.24, pp. 13,485–13,494.

Wang, M., J. E. Overland, V. Kattsov, J. E. Walsh, X. Zhang & T. Pavlova (2007). “Intrinsic versus forced variation in coupled climate model simulations over the Arctic during the twentieth century”. In: *Journal of Climate* 20.6, pp. 1093–1107.

Welch, B. L. (1947). “The Generalization of ”Student’s” Problem when Several Different Population Variances are Involved”. In: *Biometrika* 34.1/2, pp. 28–35.

Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences*. Ed. by R. Dmowska, D. Hartmann & H. T. Rossby. Third Edit. Oxford: Academic Press (Elsevier).

Wills, R. C., T. Schneider, J. M. Wallace, D. S. Battisti & D. L. Hartmann (2018). “Disentangling Global Warming, Multidecadal Variability, and El Niño in Pacific Temperatures”. In: *Geophysical Research Letters* 45.5, pp. 2487–2496.

Acknowledgements

First, I want to thank Prof Dr Martin Grosjean for his support and motivation and for providing great overview in all major decisions of the project. A big thank you goes to Dr Raphael Neukom, my main supervisor for this thesis. The lively discussions with him never failed to inspire my interest in the climate sciences. Also, he kept me from going off on tangents many a time and was always available to discuss my questions. I also want to thank Dr Jörg Franke, the co-supervisor of this thesis. Although not being strongly involved, he was always available when I had questions and provided great support when needed. Another big thank you goes to Michael Faden, my predecessor in this master's thesis position. He always took time to answer my questions, even after having left the research group. Moreover, he provided me with the necessary data and background knowledge. Furthermore, I want to thank Dr Christoph Frei (MeteoSwiss, ETH) for providing the R package "pcaXcca" that I used to compute the PCAs of the FVR data. My thanks also go to Prof Dr Stefan Brönnimann, for providing valuable insights during the project. He pointed out limitations regarding the filtering of the time series and the likely spurious nature of the trends found.

Additionally, I want to thank the members of the paleolimnology research group at Erlachstrasse 9a in Bern for their support and for providing a most friendly environment and framework for this thesis. Namely, I want to thank Dr Aurea Chiaia-Hernández, Dr Christoph Dätwyler, Dr Daniela Fischer, Stamatina Makri, Linus Rösler, Andrea Sanchini, Dr Tobias Schneider, Lu Yao Tu, Giulia Wienhues and Paul Zander.

Finally, I want to thank my family and especially my parents for their support during this year of my master's thesis. They enabled my stay in Bern and thus greatly improved my opportunity to work on this master's thesis.

Appendix

Derivation of the Correction for Ensemble Size in MEM and GM-MEM Regressions

In the following, the derivation of the correction of FVR for ensemble size used in the MEM and GM-MEM regressions is given. Let V_I denote internal variability and V_F forced variability. Further let N denote the ensemble size and R the ratio of forced to total variability (i.e. FVR). Additionally, let the asterisk (*) mark variables not corrected for ensemble size and the variables without asterisk denote the corrected ones.

From Lyu et al. 2015 we have:

$$V_F = V_F^* - \frac{1}{N-1} V_I^* \tag{6.1}$$

$$V_I = \frac{N}{N-1} V_I^* \tag{6.2}$$

$$R = \frac{V_F}{V_F + V_I} \tag{6.3}$$

Inserting 6.1 and 6.2 into 6.3 and simplifying yields:

$$R = \frac{V_F}{V_F + V_I} = \frac{V_F^* - \frac{1}{N-1} V_I^*}{V_F^* - \frac{1}{N-1} V_I^* + \frac{N}{N-1} V_I^*} = \frac{V_F^*}{V_F^* + V_I^*} - \frac{1}{N-1} \frac{V_I^*}{V_F^* + V_I^*} \tag{6.4}$$

Furthermore, we can define the identity of R^*:

$$R^* = \frac{V_F^*}{V_F^* + V_I^*} \tag{6.5}$$

Then, we can make the following derivation:

$$1 = \frac{V_F^* + V_I^*}{V_F^* + V_I^*} = \frac{V_F^*}{V_F^* + V_I^*} + \frac{V_I^*}{V_F^* + V_I^*} \tag{6.6}$$

Inserting 6.5 into 6.6, we get:

$$1 = R^* + \frac{V_I^*}{V_F^* + V_I^*} \Leftrightarrow 1 - R^* = \frac{V_I^*}{V_F^* + V_I^*} \tag{6.7}$$

Finally, we can insert 6.5 and 6.7 into 6.4 and simplify to get:

$$R = R^* - \frac{1}{N-1}(1 - R^*) = \frac{NR^* - 1}{N-1} \qquad \square$$

Declaration of consent

on the basis of Article 30 of the RSL Phil.-nat. 18

Name/First Name: Frey, Louis

Registration Number: 12-919-650

Study program:

Bachelor ☐ Master ☑ Dissertation ☐

Title of the thesis: The Spatiotemporal Evolution of the Temperature Forced Variance Ratio over the Common Era

Supervisor: Dr Raphael Neukom
Prof Dr Martin Grosjean
Dr Jörg Franke

I declare herewith that this thesis is my own work and that I have not used any sources other than those stated. I have indicated the adoption of quotations as well as thoughts taken from other authors as such in the thesis. I am aware that the Senate pursuant to Article 36 paragraph 1 litera r of the University Act of 5 September, 1996 is authorized to revoke the title awarded on the basis of this thesis.

For the purposes of evaluation and verification of compliance with the declaration of originality and the regulations governing plagiarism, I hereby grant the University of Bern the right to process my personal data and to perform the acts of use this requires, in particular, to reproduce the written thesis and to store it permanently in a database, and to use said database, or to make said database available, to enable comparison with future theses submitted by others.

Place/Date

Stäfa, 6.12.2019

Signature