Universidade de Aveiro Departamento de Física
2008

Paulo de Melo Gonçalves

Forçamento da Atmosfera Euro-Atlântica pelo Oceano Tropical

Tropical Ocean Forcing of the Euro-Atlantic Atmosphere

Universidade de Aveiro Departamento de Física
2008

Paulo de Melo Gonçalves

Forçamento da Atmosfera Euro-Atlântica pelo Oceano Tropical

Tropical Oceanic Forcing of the Euro-Atlantic Atmosphere

dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Física, realizada sob a orientação científica do Prof. Doutor Alfredo Moreira Caseiro Rocha, Professor Associado do Departamento de Física da Universidade de Aveiro

Ciência.Inovação 2010 Programa Operacional Ciência e Inovação 2010
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

o júri

presidente

Professor Doutor Manuel João Senos Matias
professor catedrático da Universidade de Aveiro

Doutor João Alexandre Medina Corte-Real
professor catedrático da Universidade de Évora

Prof. Doutora Maria Solange Mendonça Leite
professor associado com Agregação da Universidade de Trás-os-Montes e Alto Douro

Doutor António Pedro Viterbo de Sousa Azevedo
investigador do Instituto de Meteorologia

Prof. Doutor Alfredo Moreira Caseiro Rocha (Orientador)
professor associado da Universidade de Aveiro

Prof. Doutor José Manuel Henriques Castanheira
professor auxiliar da Universidade de Aveiro

agradecimentos

Ao Prof. Doutor Alfredo Moreira Caseiro Rocha agradeço o convite para a realização deste trabalho, o constante acompanhamento e apoio científico. Agradeço também a realização das simulações numéricas, sem as quais este estudo não seria possível.
Ao Prof. Doutor José Manuel Henriques Castanheira agradeço a permanente disponibilidade, as frutíferas discussões e preciosas sugestões.
Agradeço ao Álvaro Peliz os sábios conselhos que me foram guiando ao longo deste percurso.
Aos colegas Carlos Fernandes, Juan Ferreira, Nuno Vaz, Martinho Almeida, Paulo e Pedro Ferreira agradeço a cumplicidade e o companheirismo.
Esta investigação foi financiada por uma bolsa de assistente de investigação da Universidade de Aveiro e por uma bolsa de doutoramento da Fundação para a Ciência e Tecnologia (SFRH/BD/17825/2004).

palavras-chave AMIP II, MUGCM, SST, MSLP, variabilidade inter-anual

resumo

Um grande ensemble de simulações AMIP (Model Intercomparison Project) II geradas pelo MUGCM (Melbourne University General Circulation Model) são usadas para estudar o forçamento da pressão média ao nível das águas do mar (MSLP, mean sea level pressure) na região Euro-Atlântica (EA) pelas anomalias da temperatura superficial do oceano (SST, sea level temperature).

Uma Análise de Variância (ANOVA) mostra que a variabilidade de médias sazonais de MSLP na região EA, e o seu maior modo de variabilidade - a Oscilação do Atlântico Norte (NAO, North Atlantic Oscillation) – são significativamente forçadas pelas SSTs no Inverno e na Primavera.

Os dois primeiros modos de variabilidade forçada das anomalias sazonais de MSLP na região EA são estimados usando uma Análise de Componentes Principais de Detecção Optimizada. Análises de regressão e correlação usando anomalias sazonais de SST e as séries temporais associadas aos padrões forçados fornecem evidência estatística de que: (i) uma fase positiva/negativa do El Niño - Oscilação Austral (ENSO) induz uma fase negativa/positiva da NÃO no Inverno e na Primavera; (ii) uma fase positiva/negativa do Gradiente Inter-hemisférico de SST no Atântico induz uma fase negativa/positiva da NAO.

A sensibilidade da NAO à polaridade e intensidade do ENSO é também analisada. Os resultados revelam sinais das fases do ENSO quer na intensidade média da NAO, quer na sua variabilidade interna. Durante a fase fria do ENSO, a Função de Densidade de Probabilidade (PDF) do índice da NAO evidencia um pequeno mas positivo valor médio, enquanto que este é negativo na fase quente do ENSO. Além disso, a variabilidade da NAO tem um comportamento diferente para cada fase do ENSO: durante a fase quente, a PDF apresenta maior variância e sugere uma bimodalidade, enquanto que na fase fria nenhuma bimodalidade é sugerida.

keywords AMIP II, MUGCM, SST, MSLP, interannual variability

abstract

A large Atmospheric Model Intercomparison Project II ensemble of the Melbourne University General Circulation Model is used to asses sea surface temperature (SST) anomaly forcing of mean sea level pressure (MSLP) seasonal mean anomalies over the Euro-Atlantic (EA) region. Tropical SST forcing is focused.

An Analysis of Variance shows that seasonal mean MSLP variability in the EA region, and its major mode of variability (the North Atlantic Oscillation, NAO), are significantly SST-forced in winter and spring.

The two leading SST-forced variability modes of MSLP seasonal mean anomalies in the EA region are estimated, using Optimal Detection Principal Component Analysis. Regression and correlation analysis using SST anomalies and the time series associated to the forced patterns, give statistical evidence that: (i) a warm (cold) phase of the El Niño-Southern Oscillation (ENSO) induces a negative (positive) phase of the NAO in winter and spring; and (ii) a positive (negative) phase of the Atlantic Inter-hemispheric SST Gradient induces a negative (positive) phase of the NAO in spring.

The sensitivity of the NAO to ENSO polarity and strength is also analysed. The results show signals of the ENSO phases in both the mean strength of the NAO as well as in its internal variability. During the cold ENSO phase, the Probability Density Function (PDF) of the NAO index presents a small but positive mean value, whereas it is negative during the warm ENSO phase. Also, the NAO variability associated with each ENSO phase shows a different behaviour: during the warm phase, the PDF presents a larger variance and suggests bimodality, whereas no bimodality is suggested in the cold ENSO phase.

Contents

List of Figures

List of Tables

List of Acronyms

ACYC	Annual CYCle
AGCM	Atmospheric General Circulation Model
AISG	Atlantic Interhemispheric SST anomaly Gradient
AMIP	Atmospheric Model Intercomparison Project
ANOVA	Analysis of Variance
CMAP	Climate Prediction Center (CPC) Merged Analysis of Precipitation
DIV WIND	divergent wind
DJF	December, January and February
EA	Euro-Atlantic
EATL	East Atlantic
ENSO	El Niño-Southern Oscillation
EOF	Empirical Orthogonal Function
EPAC	East Pacific
GOGA	Global Ocean Global Atmosphere
HGT	geopotential height

IND Indonesia
I-POGA Indo-Pacific Ocean Global Atmosphere

JJA June, July and August

LHF latent heat flux

MAM March, April and May
MOGA Midlatitude Ocean Global Atmosphere
MSLP mean sea level pressure
MUGCM Melbourne University General Circulation Model

NAO North Atlantic Oscillation
NCEP National Centers for Environmental Prediction
NPA North Pacific-American

PC Principal Component
PCA Principal Component Analysis
PDF Probability Density Function
PNA Pacific North American
POGA Pacific Ocean Global Atmosphere
PRECIP precipitation

SIC sea ice concentration
SON September, October and November
SST sea surface temperature

SVDA	Singular Value Decomposition Analysis
TAGA	Topical Atlantic Global Atmosphere
TMP	temperature
TMP-SFC	surface temperature
TOGA	Tropical Ocean Global Atmosphere
SST Tripole	North Atlantic SST anomaly Tripole
UWIND	zonal wind
VEL POT	velocity potential
VVEL	vertical velocity
WATL	West Atlantic
WPAC	West Pacific

Chapter 1

Introduction

Lorenz [1963, 1965] demonstrated that, due to the inherent nature of instability and nonlinearity, atmospheric flows with only slightly different initial states will depart from each other and evolve eventually to flows that are just randomly related. Because of this, daily weather variations, which are due primarily to the internal dynamics of the atmosphere, cannot be predicted in detail by more than 2-3 weeks in advance [Lorenz, 1982; Chen, 1989]. In other words, *synoptic predictability* is an *initial condition problem* bounded by the time the atmosphere can "remember" its initial state.

Weather variability induces unpredictable variability on interannual variability of seasonal mean quantities [Leith, 1973; Madden, 1976; Chervin, 1986], usually referred to as *natural* or *internal variability*. On the other hand, slowly varying *external* boundary conditions, such as anomalies of sea surface temperature (SST) and sea ice concentration (SIC) can cause predictable variations in seasonal mean quantities, referred to as *external* or *forced variability*. Consequently, interannual climate variability is formed by two components: a component due to the internal dynamics of the atmosphere (internal variability) and a component forced by the slowly varying anomalous boundary conditions external to the atmospheric climate system (external variability). Since the internal component is random and the external component is predictable, *interannual climate predictability* is a *boundary condition problem* [Chervin, 1986; Branković et al., 1994; Chen and Van den Dool, 1997].

As an external boundary condition problem, interannual climate predictability requires, in the first place, the prediction of the external boundary evolution. Because of this, it is usually referred to as *interannual climate potential predictability*. Estimation of potential predictability in the globe has been a topic of ongoing research in the climate community. This estimation requires the estimation of at least one of the two components of climate variability, internal or external variability, and is, therefore, difficult to achieve using the observational record alone [Madden, 1976]. This difficulty has led the climate community to use Atmospheric General Circulation Model (AGCM) simulations to estimate potential predictability. Several experiments have been conducted with AGCMs to achieve this task [Chervin and Schneider, 1976; Chervin, 1986] but the one which has proven to be more suitable consists of performing an ensemble of integrations all forced by the same observed boundary conditions (SST and SIC) but started from different initial conditions [Dix and Hunt, 1995; Stern and Miyakoda, 1995; Harzallah and Sadourny, 1995; Kumar and Hoerling, 1995; Kumar et al., 1996], such as the experiments following the rules of the Atmospheric Model Intercomparison Project (AMIP) and AMIP II [Gates, 1992]. The philosophy is that sensitivity to initial atmospheric conditions can be used to quantify internal variability whereas the relative similarity between ensemble members can be used to quantify external variability. Since the works of Rowell et al. [1995]; Zwiers [1996]; Davis et al. [1997]; Rowell [1998] and Wang and Zwiers [1999], the Analysis of Variance (ANOVA) [Scheffé, 1959] became the standard statistical technique to extract the internal and external components of variability from an ensemble of simulations, and thus to obtain estimates of potential predictability. Of course, the reliability of these estimates rely themselves on the ability of the AGCM to simulate the observed interannual variability (e.g., Smith [1995]) .

From the potential predictability studies above mentioned there is a general consensus that it is high in the tropics and rather low in the extratropics. In the tropics, the high potential predictability results not only from the low internal variability but also from the high forced variability which results from the direct thermodynamic response to SST variations. On seasonal-to-interannual timescales, the high values of potential pre-

dictability are associated with the El Niño-Southern Oscillation (ENSO) phenomenon. In the extratropics, the North Pacific-American (NPA) sector has the highest values of potential predictability [Barnett et al., 1997; Rowell, 1998; Renshaw et al., 1998; Feldstein, 2000; Zwiers et al., 2000], while the Euro-Atlantic (EA) region has low but still significant values of potential predictability [Davis et al., 1997; Cassou and Terray, 2001]. As in the tropics, ENSO events cause higher values of potential predictability in the NPA [Branković et al., 1994; Barnett et al., 1997; Chen and Van den Dool, 1997; Renshaw et al., 1998] and EA [Branković et al., 1994; Chen and Van den Dool, 1997; Mathieu et al., 2004] sectors. In both sectors, most locations have their highest potential predictability during winter or spring [Barnston, 1994; Branković et al., 1994; Chen and Van den Dool, 1997; Rowell, 1998; Zwiers et al., 2000]. A partial explanation for this seasonality was given by Opsteegh and den Dool [1980] and Webster [1982], whose linear models show that only when the extratropical westerly flow is far enough south can the Rossby waves communicate predictable signals from the tropics to the extratropics.

ENSO is an atmosphere-ocean coupled mode of interannual variability in the equatorial Pacific. Literature on the atmospheric component of ENSO, the Southern Oscillation, dates back to the classical series of papers of Walker, named *World Weather* (Walker [1928] and Walker and Bliss [1932], among others), followed by the observational work of Chen [1982]. The oceanic component of ENSO the El Niño, has been extensively documented since the observational works of Wyrtki [1975], Weare et al. [1976] and Weare [1982]. After the works of Bjerknes (Bjerknes [1969] and Bjerknes [1972], among others), Julian and Chervin [1978] and Barnett [1981] reporting the physical link between the Southern Oscillation and El Niño, the two phenomena has been taken as two components of the same phenomenon, the ENSO [Philander, 1981; Rasmunsson and Carpenter, 1982; Philander, 1985; Wang, 1992; Neelin et al., 1998; Trenberth and Caron, 2000].

Since ENSO is the strongest mode of interannual SST variability in the globe and is predictable several months in advance[1], a large number of observational works have been done to determine its relationship with worldwide atmospheric changes [Bjerknes, 1969; van Loon and Madden, 1981; van Loon and Rogers, 1981; Horel and Wallace, 1981; Pan and Oort, 1983; Ropelewski and Halpert, 1987; Kiladis and Diaz, 1989]. Especial attention has been given to the ENSO forcing of the two major patterns of atmospheric variability in Northern Hemisphere: the Pacific North American (PNA) pattern [Dickson and Namias, 1976; Wallace and Gutzler, 1981; Barnston and Livezey, 1987] in the NPA sector, and the North Atlantic Oscillation (NAO) [Walker and Bliss, 1932; van Loon and Rogers, 1978; Wallace and Gutzler, 1981; Barnston and Livezey, 1987] in the EA region. Note that both the PNA and the NAO patterns are natural modes of variability in the northern hemisphere. They are identified as preferred modes of variability in AGCMs forced with climatological SSTs [Barnett, 1985; Glowienka-Hense, 1990; Cassou and Terray, 2001], in AMIP experiments using internal variability data [Harzallah and Sadourny, 1995], and, since internal variability is higher than the forced variability in the extratropics, also using total variability data [Zwiers et al., 2000].

A theoretical explanation for the ENSO forcing on the NPA region was given by Hoskins and Karoly [1981] (see also Opsteegh and den Dool [1980]; Webster [1981, 1982] and Lau and Lim [1984]). These authors showed that the PNA pattern found in the mid-tropospheric geopotential height field of the northern hemisphere bears a strong qualitative resemblance to the steady-state solutions of the linearised primitive equations on a sphere, forced by a tropical heat source. Anomalous SST in the tropical Pacific forces anomalies in convection and large-scale overturning with subsidence in the descending branch of the local Hadley circulation. The resulting strong upper tropospheric divergence in the tropics and convergence in the subtropics act as a Rossby

[1]Barnett [1993] proposed that forecasts of winter SST at lead times of at least 6 months are good enough to be used with atmospheric models to attempt long-range winter forecasts for the North American continent. Collins et al. [2002] found that ENSO could be usefully predicted, on average, up to 8 months in advance.

wave source. The emanating wave trains carry energy into the extratropics in a great circle path that projects on the PNA pattern. Sardeshmukh and Hoskins [1988] showed that the climatological stationary waves and associated jet streams can make the total Rossby wave sources somewhat insensitive to the position of the tropical heating that induces them and thus can create preferred teleconnection response patterns, such as the PNA.

There is a general agreement that the impact of ENSO on the EA region is weaker and less robust than on the PNA sector but still significant [Pozo-Vásquez et al., 2001; Cassou and Terray, 2001; Gouirand and Moron, 2003; Mathieu et al., 2004]. Although largely studied, the response of the EA atmosphere to ENSO forcing is not consensual and the physical mechanisms involved are not yet fully understood. One of the objectives of this work is, thus, to improve the knowledge on this subject, focusing on the non-linear behaviour of the response. It will be showed that the impact on the EA region can be characterised, at least in part, in terms of changes in the frequency of occurrence of the NAO regimes Melo-Gonçalves et al. [2005].

In the Atlantic sector, it is consensual that the NAO forces the underlying SSTs resulting in the North Atlantic SST anomaly Tripole pattern [Deser and Blackmon, 1993], but recent works suggest that the latter may feedback on the former at longer timescales. Several works have also been published reporting the response of the EA atmosphere to the tropical Atlantic SSTs at the decadal timescale Venzke et al. [1999]; Sutton and Hodson [2003]; Hodson et al. [2003]. However, the response of the atmospheric circulation to both tropical and extratropical SSTs on interannual timescales is less studied and understood [Czaja and Frankignoul, 2002; Frankignoul et al., 2003], probably because of its blending with the remote signal from the tropical Pacific associated with ENSO events, and the influence of the latter on the tropical Atlantic SSTs through the so called *atmospheric bridge* [Klein et al., 1999; Saravanan and Chang, 2000; Giannini et al., 2001; Alexander et al., 2002; Huang, 2004]. For this reason, another purpose of this work is the study of the *interannual* Atlantic SST forcing on the EA atmospheric circulation.

In order to achieve the above mentioned objectives, two experiments were conducted with the Melbourne University General Circulation Model (MUGCM). The first experiment conducted with the MUGCM followed the rules specified by AMIP II. In the second experiment, the MUGCM was forced by the climatological annual cycle of monthly mean SST and SIC.

The layout of this thesis is as follows. Chapter 2 describes the datasets assumed to represent the observed atmosphere and the model experiments performed with the MUGCM to obtain the simulated data. In chapter 3, the ability of the MUGCM to simulate the observed climate and variability is assessed, using the reanalysis from the National Centers for Environmental Prediction (NCEP) as representative of the observed atmosphere. Chapter 4 presents estimates of internal and forced variabilities at each grid point of the EAregion, obtained by performing an ANOVA to the MUGCM AMIP II ensemble. Potential predictability estimates are also provided for each grid point of the EA sector, and the regions with significant SST-forcing are identified. The total variability of two indices of the NAO is also decomposed into its internal and forced components, and its potential predictability is estimated for consecutive and overlapping three-month seasons throughout the year. In chapter 5, the leading modes of mean sea level pressure (MSLP) variability in the EA region forced by SST and SIC variability are estimated. The regions of the global ocean where SST anomaly variability is responsible for the detected MSLP forced modes are localised, and the forcing SST modes of variability are identified. Chapter 6 addresses the forcing of the tropical Pacific SST anomalies on the MSLP variability in the EA sector by studying the sensitivity of the NAO to ENSO polarity and strength. Finally, chapter 7 provides a summary of the analyses and the major conclusions of the this work.

Chapter 2

Observed and Simulated Data

This chapter describes the observed data and the model experiments performed with the MUGCM to obtain the simulated data used in this work.

2.1 Observed data

It is assumed here that the observed data will represent the "'real" atmospheric circulation and sea surface temperatures. The following datasets were used:

- NCEP datasets - monthly mean data of several variables on a global 2.5º latitude by 2.5º longitude grid, from 1950 to 2001, extracted from the NCEP reanalysis [Kalnay and co authors, 1996].

- CMAP dataset - monthly means of precipitation from 1979 to 2001, on a 2.5º latitude by 2.5º longitude grid, from the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) [Xie and Arkin, 1996]).

2.2 Model and numerical experiments

The AGCM used was the version VIII.2 of the MUGCM. The MUGCM is a spectral atmospheric model with rhomboidal truncation at wave 31, which corresponds to a horizontal resolution of approximately 3.75º longitude by 2.25º latitude. The MUGCM

uses nine vertical sigma levels: 0.991, 0.926, 0.811, 0.664, 0.500, 0.336, 0.189, 0.074 and 0.009. Both the diurnal and the seasonal cycle are included and radiation is allowed to interact with CO_2, ozone, water vapour and clouds centre. The model includes prognostic clouds, SIC and prescribed SSTs. The model was derived from a hemispheric version of the model described by Bourke et al. [1977] and McAvaney et al. [1978]. Several modifications were made to the physics of the model and the structure of the code, and description of some of these may be found in Simmonds [1985].

Two experiments have been performed with the MUGCM. The first experiment conducted with the MUGCM followed the rules specified by AMIP II. AMIP II is the successor of AMIP [Gates, 1992; Gates et al., 1998], but uses improved SST and SIC boundary conditions [Taylor et al., 2000], and the time period was extended from January 1979 - December 1988 to January 1979 - February 1996. The experimental design proposed by the AMIP projects consists in performing, with an AGCM an ensemble of integrations all forced by the same monthly varying observed SST and SIC boundary conditions, specified for a determined period of time, and random initialised. In our experiment with the MUGCM we performed twenty nine integrations, all forced by the SST and SIC boundary conditions specified by AMIP II for the period January 1979 - February 1996 [Taylor et al., 2000], and randomly initialised with initial conditions obtained by a previous control run.

Our AMIP II experiment with the MUGCM results in twenty nine realisations of the climate's path through its phase space from January 1979 to February 1996. Since the atmosphere forgets its initial state very quickly [Lorenz, 1963], these realisations are independent. Despite the model deficiencies in simulating the climate, the availability of several independent realisations of the climate is a major advantage of this kind of model data over the observed datasets, since these can only offer one realisation of the climate.

For some years now, the term Global Ocean Global Atmosphere (GOGA) appeared in the literature [Lau and Nath, 1994] to name an AGCM experiment where monthly

varying SSTs are used for the entire global ocean to force *one* AGCM simulation over the global atmosphere. Neither the database of the boundary conditions nor the time period of the simulation are specified. Usually, SSTs are prescribed to the monthly values observed for a particular period of time. Such nomenclature arose to distinguish this experiment from those where SSTs are monthly varying only in some part of the global ocean and kept constant and equal to the monthly climatology elsewhere. Examples of such experiments are the Tropical Ocean Global Atmosphere (TOGA)[1] and the Midlatitude Ocean Global Atmosphere (MOGA).

The purpose of experiment designs to force AGCMs, such as TOGA, MOGA, and even Pacific Ocean Global Atmosphere (POGA), Indo-Pacific Ocean Global Atmosphere (I-POGA) and Topical Atlantic Global Atmosphere (TAGA), is to isolate the SST forcing of a particular ocean region of the global ocean, or to study the relative importance of different ocean regions in SST-forcing the atmospheric circulation. For example, Lau and Nath [1994], performed GOGA and TOGA experiments in order to study the relative importance of tropical *versus* extratropical Pacific SST anomalies in forcing the midlatitude atmospheric circulation (see also Lau [1997]). The TAGA experiment has been successful in determining the atmospheric response over the tropical Atlantic and over the EA region to SST anomalies in the tropical Atlantic without the interference of the ENSO signal [Chang et al., 2000; Saravanan and Chang, 2000; Giannini et al., 2001].

By the GOGA definition, an AMIP experiment can be considered as an ensemble of GOGA simulations and, for this reason, we will often use the term GOGA to refer to our AMIP II simulations.

[1]TOGA is also an international research programme designed to study the short-term climate variations (time scales of months to years) using the 1985-1994 period - the TOGA decade. The TOGA programme focused on the interannual variability of the coupled ocean-atmosphere system associated with ENSO. See National Research Council [1996] and the especial issue of the Journal of Geophysical Research (No. C7, Vol. 103, June 1998), for example, McPhaden et al. [1998], Trenberth et al. [1998], Wallace et al. [1998] and Neelin et al. [1998].

In the second experiment, the MUGCM was forced by the climatological annual cycle of monthly mean SSTs and SIC. The climatology for each month was computed from the SST and SIC data of used to forced the AMIP II simulations. In this experiment, the MUGCM also includes radiative forcing from sulphate aerosols. The geographical distribution of the concentration of aerosols is the same used by the Hadley Centre Coupled Climate Model 2 [Johns et al., 1997]. An ensemble of thirty three integrations, all forced by the same monthly mean SST and SIC annual cycle and random initialised, were performed. We will refer to these simulations as the Annual CYCle (ACYC) simulations. The simulations will be used to estimate the internal variability of the MUGCM which will be compared with the one obtained by the AMIP II ensemble.

2.3 Seasonal mean data

The analysis in this work was performed using seasonal mean data (chapter 3, sections 3.1 to 3.3) and seasonal mean *anomaly* data (rest of the work) for the each of the four standard seasons: winter - December, January and February (DJF), spring - March, April and May (MAM), summer - June, July and August (JJA), and autumn - September, October and November (SON). Seasonal mean data for each season was calculated by averaging monthly mean data of the corresponding calendar months. Seasonal mean *anomaly* data was obtained by averaging monthly mean *anomaly* data of the corresponding calendar months. Monthly mean *anomaly* data was computed by removing the annual cycle from monthly mean data. January climatology was subtracted from January data of all years, February climatology was subtracted from February data of all years, and so on.

Seasonal mean data and seasonal mean anomaly data for the four standard seasons were obtained from (i) the MUGCM AMIP II monthly mean dataset (29 simulations of monthly means from January 1979 to February 1996), (ii) the NCEP monthly mean dataset (monthly means from January 1950 to December 2001), and (iii) the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) monthly mean

Table 2.1: Time range and number of seasonal means (N) of the MUGCM AMIP II, NCEP and CMAP datasets.

Dataset	Winter (DJF)	Spring (MAM)	Summer (JJA)	Autumn (SON)
MUGCM AMIP II	1980 - 1996	1979 - 1995	1979 - 1995	1979 - 1995
N	$29 \times 17 = 493$	$29 \times 17 = 493$	$29 \times 17 = 493$	$29 \times 17 = 493$
NCEP	1951 - 2001	1950 - 2001	1950 - 2001	1950 - 2001
N	51	52	52	52
CMAP	1980-2001	1979 - 2001	1979 - 2001	1979 - 2001
N	22	23	23	23

dataset (monthly means from January 1979 to December 2001). The time range and number of seasonal means of each dataset and season are presented in table 2.1.

Since seasonal mean data is only used in chapter 3, sections 3.1 to 3.3, where the seasonal climatology of the model is compared with the NCEP and CMAP climatologies, the term *seasonal mean anomalies* are often simply referred to as *seasonal means* in the rest of the work.

Chapter 3

Model Validation

The reliability of the results obtained by an experiment performed with an AGCM depends on the capability of the model to simulate the *true* atmospheric behaviour. It is thus mandatory to assess the performance of the model. In the case of an AMIP II experiment, in which the model is forced by observed SST and SIC, the model's climate variability must be compared with the observed climate variability. Furthermore, much of the variability from interannual to decadal timescales is closely linked to the location and intensity of the mean atmospheric features. The model's ability to reproduce them as well as their seasonal fluctuations thus appears essential [Kumar et al., 1996; Smith, 1995; Cassou and Terray, 2001]. For the MUGCM, much of this work has already been done and published. An extensive atlas of its climatology can be found in Simmonds et al. [1988]. It's performance has been compared with other AGCMs [Simmonds, 1990; Boer et al., 1991, 1992], and it has been shown to simulate well the interannual variability over the globe [Simmonds and Smith, 1986; Castanheira, 2000], Northern Hemisphere [Walland and Simmonds, 1997], world's oceans [Simmonds and Dix, 1989], Antarctic [Simmonds, 1990], Australia [Simmonds and Lynch, 1992], south-eastern Africa [Rocha, 1992; Rocha and Simmonds, 1997], and the Iberian Peninsula [Alves and Rocha, 2003].

Despite the numerous cited works supporting the ability of the MUGCM to simulate the observed climatology and variability, the validation of MUGCM's climatology and

variability is presented in this chapter, using the NCEP reanalysis. The validation is performed using seasonal means for winter, spring, summer and autumn (see table 2.1).

Although the time range of the MUGCM simulations is confined to the January 1979 to February 1996 period, we use the entire time range of the NCEP data (January 1950 to December 2001) because we consider the NCEP data set as a sample of realisations of the atmospheric random variables. The statistics of this sample, namely the sample mean (climatology) and the sample variance-covariance matrix (variability) are estimators of the *true*, that is populational, mean and variance-covariance matrix of the atmospheric random variables. Since the mathematical expectation of these sample statistics are the populational statistics, it would be a waste not to use all the observations available.

The layout of this chapter is as follows. In section 3.1, MUGCM's climatology in the EA region is compared to the NCEP climatology using winter, spring, summer and autumn seasonal means of MSLP, surface temperature (TMP-SFC), precipitation (PRECIP), and geopotential height (HGT) at 850 (HGT-850), 500 (HGT-500) and 200 (HGT-200) mb.

In section 3.2, modelled and observed vertical structures of zonal-mean zonal circulation, in winter and summer, are compared. Departures of winter and summer upper tropospheric zonal wind from zonal-mean symmetry, in particular the jet streams, are discussed in section 3.2.1. Section 3.2.2 discusses the model upper tropospheric planetary standing waves in winter, and the roles of orographic and midlatitude thermal forcings.

In section 3.3, the simulated winter climatology of meridional and zonal mass overturning circulations in the Pacific and Atlantic (Hadley, Ferrel and Walker circulation) are compared to the corresponding NCEP climatologies.

In section 3.4, MUGCM's variability is compared to the observed variability using two different methodologies. In section 3.4.1, the same variables used to test the climatology of the model (MSLP, TMP-SFC, PRECIP, HGT-850, HGT-500 and HGT-200) are used to compute the correlation between the MUGCM AMIP II ensemble mean and NCEP, at each grid point of the MUGCM's grid in the EA sector, for each of the four standard seasons. In section 3.4.2, we compare the spatial patterns of variability of MUGCM and NCEP seasonal mean *anomalies* of MSLP, in the EA region, obtained by Principal Component Analysis (PCA).

Since the purpose of this chapter is the validation of the model, objective quantitative comparisons between modelled and observed features are done. Is is tried, however, to go beyond a merely quantitative approach by also providing a qualitative description of the atmospheric phenomena involved. Besides the published works referenced in text, much of the descriptions and discussions were done with the aid of the exceptional books of Peixoto and Oort [1992] and Holton [1992].

3.1 Climatology of the Euro-Atlantic atmosphere

MUGCM and NCEP climatologies in the EA region of MSLP, TMP-SFC, PRECIP, HGT-850, HGT-500 and HGT-200, in the four standard seasons (see table 2.1), are compared and discussed in sections 3.1.1 to 3.1.4. In order to assist this comparison with a quantitative measure, the following hypothesis test (see, for example, Milton and Arnold [1995]) was performed:

$$H_0 : \mu_1 = \mu_2 \; vs \; H_1 : \mu_1 \neq \mu_2, \tag{3.1}$$

where μ_1 and μ_2 are the populational (*true*) means of a particular atmospheric variable from the MUGCM and NCEP dataset, respectively, at each grid point of the MUGCM EA domain. Note that the NCEP data was regridded to the MUGCM's grid, using bilinear interpolation, prior to the computation of the test. In the test 6.14, it is not assumed that the populational variances of MUGCM and NCEP are equal. A Z random variable, with the populational variances replaced by the sample variances, is used as

the test statistic,

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \tag{3.2}$$

where $\bar{X}$, S^2 and n are the sample means, sample variances and the number of observations, respectively, and the subscripts 1 and 2 refer to MUGCM and NCEP data respectively. The number of observations is 29 simulations $\times$ 17 years = 493 for MUGCM data and 51 (for winter) or 52 (for spring, summer and autumn) years for NCEP data, respectively (see table 2.1). Note that these observations are independent, as required by the hypothesis test. Invoking the Central Limit Theorem, this test statistic has, if H_0 is true, a t-Student distribution with a number of degrees of freedom that can be estimated by the Smith-Satterthwaite procedure (see, for example, Milton and Arnold [1995]),

$$T \sim t_\gamma, \qquad \gamma = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_1^2/n_1)^2}{n_2 - 1}}. \tag{3.3}$$

The hypothesis test 6.14 was performed, at a significance level of $\alpha = 0.001$, at each grid point of the MUGCM EA domain, for each variable in each season. The rejection of the null hypothesis at a significance level of α, at a particular grid point, which occurs when

$$|t| > t_{\gamma,1-\alpha/2}, \tag{3.4}$$

where t[1] is a realisation of the test statistic T of equation 3.3, implies that we can not consider that the means of the MUGCM and NCEP populations are equal, which means that the model does not simulate accurately the observed data at that grid point. The percentage of the EA area where the null hypothesis is rejected, for each variable and season, is presented in table 3.1 (value on the top left-hand corner of each entry of the table). These percentages are also shown in figures 3.1 to 3.7 (value on the top right-hand of the left panels).

The agreement between the MUGCM and NCEP climatologies is also assessed, as in Davis et al. [1997], computing the spatial correlation between MUGCM and NCEP climatology patterns for each variable and each season. Note again that the NCEP

[1] In equation 3.4, $t_{\gamma,1-\alpha/2}$ is such that $P[T \leq t_{\gamma,1-\alpha/2}] = 1 - \alpha/2$.

Table 3.1: Percentage of the EA area where MUGCM and NCEP seasonal climatologies are significantly different, at the significance level of 0.001 (hypothesis test 6.14), and spatial correlations between MUGCM and NCEP climatologies. In each entry of the table, the left values are area percentages and the right values are correlations. The top values were obtained when all NCEP years were used and the bottom values were obtained when only the AMIP II years of the NCEP data were used.

Variable	DJF		MAM		JJA		SON	
SLP	80.7	0.9998	89.7	0.9998	96.0	0.9999	82.4	0.9999
	52.6	0.9999	81.7	0.9999	87.7	0.9999	57.4	0.9999
TMP-SFC	61.0	0.9724	61.1	0.9810	60.4	0.9860	54.1	0.9896
	46.5	0.9740	43.1	0.9816	46.2	0.9858	41.4	0.9891
PRECIP	82.5	0.7427	75.0	0.6855	80.0	0.4306	80.4	0.7010
	71.6	0.7608	66.7	0.6963	67.6	0.4280	69.7	0.6986
HGT-850	75.7	0.9973	89.2	0.9970	91.9	0.9981	77.6	0.9991
	43.7	0.9974	79.3	0.9972	86.6	0.9978	59.8	0.9990
HGT-500	70.3	0.9995	87.8	0.9995	92.5	0.9999	86.9	0.9999
	45.5	0.9995	73.3	0.9996	89.9	0.9999	75.4	0.9999
HGT-200	81.3	0.9997	91.2	0.9996	74.1	0.9999	70.5	0.9999
	58.0	0.9997	81.5	0.9997	58.1	0.9999	55.5	0.9999

data was regridded to the MUGCM's grid, using bilinear interpolation, prior to the computation of the correlations. The significance of these correlations was not tested because the observations are not independent (values at the grid points of the EA region). The spatial correlation, for each variable and season, is shown in table 3.1 (value on the top right-hand corner of each entry of the table).

The percentage of the EA area where the null hypothesis of test 6.14 is rejected and the spatial correlation between MUGCM and NCEP data, were also calculated using only the AMIP II years of the NCEP data. The percentages are shown in table 3.1 on the bottom left-hand corner of each entry of the table, and the spatial correlations are presented on the bottom right-hand corner. These percentages and correlations are higher for all cases, as expected, since the time period used to compute the climatologies is the same.

The percentage of the EA area where modelled and observed climatologies are significantly different from zero, at a significance level of 0.001, is high for all variables and seasons. The percentages for TMP-SFC are much lower than the others because the temperatures on the surface of the model oceans are observed SSTs. All the percentages are, however, much lower when only the AMIP II years are used to compute NCEP climatologies. Still, they are around 50%. These results suggest that the model does not reproduce well the NCEP climatology. However, it should be noted that the hypothesis test 6.14 compares the climatologies at one grid point at a time. It does not compare the spatial patterns of the two climatologies as a whole. For example, figure 3.1 shows that the model is able to reproduce the Icelandic Low and the Azores High, but the position of these centres are biased relative to the observed ones. These biases cause that test 6.14 compares, for example, the highest pressure grid point of the observed Azores High with a grid point that is not the highest pressure grid point of the modelled Azores high, and, consequently, rejects the null hypothesis at that grid point. Note also that figures 3.1 to 3.7 and the high spatial correlations between modelled and observed climatologies support that the model is able to reproduce the main features of the observed climatology spatial patterns.

3.1.1 Mean sea level pressure

For MSLP, the model provides a good simulation of the observed seasonal spatial pattern including the Icelandic Low and Azores High, the meridional southward pressure gradient and also its seasonal variation. The simulated meridional MSLP gradient reaches its maximum in the winter season, as expected, because it is when the polar-to-equator temperature gradient is higher (section 3.1.2, figure 3.2). However, it is slightly overestimated between the Azores High and the Icelandic Low because the simulated pressure cells are too intense in winter. Another bias of the MSLP MUGCM climatology is found in the direction of the isobars between the two large-scale pressure cells: in all seasons except summer, modelled isobars are almost parallel to the latitude circles while observed isobars cross the them from south to north.

3.1.2 Surface temperature

Figure 3.2 shows that the MUGCM simulates well the climatologies of the surface temperature and also the pole-to-equator gradient and its seasonal variation. As expected, the simulated pole-to-equator temperature gradient is highest in winter, consistent with the seasonal variation of the differential heating.

3.1.3 Precipitation

Precipitation (figure 3.3) is less well reproduced than the other analysed variables, a characteristic common to all models because it is a difficult variable to model due to its highly irregular spatial and temporal behaviour. Nevertheless, the main features of the observed distribution are reproduced by the model. These include low precipitation rate associated to low moisture content at high latitudes where temperature is very low, specially in Greenland (see figure 3.2) and in the region under the influence of the semi-permanent Azores High and associated subsidence (see figure 3.1).

Roughly, the simulated precipitation is underestimated over the western Atlantic ocean and Mediterranean sea in summer and spring and overestimated (underestimated)

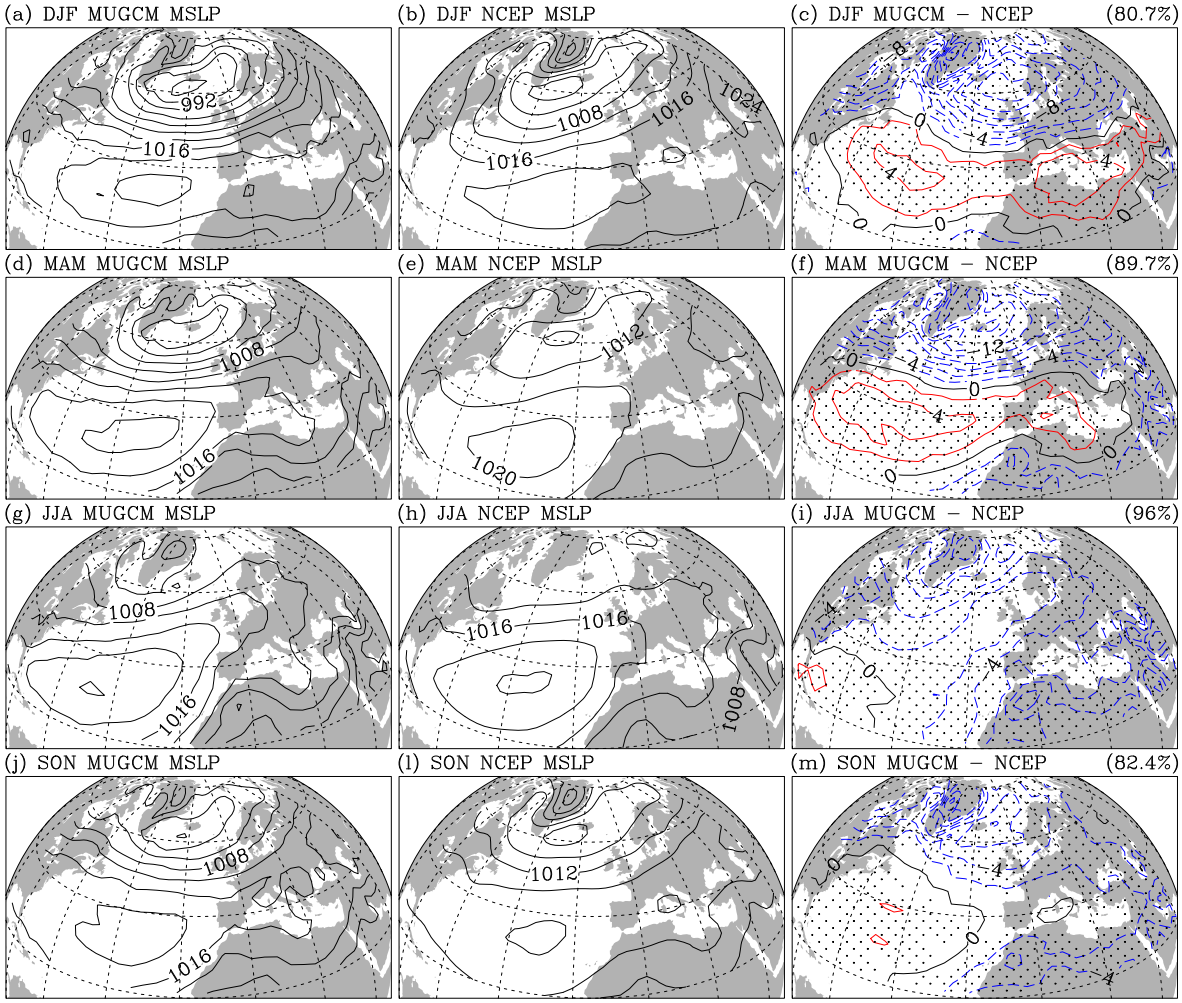

Figure 3.1: MSLP seasonal climatologies in the EA region. From the MUGCM ensemble mean: (a) DJF, (d) MAM, (g) JJA and (j) SON. From NCEP data: (b) DJF, (e) MAM, (h) JJA and (l) SON. Difference MUGCM - NCEP: (c) DJF, (f) MAM, (i) JJA and (m) SON. Units in mb. Isobar spacing of 4 mb for (a,b,d,e,g,h,j,l) and 2 mb for (c,f,i,m). The area is stippled where the difference between MUGCM and NCEP pressure is significantly different from zero, at a significance level of 0.001 (hypothesis test 6.14). The percentage of total area that is stippled is printed on the top right-hand corner of each panel. Grid lines represented every 20^{o}.

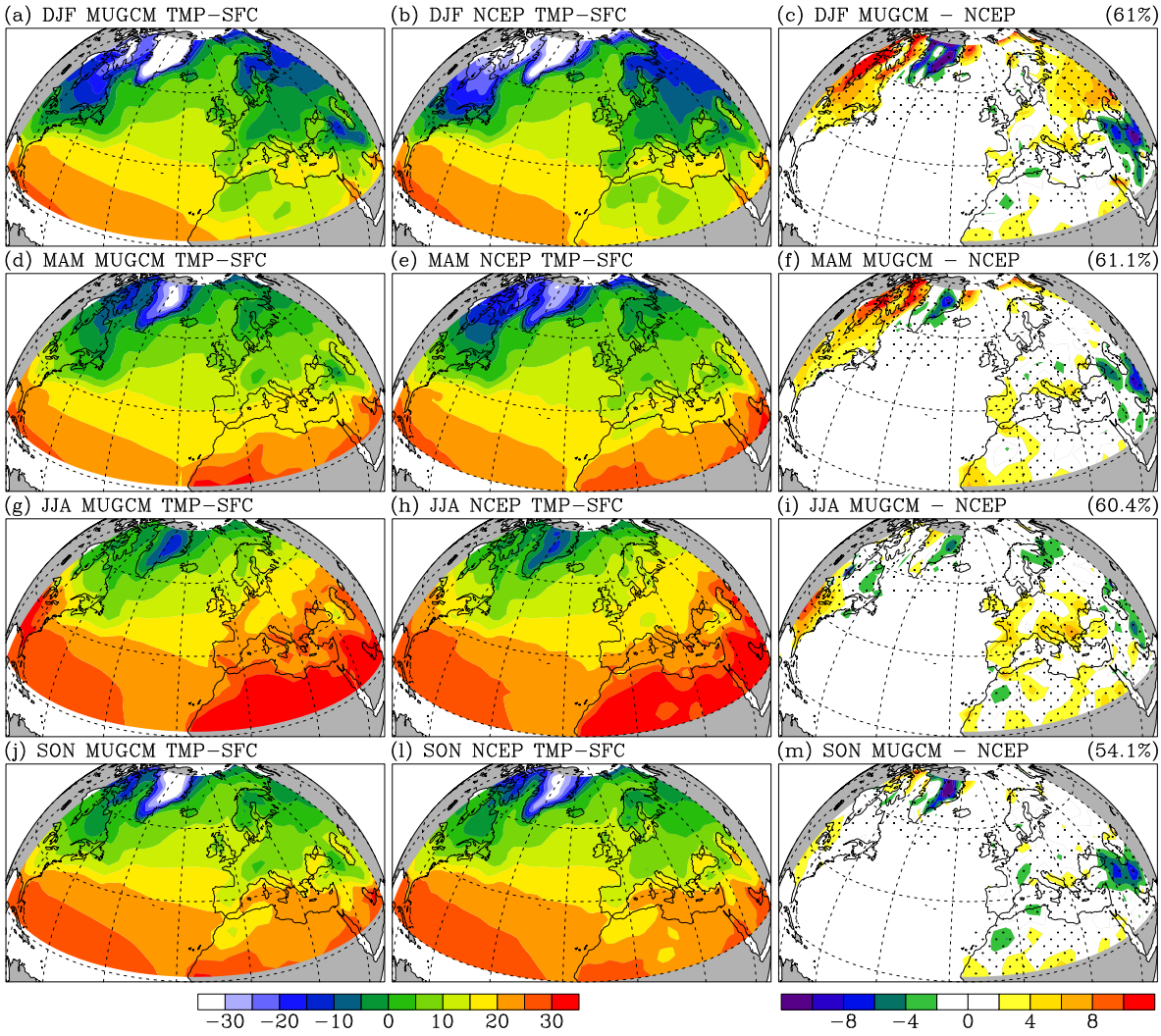

Figure 3.2: As in figure 3.1, but for TMP-SFC. Units in °C. Isotherm spacing of 5° C for (a,b,d,e,g,h,j,l) and 2° C for (c,f,i,m).

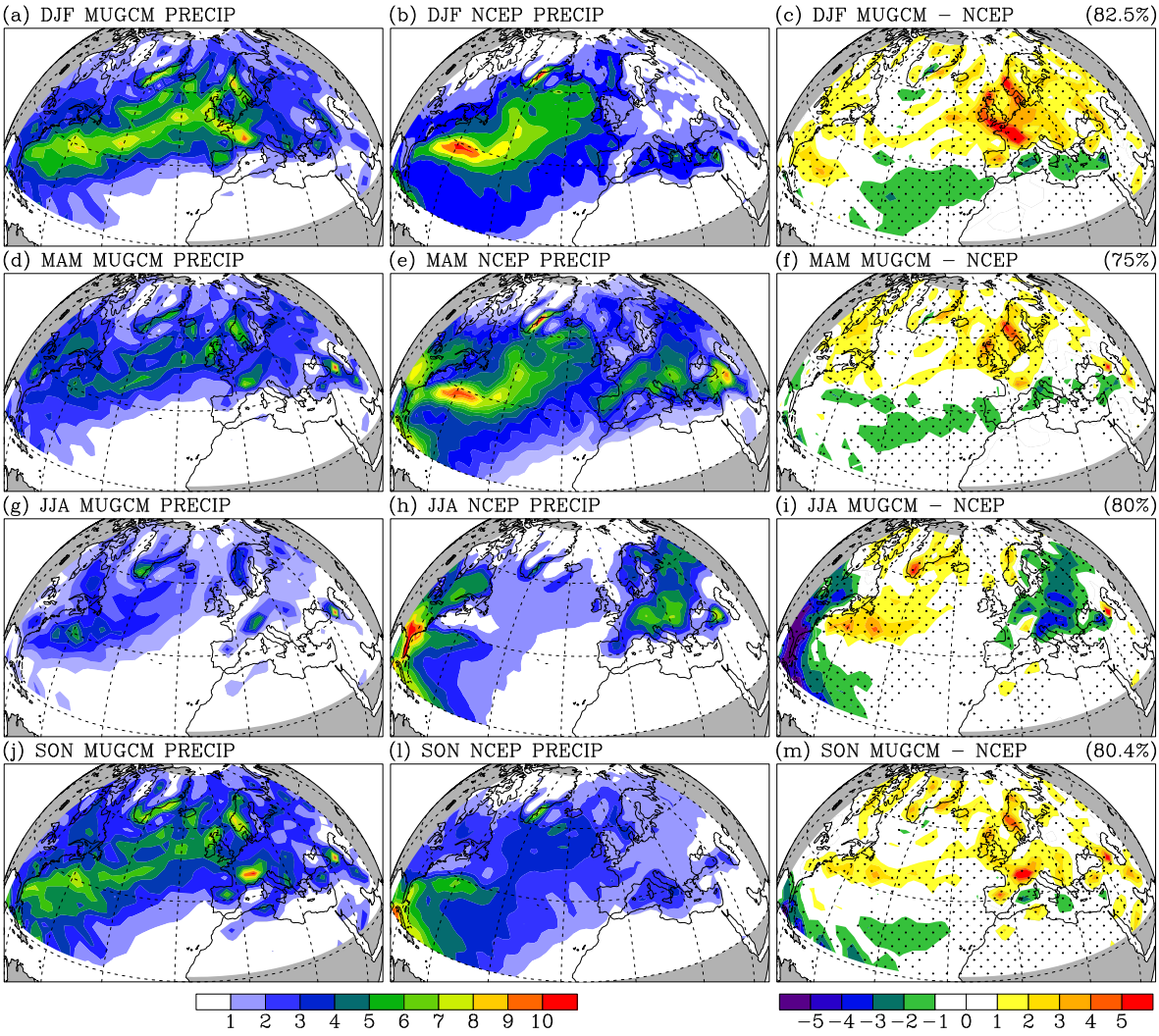

Figure 3.3: As in figure 3.1, but for PRECIP. Units in mm/day. Isoline spacing of 2 mm/day for (a,b,d,e,g,h,j,l) and 1 mm/day for (c,f,i,m).

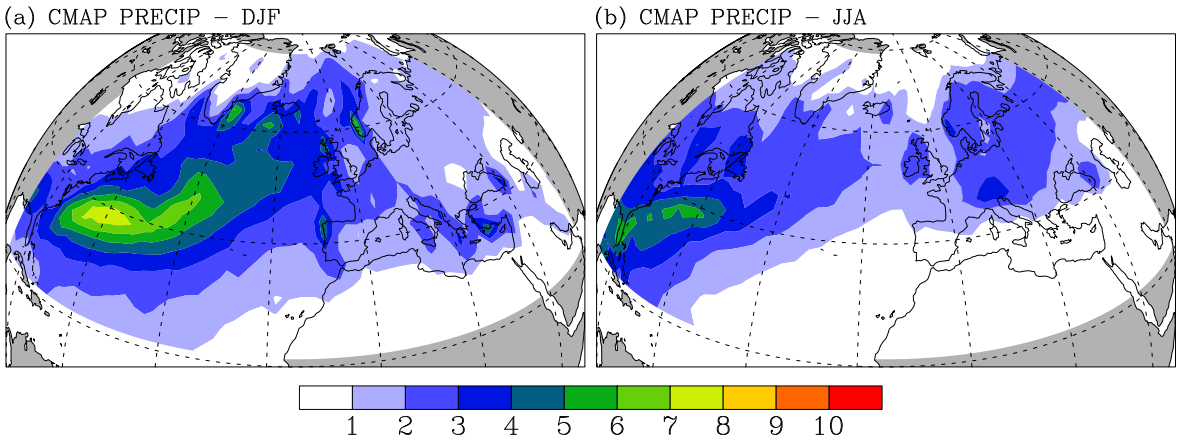

Figure 3.4: CMAP seasonal climatologies (1979-2001) in the EA region in (a) DJF and (b) SON. Units in mm/day. Isoline spacing of 1 mm/day. Grid lines represented every 20º.

over the central and eastern Europe in winter (summer).

Note that the MUGCM winter and summer climatology is considered biased in central and eastern Europe because the NCEP precipitation rate in this region is higher in summer than in winter. Since this winter/summer contrast is the opposite of that found in the rest of the Euro-Atlantic sector, it is pertinent to wonder if it is the NCEP climatology, rather than the MUGCM climatology, that is biased in this region. This is not, however, the case, as it may be confirmed by the winter and summer climatologies computed with the CMAP [Xie and Arkin, 1996] dataset (see chapter 2) shown in figure 3.4.

3.1.4 Geopotential height

Observed geopotential height climatologies at 850, 500 and 200 mb (figures 3.5 to 3.7) are well simulated by the model, specially in middle and upper tropospheric levels. The patterns of geopotential height at 850 mb are, as expected, similar to the MSLP patterns, as well as the biases relative to the observed patterns.

Since extratropical midlatitude atmospheric large-scale horizontal[2] circulation is nearly geostrophic, the wind field at pressure level p can be approximated to by the geostrophic wind field at the same level,

$$\mathbf{V}_g = \mathbf{k} \times \frac{1}{f} \nabla_p \Phi \tag{3.5}$$

$$= \mathbf{k} \times \frac{1}{f} \nabla_p \Big(g_o Z \Big), \tag{3.6}$$

where and Φ is the geopotential, $Z = \Phi / g_o$ is the geopotential height, f is the Coriolis parameter and g_o is the global average of gravity at mean sea level.

From equation 3.6 follows that $g_o Z$ is proportional to the streamfunction of the geostrophic wind and thus isolines of geopotential height are a good approximation to the streamlines of the actual wind field at the corresponding pressure level. Because of this, the maps of figures 3.5 to 3.7 can be used to qualitatively diagnose the seasonal climatology of the tropospheric circulation.

The maps of MSLP (figure 3.1) and 850 mb HGT (figure 3.5) show that the model reproduces fairly well the observed low-tropospheric circulation in all seasons including the cyclonic (anticyclonic) circulation associated to the Low (High) pressure cells and the intensification of the westerly wind between the two pressure systems during winter. However, since the model overestimates the pressure/height gradient between the two pressure systems in all seasons and specially in winter, the same overestimation can be attributed to the low-tropospheric wind.

The model overestimation of the low-tropospheric winter circulation is consistent with the positive bias of the modelled precipitation in Europe during winter: an intensified westerly wind increases the advection of moisture into Europe which combined with the ocean/land temperature contrast (land colder than ocean in winter, figure 3.2(a)) leads to an intensification of the precipitation winter climatology.

[2]Throughout this work *horizonal fields* is used interchangeably to refer to isoheight or isobaric surface fields.

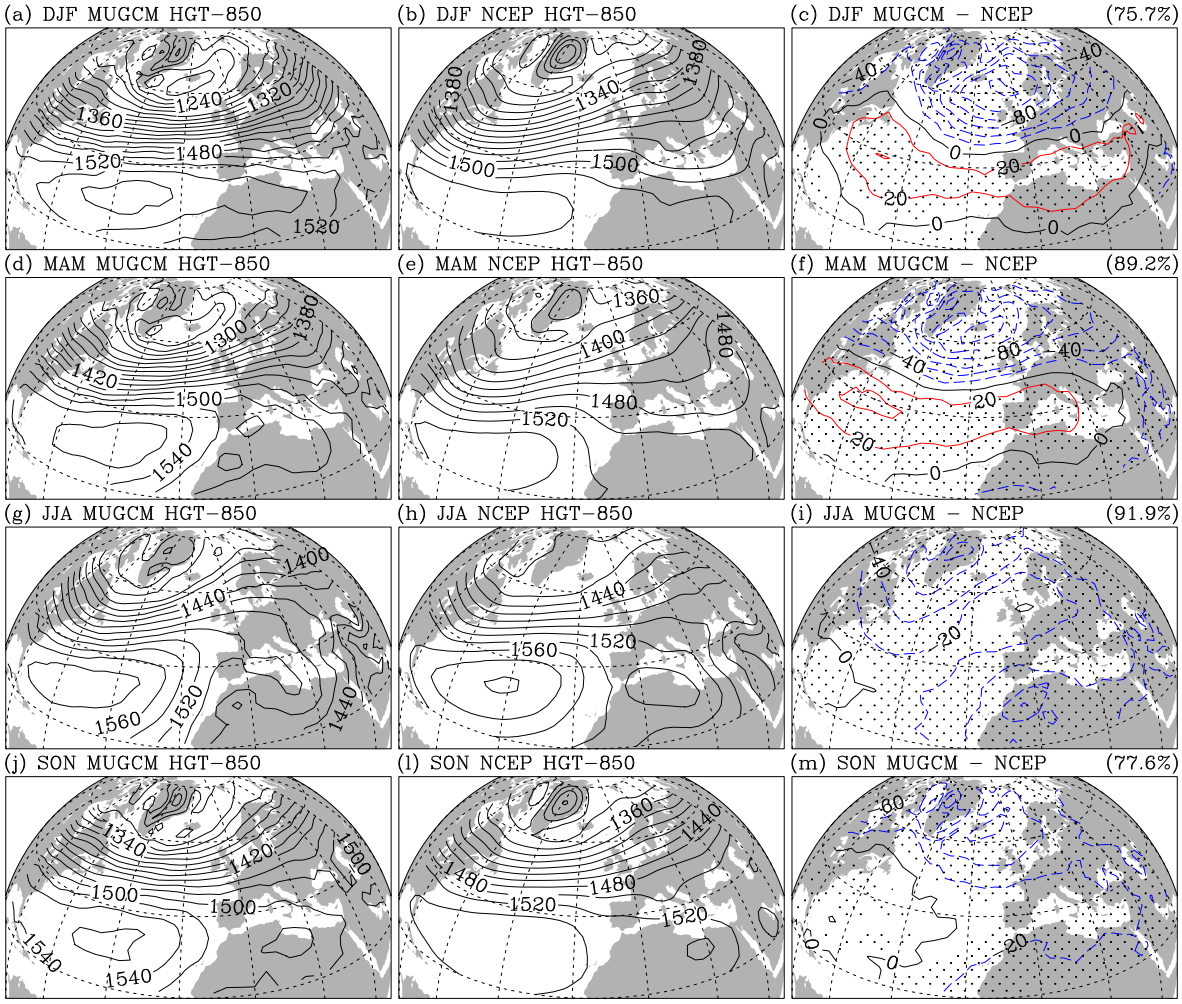

Figure 3.5: As in figure 3.1, but for 850-mb HGT. Units in gpm and isoline spacing of 20 gpm.

Finally, and using again the geostrophic approximation and the maps of 500 and 200 mb HGT (figures 3.6 and 3.7), we note that the observed nearly zonal westerly circulations at the middle and upper troposphere are also fairly reproduced by the model.

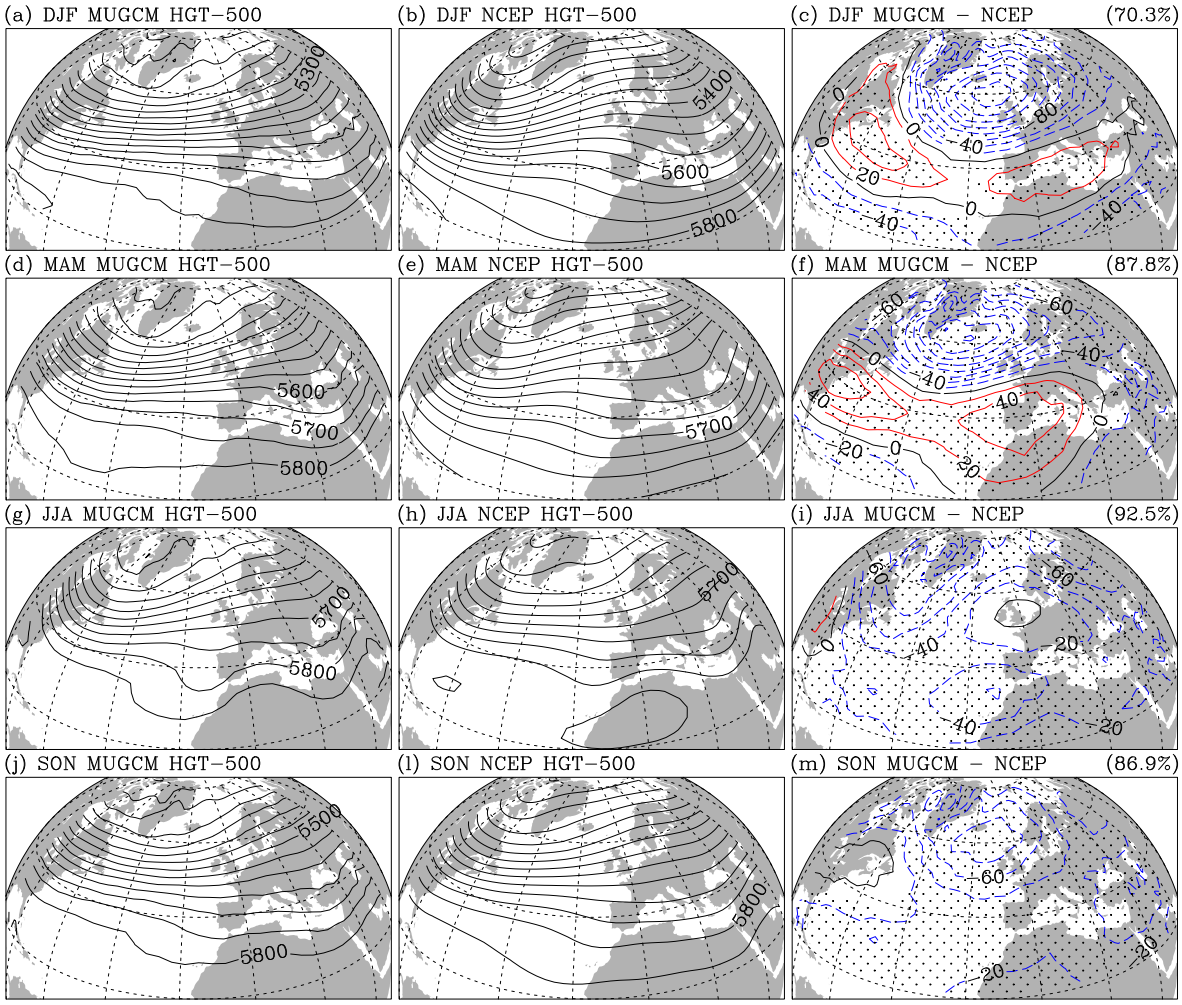

Figure 3.6: As in figure 3.1, but for 500-mb HGT. Units in gpm. Isoline spacing of 50 gpm for (a,b,d,e,g,h,j,l) and 20 gpm for (c,f,i,m).

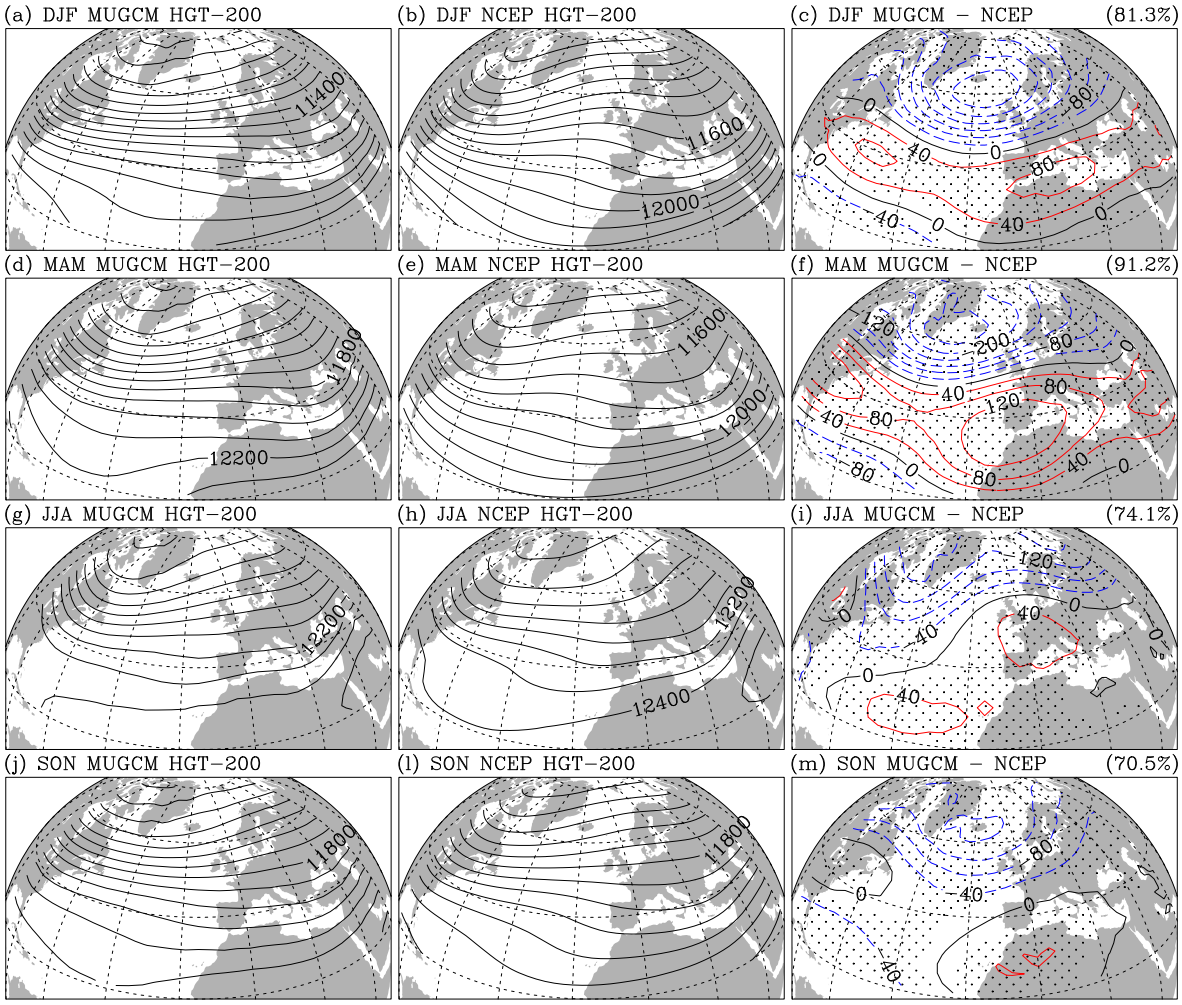

Figure 3.7: As in figure 3.1, but for 200-mb HGT. Units in gpm. Isoline spacing of 100 gpm for (a,b,d,e,g,h,j,l) and 40 gpm for (c,f,i,m).

3.2 Climatology of the zonal-mean zonal circulation

An AGCM midlatitude response to SST forcing depends sensitively on the details of the climatological flow [Kumar et al., 1996]. It is thus essential to validate the MUGCM's flow climatology. The vertical structure of the zonal-mean circulation simulated by the model is now compared to the NCEP structure, using cross sections of longitudinally averaged zonal wind (UWIND) and temperature (TMP) for the solstice seasons, winter and summer (figure 3.8). Although the vertical-meridional planes in figure 3.8 are displayed from 90°S to 90°N, only the Northern Hemisphere is discussed here.

At midlatitudes, the zonal-mean zonal wind is nearly geostrophic,

$$\bar{u}_g = -\frac{1}{f}\frac{\partial \bar{\Phi}}{\partial y}, \tag{3.7}$$

and, hence, satisfies to a high degree of accuracy the *thermal wind* relationship

$$\frac{\partial \bar{u_g}}{\partial \ln p} = \frac{R}{f}\frac{\partial \bar{T}}{\partial y}, \tag{3.8}$$

where $R = 287$ J kg^{-1} K^{-1} is the gas constant for dry air, and the overbar represents zonally averaged quantities.

The model is able to simulate the general structure of the observed zonal-mean wind and also its seasonal variation. In particular, the observed behaviour of the jet stream is reproduced by the model. It is located just below the tropopause at the latitude where the thermal wind (equation 3.8) integrated through the troposphere reaches its maximum. This maximum is higher and further south in winter than in summer, according, by thermal balance, to the seasonal variation of the pole-to-equator gradient. Some biases are present, however. In both winter and summer, the modelled jet stream axis lies within the 100-150 mb layer whereas the observed axis is centred around 200 mb. The modelled jet oscillates between 40°N (winter) and 50°N (summer) while the observed jet changes between 30°N (winter) and 40°N (summer). The intensity of the jet is accurately simulated in winter, reaching 40 m s^{-1}, but is 8 m s^{-1} higher than the observed value (16 m s^{-1}) in summer.

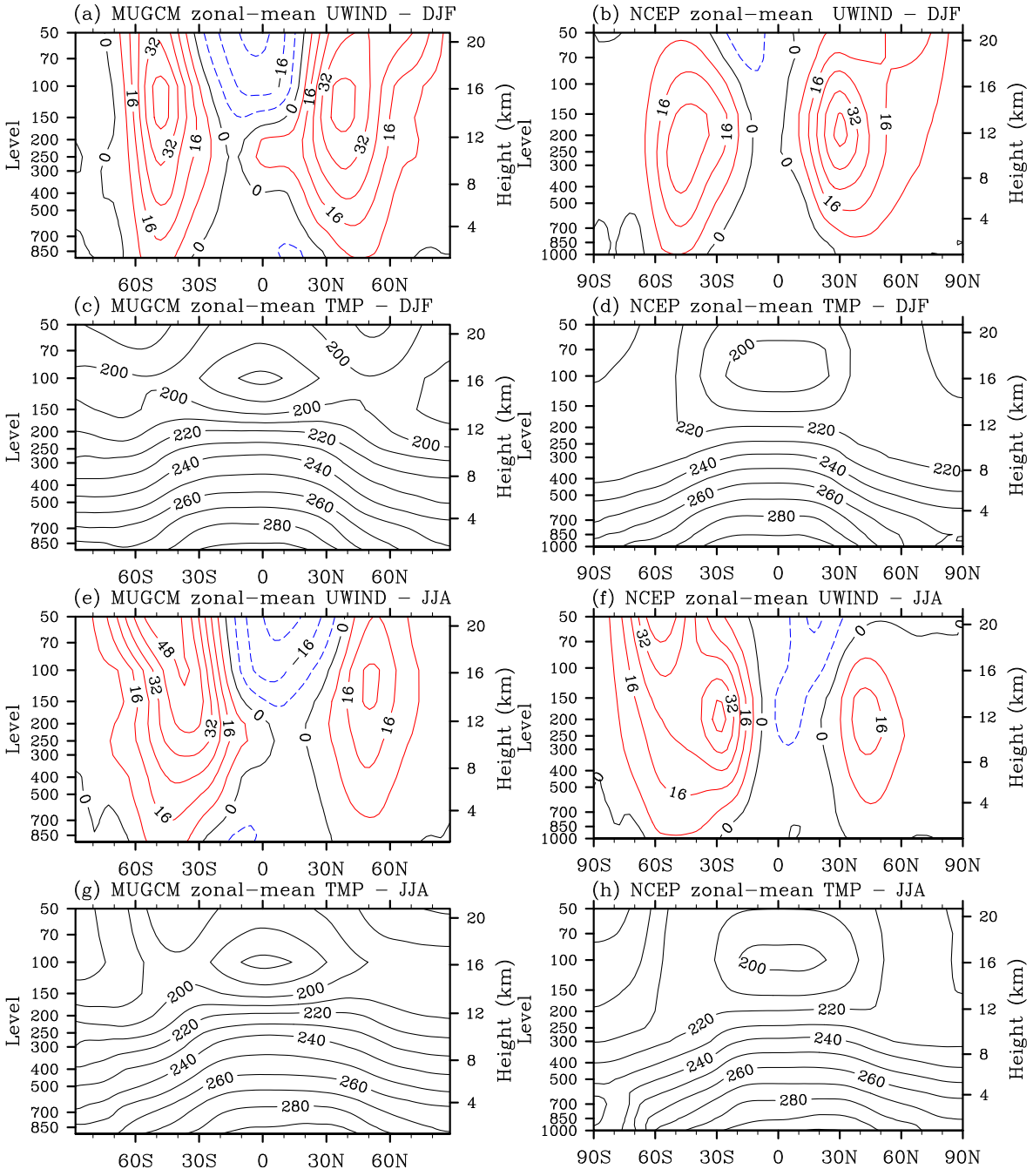

Figure 3.8: Meridional cross sections of zonal-mean zonal wind (UWIND) and temperature (TMP) for (a-d) winter and (e-f) summer climatologies computed from MUGCM (left panels) and NCEP (right panels) data. Isotach spacing of 8 m s^{-1} and isotherm spacing of 10° K.

It also worth to note that, during winter, the simulated lower tropospheric zonal-mean zonal wind is stronger than observed.

3.2.1 Zonal asymmetries. Jet streams

Once the level of the zonal-mean modelled jet stream has been identified (see last section), its zonal asymmetries can be diagnosed. Figure 3.9 presents winter and summer climatologies of upper tropospheric zonal wind. Modelled and observed wind contours are plotted at 150 mb and 200 mb, respectively.

The model is able to simulate the three major departures from the zonal-mean zonal circulation observed in winter, located over eastern Asia and western Pacific (*Pacific jet stream* or *Asian jet stream*), over eastern North America and western North Atlantic (*Atlantic jet stream* or *North America jet stream*), and also over northeastern Africa and Middle East. For easier comparison between modelled and observed jets and their seasonal changes, the location and maximum speed of the jets in winter and summer are presented in table 3.2.

Both observed Pacific and Atlantic jet streams share the main features of the zonal-mean jet, namely its equatorward displacement and intensification in speed from summer to winter. This behaviour is also present in the modelled Pacific and Atlantic jets. See details in table 3.2. In particular, both modelled and observed Pacific and Atlantic jets have the same position of the corresponding zonal-mean jet in summer (with the exception of the observed Atlantic jet which is 5º to the north), but are 5º further north in winter.

Both Pacific and Atlantic modelled jet streams are biased 10º to the north in winter. In summer, the Pacific (Atlantic) jet is displaced 10º (5º) to the north.

In winter, the observed Pacific jet stream is located around 140ºE reaching 60 m s^{-1}, retrograding to 90ºE and weakening to 30 m s^{-1} in summer. The modelled jet is

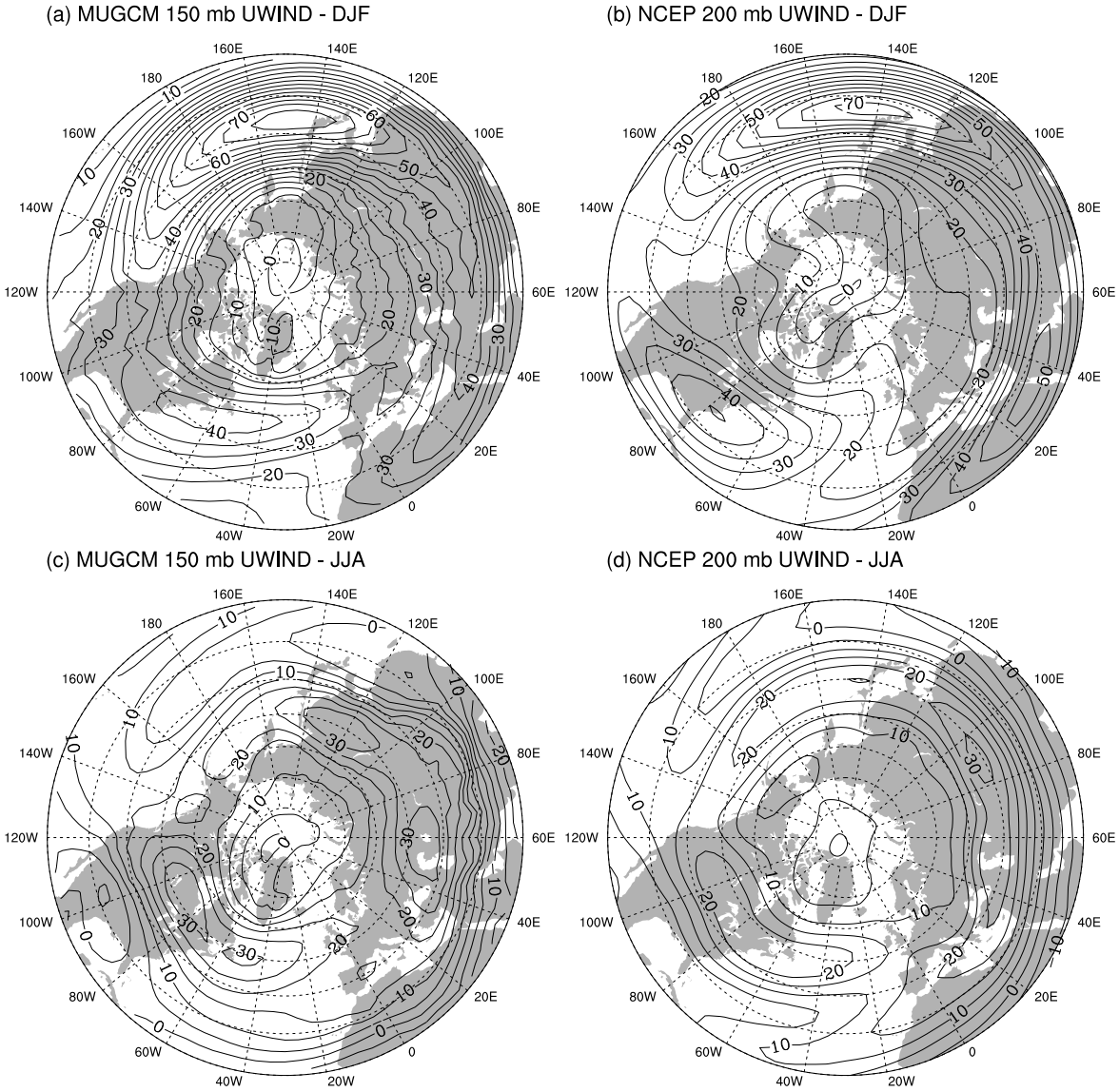

Figure 3.9: Winter (upper panels) and summer (lower panels) climatologies of upper tropospheric zonal wind (UWIND) computed from MUGCM (left panels) and NCEP (right panels) data. Isotach spacing of 5 m s^{-1}. Grid latitude (longitude) lines represented every 10° (20°).

biased 10° (30°) downstream in winter (summer). The speed of the observed jet is well simulated by the model in summer but it is 15 m s^{-1} stronger in winter.

The Atlantic jet stream is located around 80°W reaching 40 m s^{-1} in winter, progressing to 70°W and weakening to 20 m s^{-1} in summer. The modelled jet is biased 20° downstream in winter and about 10° upstream in summer. The speed of the observed jet is well reproduced by the model in winter but it is overestimated 10 m s^{-1} in summer.

Table 3.2: Location and speed of zonal-mean, Pacific and Atlantic jet streams.

	Location	(lat,lon)	Speed	(m s^{-1})
	DJF	JJA	DJF	JJA
NCEP $\bar{u}_{200}$	(30N,-)	(40N,-)	40	16
MUGCM $\bar{u}_{150}$	(40N,-)	(50N,-)	40	24
NCEP Pacific u_{200}	(35N,140E)	(40N,90E)	60	30
MUGCM Pacific u_{150}	(45N,150E)	(50N,120E)	75	30
NCEP Atlantic u_{200}	(35N,80W)	(45N,70W)	40	20
MUGCM Atlantic u_{150}	(45N,60W)	(50N,80W)	40	30

3.2.2 Orographic and thermal forcing

It seems clear from the last section that departures of upper tropospheric zonal wind from zonal symmetry are linked to the distribution of continents and oceans. In particular, the pronounced departures over eastern Asia and eastern North America (the Pacific and Atlantic jet streams) are linked to the Hymmalias and Rocky mountains, respectively.

Asymmetries of upper tropospheric westerly flow can also be inferred from upper tropospheric geopotential heights. In the upper panels of Figure 3.10, modelled and observed winter climatologies of geopotential height at 150 mb and 200 mb, respectively,

are presented. They exhibit a planetary wave pattern with a preferred wave-number-two configuration with two troughs (just to the east of the asian and american continents) and two ridges (just to the west of Europe and North America).

Atmospheric waves are usually analysed using the *linear perturbation* method in which all quantities are decomposed into a *basic state* portion and a *perturbation* portion. For stationary planetary waves, the basic state and the perturbation portion of a quantity A are taken as the zonal-mean $\bar{A}$ and the longitudinal deviation from this zonal-mean A', referred as *eddy*, respectively, such that $A = \bar{A} + A'$. The eddy wave components of the stationary planetary waves shown in the upper panels of figure 3.10 are shown in the lower panels of the same figure. Meridional averages between 30°N and 60°N of these eddy fields are shown in figure 3.11. The observed eddy winter climatology has an hemispheric wave-pattern of alternating lows and highs: two lows just to the east of the asian and american continents, and two highs just to the west of Europe and North America. The highs are well reproduced by the model but the asian low is too deep and, consequently, the american low vanishes.

The eddy wave patterns of figure 3.10 are in qualitatively agreement with the conservation of Ertel's potential vorticity,

$$P = (\zeta_\theta + f)\Big(-g\frac{\partial\theta}{\partial p}\Big), \tag{3.9}$$

where ζ_θ is the vertical component of the relative vorticity evaluated on an isentropic surface, by a steady westerly adiabatic and frictionless flow over a large-scale mountain barrier. As the air column begins to cross the barrier its vertical extent decreases ($-\partial\theta/\partial p > 0$) and must acquire anticyclonic vorticity ($\zeta_\theta < 0$) turning equatorward. When the air column has passed over the mountain and return to its original depth it will be south of its original latitude so that f will be smaller and the flow must acquire cyclonic vorticity ($\zeta_\theta > 0$) deflecting poleward. When the flow returns to its original latitude, it still have a poleward velocity component and will continue poleward gradually acquiring anticyclonic curvature until its its direction is again reversed. Therefore, the resulting flow will have a wave-like pattern with a trough immediately downstream

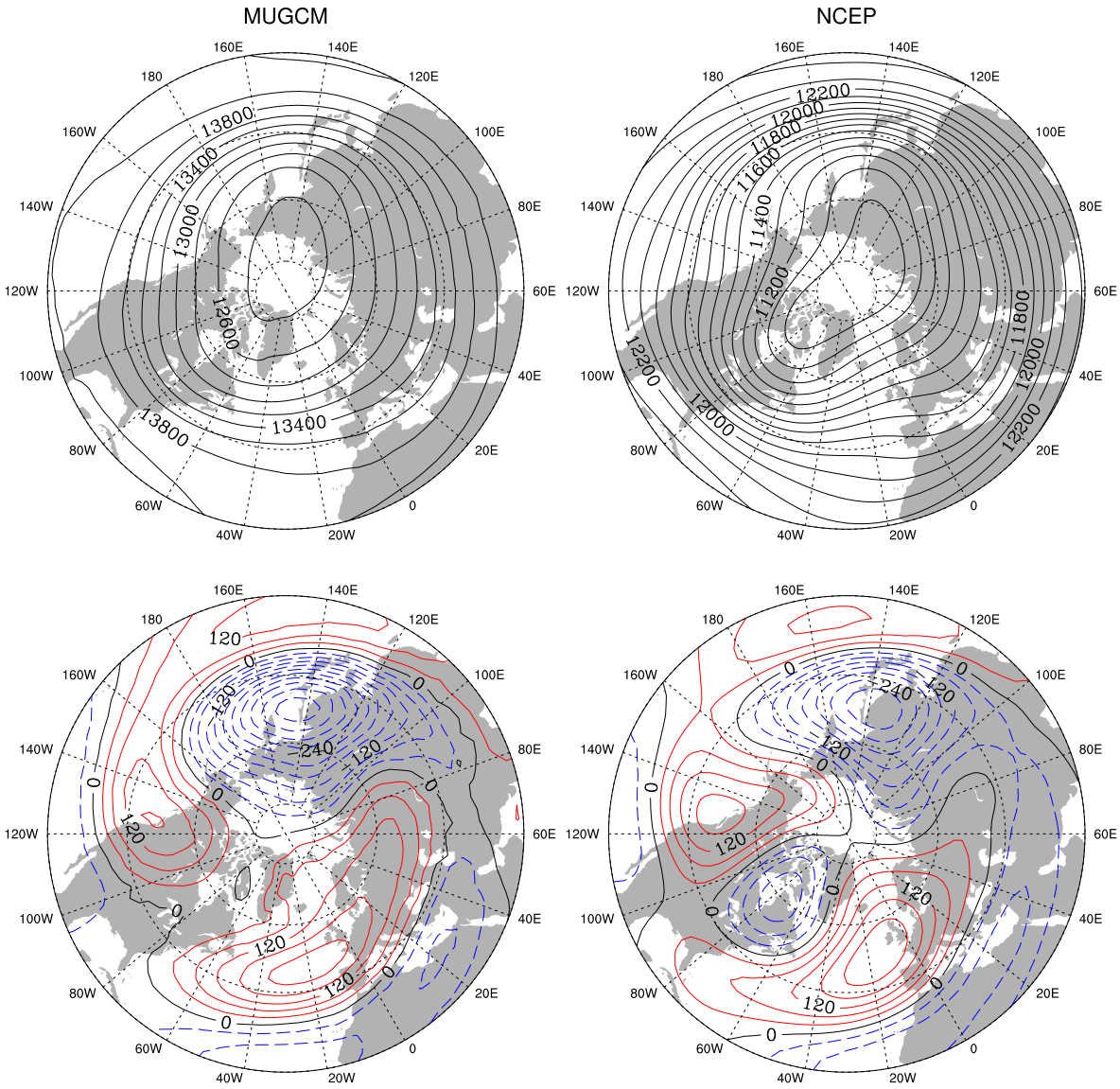

Figure 3.10: Winter climatology of MUGCM 150 mb HGT (upper left panel) and NCEP 200 mb HGT (upper right panel), and corresponding eddy component (lower panels). Isolines spaced by 200 gpm (upper panels) and 50 gpm (lower panels). Grid lines represented every 20°.

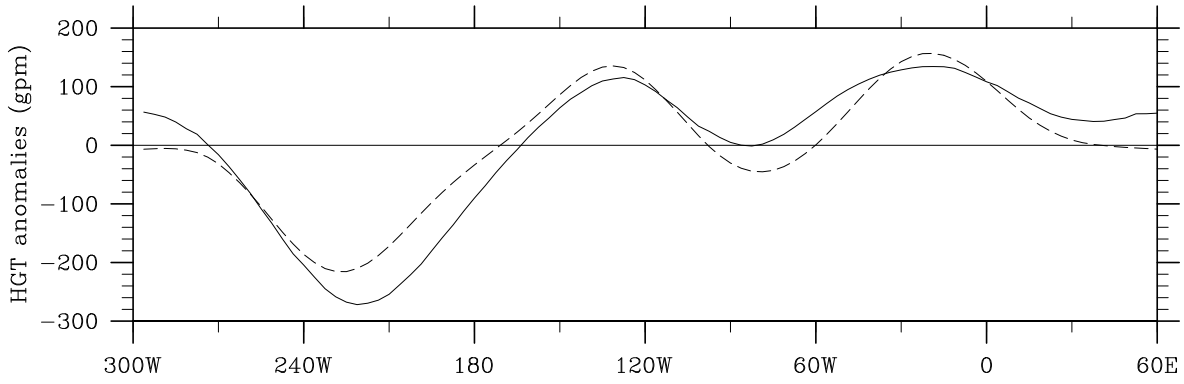

Figure 3.11: Meridional average between 30°N and 60°N of upper tropospheric HGT eddy climatology (shown in the lower panels of figure 3.10). MUGCM (NCEP) height anomalies in gpm plotted with a solid line (dashed line).

the barrier (the lee side) and ridge further downstream.

The effect of orography upon the stationary planetary-scale flow field in the atmosphere was first studied by Charney and Eliassen [1949]. Using a one-dimensional barotropic model, they concluded that the large-scale quasi-stationary disturbances of the middle latitudes are produced mainly by forced ascent of the westerly flow over the continental land masses. On the other hand, Smagorinsky [1953] evaluated the effects of heat sources and sinks upon stationary flow using a baroclinic model, concluding that the thermal effects account for the essential features of the observed sea level pressure, whereas the relative importance of mountain effects increases with increasing altitude.

Derome and Wiin-Nielsen [1971] used a β-plane quasi-geostrophic model to study the response to the effects of earth topography and diabatic heating for January 1962, near 45°N. They showed that the standing waves forced by topography are in about the same position as those forced by the diabatic heating, and that the former have somewhat larger amplitudes than the latter. They also showed that the troughs tend to occur near the regions of large-scale heating and the ridges consequently occur near the regions of large-scale cooling. This result agrees with the patterns of figure 3.10 (lower panels) and corresponding middle latitude meridional averages (figure 3.11), and

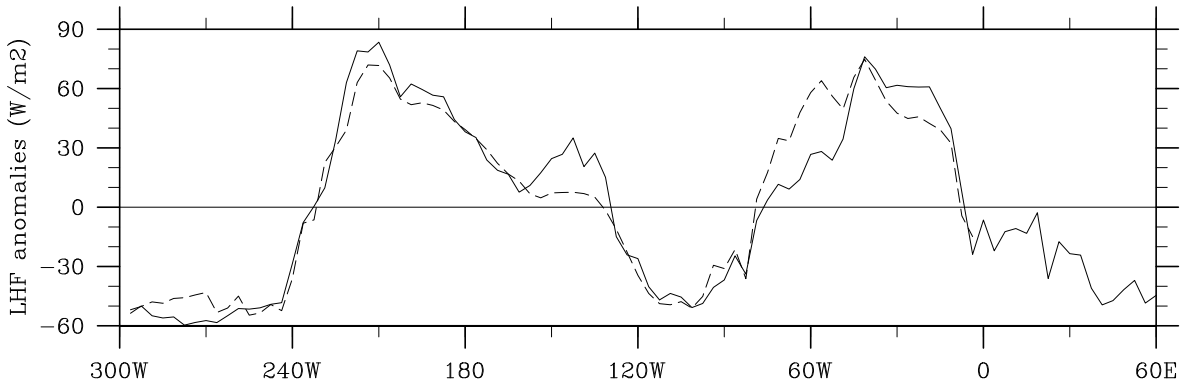

Figure 3.12: Meridional average between 30°N and 60°N of latent heat flux (LHF) eddy climatology. MUGCM (NCEP) anomalies in W m^{-2} plotted with a solid line (dashed line).

to the winter middle latitude zonal distribution of latent heat flux (LHF) presented in figure 3.12.

General circulation models have also been used to try to explain how the large-scale atmospheric standing waves are maintained against dissipative forces. Manabe and Terpstra [1974] ran a general circulation model with and without mountains and confirmed that it is necessary to consider the effects of mountains for the successful simulation of the stationary flow field in the atmosphere, particular in the upper troposphere and stratosphere as indicated by Smagorinsky [1953].

3.3 Winter climatology of the atmospheric circulation cells

A theorem of Helmholtz states that any horizontal velocity field $\mathbf{V}$ can be divided into a *nondivergent* (or *rotational*) part $\mathbf{V}_\Psi$ and a *irrotational* (or *divergent*) part $\mathbf{V}_\Phi$:

$$\mathbf{V} = \mathbf{V}_\Psi + \mathbf{V}_\Phi, \tag{3.10}$$

where

$$\nabla \cdot \mathbf{V}_\Psi = 0 \tag{3.11}$$

and

$$\nabla \times \mathbf{V}_\Phi = 0. \tag{3.12}$$

From equation 3.12 follows that

$$\frac{\partial v_\Phi}{\partial x} = \frac{\partial u_\Phi}{\partial y} \tag{3.13}$$

where u_Φ and v_Φ are the components of the horizonal irrotational velocity $\mathbf{V}_\Phi$. The last equation implies that $u_\Phi dx + v_\Phi dy$ is an exact differential and hence we can define

$$d\Phi = u_\Phi dx + v_\Phi dy, \tag{3.14}$$

from which we have

$$u_\Phi = \frac{\partial \Phi}{\partial x} \quad \text{and} \quad v_\Phi = \frac{\partial \Phi}{\partial y},$$

that is,

$$\mathbf{V}_\Phi = \nabla \Phi, \tag{3.15}$$

which means that Φ is the *velocity potential*.

It is well known that Hadley and Walker cells are *thermally direct* cells, that is, they are thermally driven [Holton, 1992]. Atmospheric heating associated with convection induces atmospheric convergence/divergence that drives atmospheric vertical motion and circulation. Therefore, Hadley and Walker cells are usually described by the divergent part of the horizontal wind, $\mathbf{V}_\Phi$, and by vertical motion [Hastenrath, 2001; Wang, 2002b,a].

Here the three-dimensional circulation is assessed using the divergent component, $\mathbf{V}_\Phi$, of the horizontal wind, referred here as divergent wind (DIV WIND), at the lower troposphere and upper troposphere, and vertical velocity (VVEL) at the middle troposphere. The pressure levels of 800 mb, 500 mb and 200 mb are taken to represent the lower, middle at upper troposphere. Figure 3.13 (a-d) shows winter climatologies of the DIV WIND at 200 and 850 mb, along with the associated, by equation 3.15, velocity potential (VEL POT) climatologies. VVEL (figure 3.13 (e,f)) is taken here to be the negative of the pressure vertical velocity climatologies at 500 mb. Figure 3.13 (g,h) shows the precipitation climatology, which is a proxy of convective upward motion in

the tropics. NCEP precipitation results from convection activity only, but the MUGCM map shows total precipitation, and therefore, must be taken as convective only in the tropical belt.

The model reproduces reasonably well the divergent circulation in winter. Centres of high positive (low negative) velocity potential are centres of strong convergence (divergence). Figure 3.13 shows that centres of strong convergence (divergence) at the lower troposphere are mirrored by centres of strong divergence (convergence) at the upper troposphere, and connected by strong upward (downward) vertical motion. The general three-dimensional structure of the divergent circulation is well reproduced by the model, but with larger amplitudes.

The meridional overturning circulation is longitudinally asymmetric and, because of this, zonal sectors were defined using the information given by the maps of figure 3.13. These are presented in table 3.3 along with sectors used by other authors. In section 3.3.1, the meridional overturning circulation in each sector is described in terms of the zonal-mean[3] meridional mass streamfunction (streamfunction in the meridional-vertical plane). The Walker circulation is described using the meridional-mean[4] zonal mass streamfunction (streamfunction in the zonal-vertical plane) in section 3.3.2.

3.3.1 Hadley and Ferrel cells

Figure 3.13 shows that the meridional overturning, represented by the divergent meridional circulation at 850 and 200 mb and by the vertical velocity at 500 mb, is longitudinally asymmetric. Consequently, the meridional cells (Hadley and Ferrel) are also longitudinally asymmetric. Because of this, zonal sectors with approximately zonal symmetry must be defined in order to capture the zonal variations of the overturning circulation. These zonal sectors, bounded by longitudes λ_1 and λ_2 ($\lambda_2 > \lambda_1$), are defined such that the zonal-mean meridional circulation in the sector approximately satisfies the continuity equation, written here in log-pressure coordinates whose vertical

[3]Zonal average over the longitude range of the sector.

[4]Meridional average over the entire equatorial band.

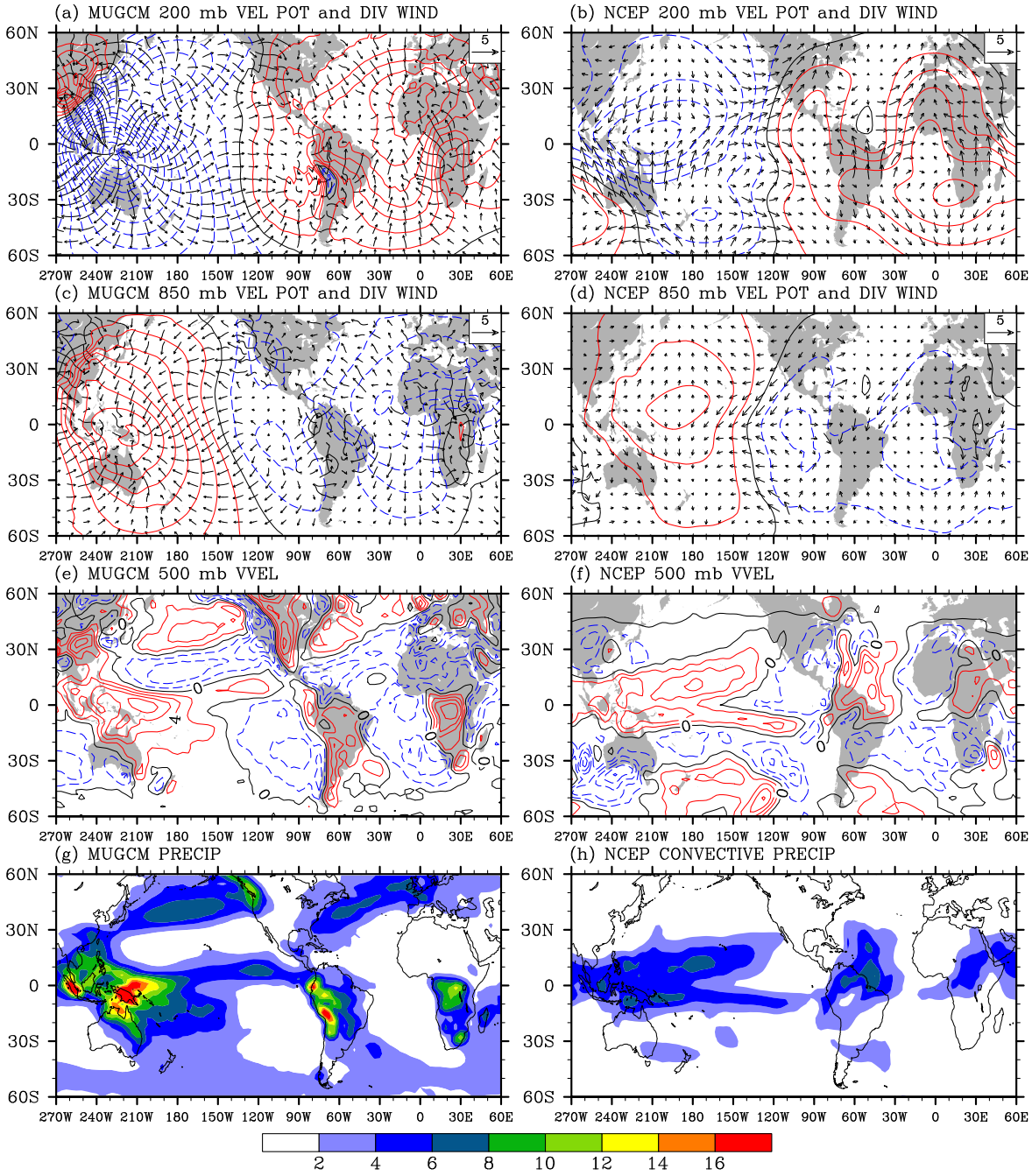

Figure 3.13: MUGCM (left panels) and NCEP (right panels) winter climatology of velocity potential (VEL POT) and divergent wind (DIV WIND) at 200 mb (a,b) and 850 mb (c,d), vertical velocity (VVEL) at 500 mb (e,f), and total (g) and convective (h) precipitation (PRECIP) (g,h). Negative, zero and positive contours in blue, black and red. Divergent wind in m.s^{-1}. Isoline spacings are 2×10^6 m^2.s^{-1} for velocity potential and 2×10^{-4} mb.s^{-1} for vertical velocity. Vertical velocity is taken to be negative of the pressure vertical velocity, i.e., positive values indicate upward motion.

Table 3.3: Zonal sectors of meridional mass overturning: Indonesia (IND), West Pacific (WPAC), East Pacific (EPAC), West Atlantic (WATL) and East Atlantic (EATL).

	IND	WPAC	EPAC	WATL	EATL
This work	270:220W	210:160W	150:100W	70:40W	20W:10E
Wang [2002b]	-	240:190W	150:100W	-	-
Wang [2002a]	-	-	-	60:40W	10W:10E
Hastenrath and Lamb [2004]	-	180:150W	120:90W	-	30W-0
Trenberth et al. [2000]	-	-	170:90W	-	30W:10E

coordinate is $z^* = -H\ln(p/p_s)$ where $H = RT_s/g$ is the scale height, p_s is a standard reference pressure (1000 mb), T_s is a global average temperature and g is gravity,

$$\frac{\partial \rho_o \bar{v}}{\partial y} + \frac{\partial \rho_o \bar{w}^*}{\partial z^*} \simeq 0, \tag{3.16}$$

where is $\bar{v}$ and $\bar{w}^*$ are the zonal averages, from λ_1 and λ_2, of the meridional and vertical velocity respectively, and $\rho_o \equiv \rho_o(z^*)$ is density.

Neglecting the errors resulting from this approximation, the continuity equation 3.16 implies that $-\rho_o \bar{v} dz^* + \rho_o \bar{w}^* dy$ is an exact differential and hence we can define

$$d\Pi = -\rho_o \bar{v} dz^* + \rho_o \bar{w}^* dy, \tag{3.17}$$

from which we have

$$\rho_o \bar{v} = -\frac{\partial \Pi}{\partial z^*} \quad \text{and} \quad \rho_o \bar{w}^* = \frac{\partial \Pi}{\partial y}, \tag{3.18}$$

or changing to pressure coordinates,

$$\bar{v} = g\frac{\partial \Pi}{\partial p} \quad \text{and} \quad \bar{\omega} = -g\frac{\partial \Pi}{\partial y}. \tag{3.19}$$

Integrating the first equation of 3.19, where $\bar{v} \equiv \bar{v}(p, \theta)$ with θ denoting latitude, from p to p_s, yields

$$\Pi(p, \theta) = \frac{1}{g}\int_p^{p_s} \bar{v}(p', \theta) dp', \tag{3.20}$$

whose units are kg s^{-1} m^{-1}, that is, it represents the rate of mass overturning per unit zonal-length. The total rate of mass overturning in a sector bounded by longitudes λ_1 and λ_2 ($\lambda_2 > \lambda_1$), with a zonal-length of $a\cos\theta(\lambda_2 - \lambda_1)$, is

$$\Pi_{[\lambda_1,\lambda_2]}(p,\theta) = \frac{a\cos\theta(\lambda_2 - \lambda_1)}{g}\int_p^{p_s} \bar{v}(p',\theta)dp', \tag{3.21}$$

whose units are kg s^{-1}.

Modelled and observed monthly mean time series of $\bar{v}$, averaged over the longitude range of each sector of table 3.3, were used to compute equation 3.21. The resulting monthly mean time series of the zonal-mean meridional mass streamfunction were then used to calculate the winter climatologies presented in figure 3.14.

Figure 3.14 clearly shows that the modelled meridional mass circulation is overestimated. In particular, the thermally direct Hadley cells are too intense, which may have important consequences on the extratropical atmospheric response to tropical SST forcing. In section 3.2.1 it was found that the Pacific jet stream is overestimated by the model. We see now that this may be a consequence of the intensified Pacific Hadley cells.

3.3.2 Walker cells

The Walker circulation is described here using the zonal circulation averaged over the equatorial band, which approximately satisfies the continuity equation, in log-pressure coordinates,

$$\frac{\partial \rho_o \bar{u}}{\partial x} + \frac{\partial \rho_o \bar{w}^*}{\partial z^*} \simeq 0, \tag{3.22}$$

where is $\bar{u}$ and $\bar{w}^*$ are the meridional averages, taken over an equatorial band, of the zonal and vertical velocity respectively, and $\rho_o \equiv \rho_o(z^*)$ is density.

Neglecting the errors resulting from this approximation, follows from the continuity equation 3.22 that $-\rho_o \bar{u} dz^* + \rho_o \bar{w}^* dx$ is an exact differential and hence we can define

$$d\Xi = -\rho_o \bar{u} dz^* + \rho_o \bar{w}^* dx, \tag{3.23}$$

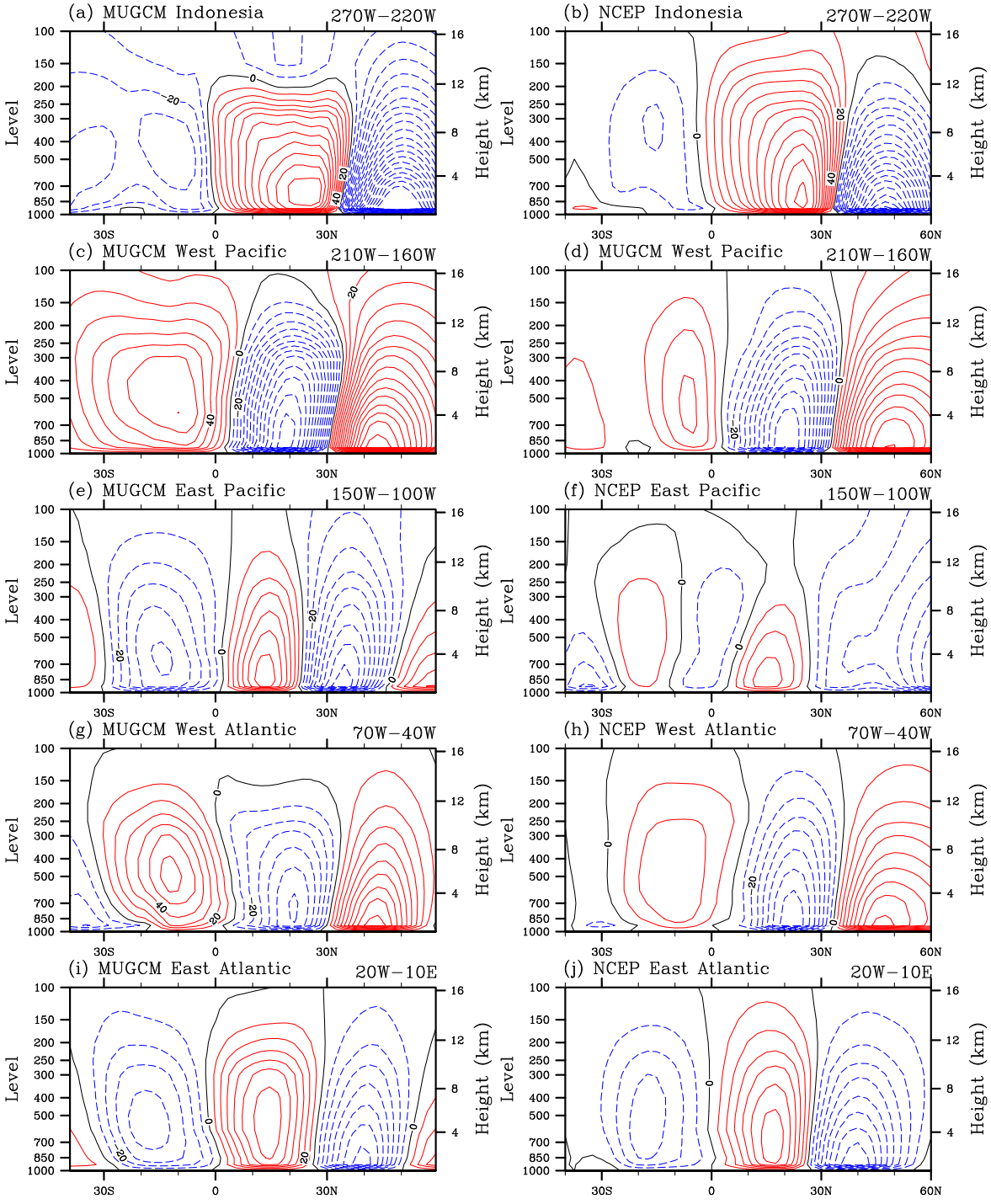

Figure 3.14: Winter climatology of meridional mass streamfunction zonally averaged in the sectors defined in table 3.3. Negative, zero and positive contours in dashed blue, solid black and solid red. Isoline spacing is 20×10^9 kg s^{-1}.

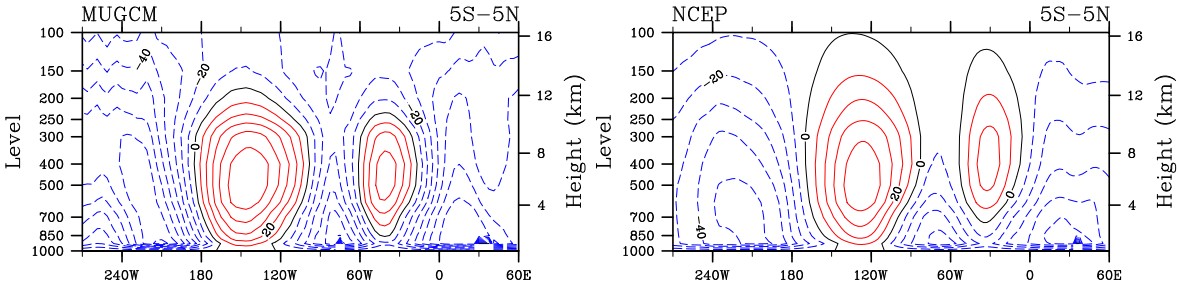

Figure 3.15: Winter climatology of zonal mass streamfunction averaged from 5°S to 5°N. Negative, zero and positive contours in dashed blue, solid black and solid red. Isoline spacing is 20×10^9 kg s^{-1}.

from which we have

$$\rho_o \bar{u} = -\frac{\partial \Xi}{\partial z^*} \quad \text{and} \quad \rho_o \bar{w}^* = \frac{\partial \Xi}{\partial x}, \tag{3.24}$$

or changing to pressure coordinates,

$$\bar{u} = g\frac{\partial \Xi}{\partial p} \quad \text{and} \quad \bar{\omega} = -g\frac{\partial \Xi}{\partial x}. \tag{3.25}$$

Integrating the first equation of 3.25, where $\bar{u} \equiv \bar{u}(p, \lambda)$ with λ denoting longitude, from p to p_s yields

$$\Xi(p, \lambda) = \frac{1}{g}\int_p^{p_s} \bar{u}(p', \lambda)dp', \tag{3.26}$$

whose units are kg s^{-1} m^{-1}, that is, it represents the rate of mass overturning per unit meridional-length. The total rate of mass overturning in the equatorial band $\Delta\theta$, with a meridional-length of $a\Delta\theta$, is

$$\Xi_{\Delta\theta}(p, \lambda) = \frac{a\Delta\theta}{g}\int_p^{p_s} \bar{u}(p', \lambda)dp', \tag{3.27}$$

whose units are kg s^{-1}.

Modelled and observed monthly mean time series of $\bar{u}$, averaged from 5°S-5°N, were used to compute equation 3.27. The resulting monthly mean time series of the meridional-mean zonal mass streamfunction were then used to calculate winter climatologies presented in figure 3.15.

Figure figure 3.13 shows strong low-level convergence over western Pacific (the Pacific warm pool) associated with strong convection and precipitation, and upper-level divergence. In the eastern Pacific, upper-level convergence and lower-level divergence are connected with downward vertical velocity. The circulation is closed with low-level easterlies and upper-level westerlies. In the equatorial Atlantic, an analogous circulation, although weaker and less clear, is also present. This type of zonal circulation is the so called Walker circulation first identified in the equatorial Pacific ocean [Bjerknes, 1969; Julian and Chervin, 1978; Trenberth et al., 1998; Wang, 2002b; Hastenrath and Lamb, 2004] and latter in the equatorial Atlantic ocean [Trenberth et al., 1998; Hastenrath, 2001; Wang, 2002a; Hastenrath and Lamb, 2004]. Figure 3.15 shows that the mass overturning in the zonal-vertical plane, associated with the Pacific and Atlantic Walker cells, is reproduced by the MUGCM but with higher amplitude in both oceans.

3.4 Variability of the Euro-Atlantic atmosphere

In this section, MUGCM's variability is compared to the observed variability using two different methodologies. In subsection 3.4.1, the correlation at each grid point of the MUGCM's grid in the EA sector, between the MUGCM-AMIP II ensemble mean and NCEP MSLP, TMP-SFC, PRECIP, HGT-850, HGT-500 and HGT-200 time series, is determined for each of the four standard seasons. In subsection 3.4.2, we compare the spatial patterns of variability of MUGCM and NCEP seasonal mean *anomalies* of MSLP, in the EA region, obtained by Principal Component Analysis (PCA).

3.4.1 Correlation patterns

The agreement between the MUGCM and NCEP time variability is here measured by the correlation between the MUGCM ensemble mean and NCEP time series of the AMIP II time range (see table 2.1), at each grid point of the MUGCM's grid in the EA sector. Note that the NCEP data was regridded to the MUGCM's grid, using bilinear interpolation, prior to the computation of the correlations. The correlations were determined for MSLP, TMP-SFC, PRECIP, HGT-850, HGT-500 and HGT-200,

at each of the four standard seasons. The patterns of correlations are presented in figures 3.16 to 3.21.

The correlations between between MUGCM ensemble mean and NCEP time series of seasonal means at each grid point of a particular variable at a particular season were subjected to the hypothesis test (see, for example, Milton and Arnold [1995]),

$$H_0 : \rho = 0 \ vs \ H_1 : \rho > 0, \tag{3.28}$$

where ρ is the population (*true*) correlation between the MUGCM ensemble mean and NCEP time series at a particular grid point. Note that a right-tailed test is used instead of a two-tailed test because the rejection of the null hypothesis of the right-tailed test implies that the correlation is positive.[5]

The test statistic of this test has, invoking again the Central Limit Theorem, a t-Student distribution with $n-2$ degrees of freedom,

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}, \tag{3.29}$$

where r is the sample correlation and $n = 17$ is the number of observations (seasonal means). Note that the seasonal means are independent observations, as required by the test. The criterion to reject the null hypothesis, at a significance level of α, is

$$t > t_{n-2,1-\alpha}, \tag{3.30}$$

where t[6] is a realisation of the test statistic T of equation 3.29.

Figures 3.16 to 3.21 show the correlation between the model ensemble-mean and the NCEP time series, at each grid point of the EA region, of MSLP, TMP-SFC, PRECIP, HGT-850, HGT-500 and HGT-200 in winter, spring, summer and autumn. Regions

[5] A two-tailed test could also be used. The rejection of the null hypothesis of a two-tailed test implies that the correlation is different from zero, and thus may be positive or negative. In this case, the grid points where the model simulates well the observed variability are those grid points where the null hypothesis is rejected *and* where the correlation is positive.

[6] In equation 3.30, $t_{n-2,1-\alpha}$ is such that $P[T \leq t_{n-2,1-\alpha}] = 1-\alpha$.

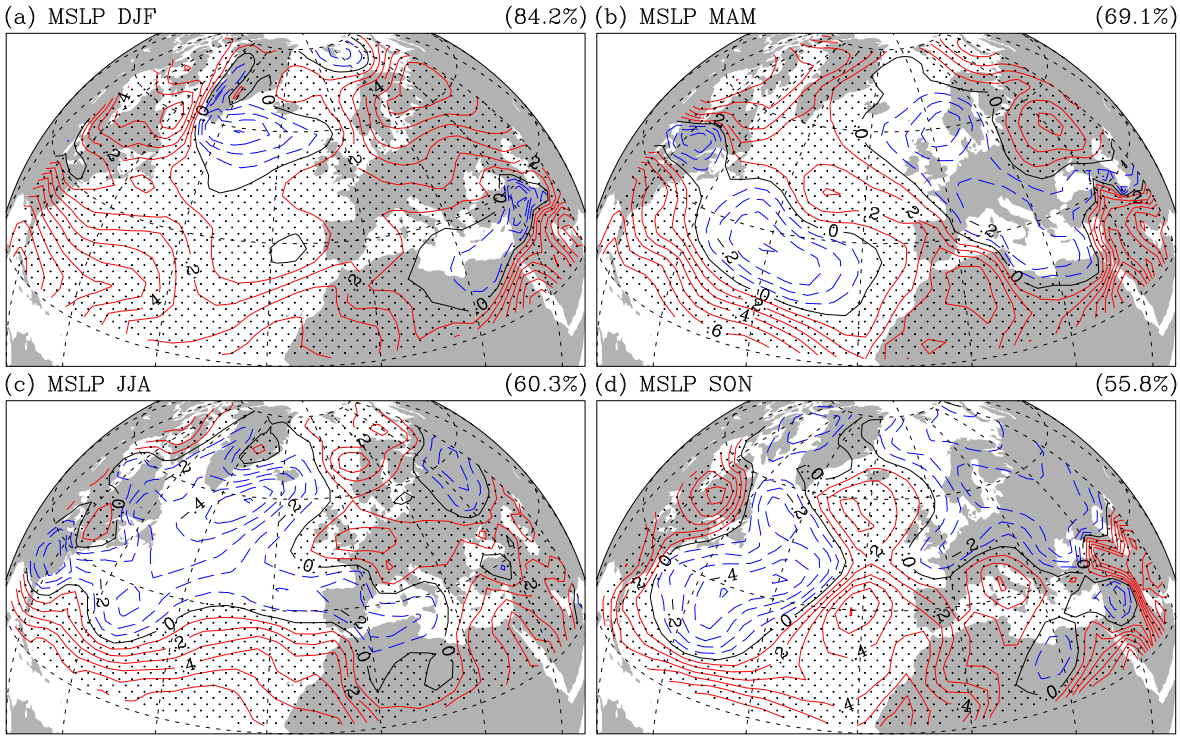

Figure 3.16: Correlations between MUGCM ensemble-mean and NCEP seasonal mean time series of MSLP in the EA region. (a) DJF, (b) MAM, (c) JJA and (d) SON. Isoline spacing of 0.1. The area is stippled where the correlation is significantly greater than zero, at a significance level of 0.01 (hypothesis test 3.28). The percentage of total area with significant positive correlation is printed on the top right-hand corner of each panel. Grid lines represented every 20°.

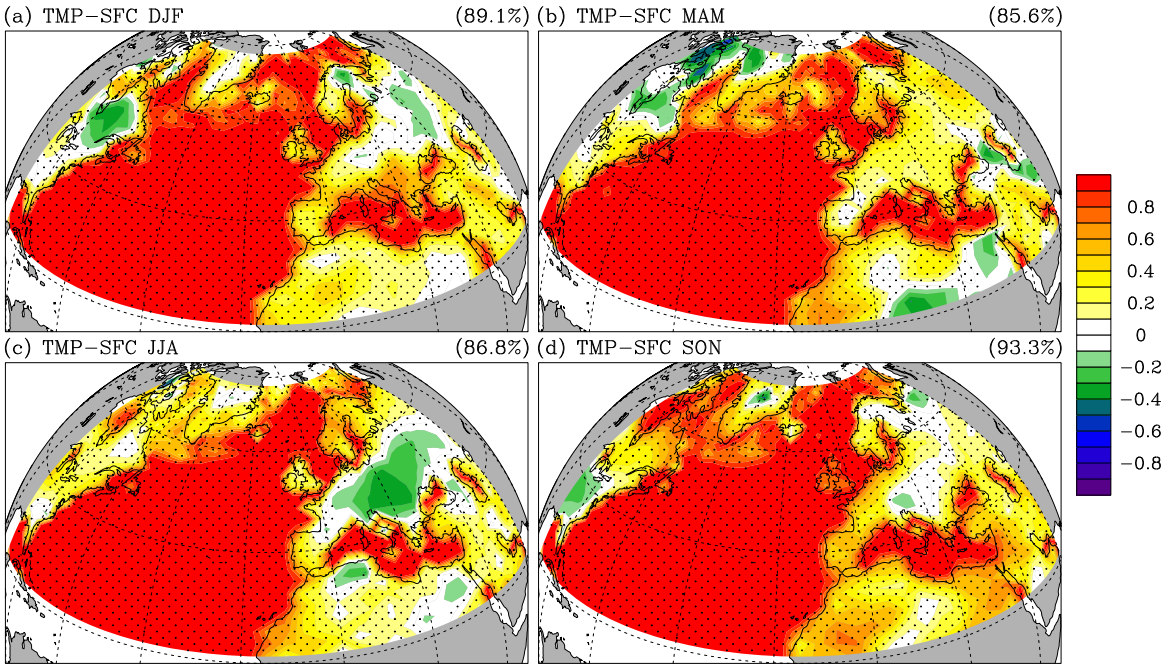

Figure 3.17: As in 3.16 but for TMP-SFC.

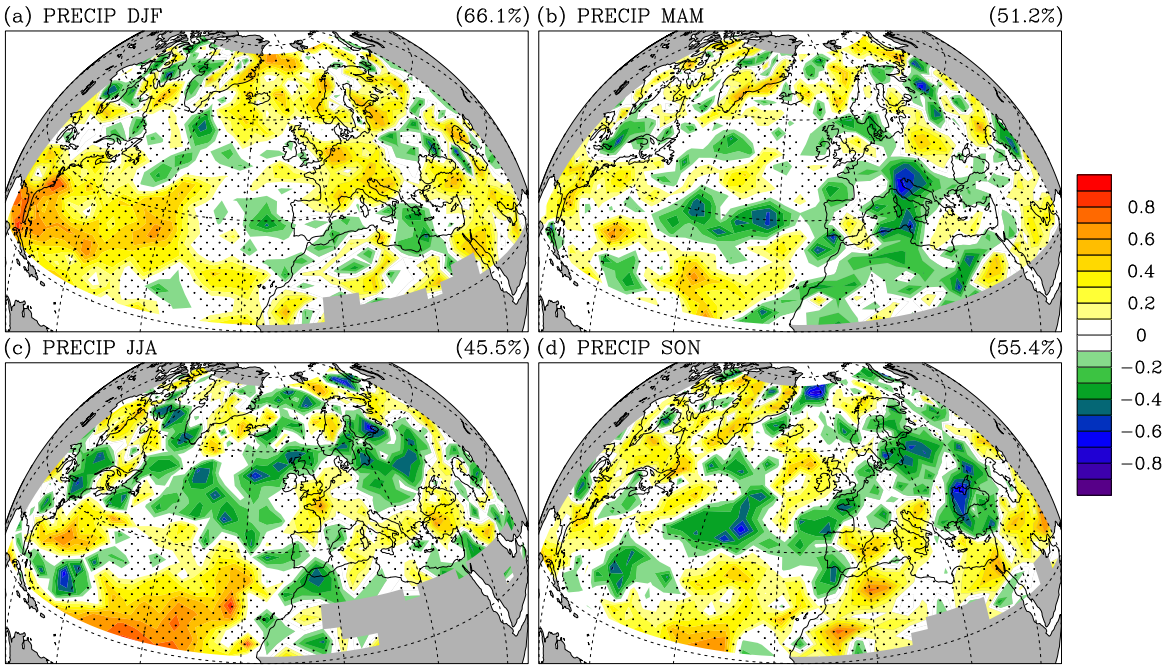

Figure 3.18: As in 3.16 but for PRECIP.

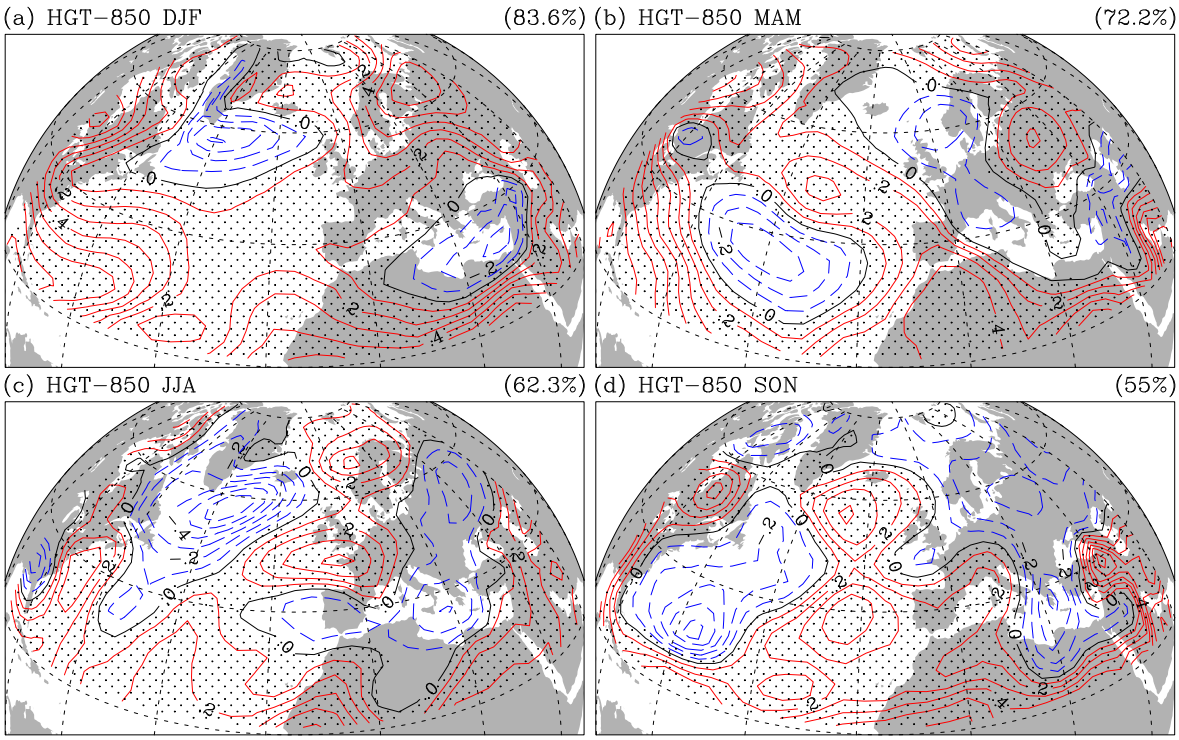

Figure 3.19: As in 3.16 but for HGT-850.

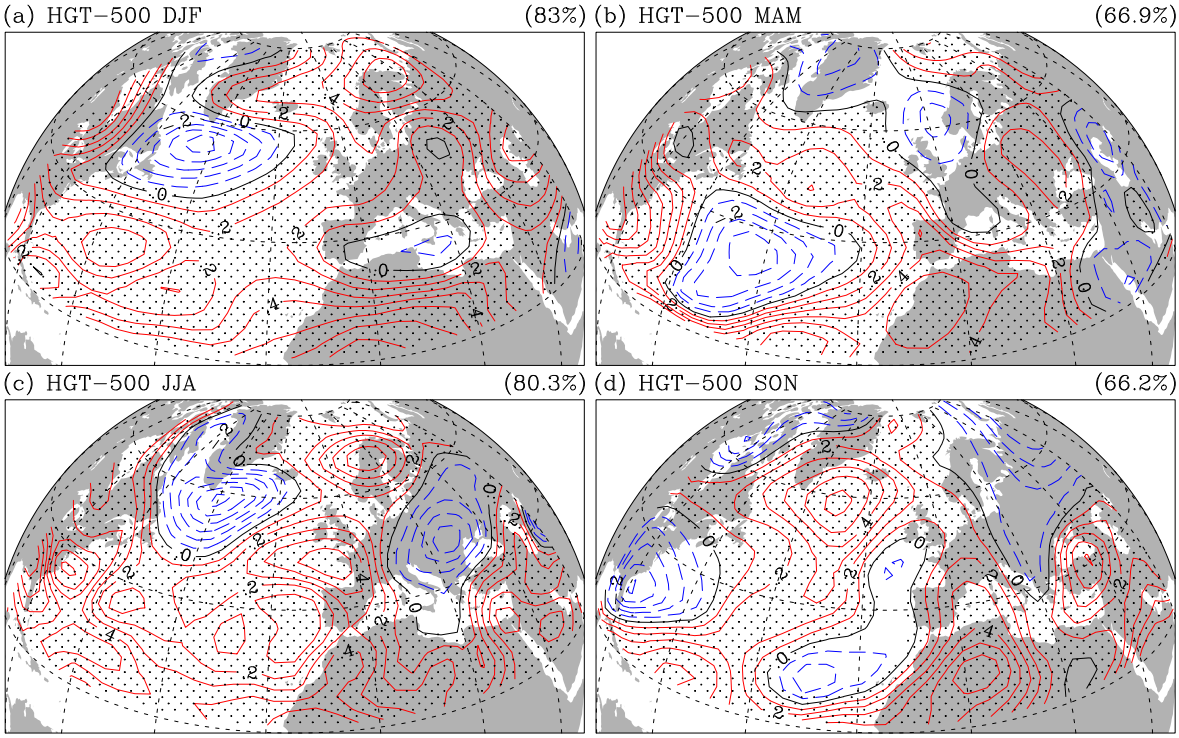

Figure 3.20: As in 3.16 but for HGT-500.

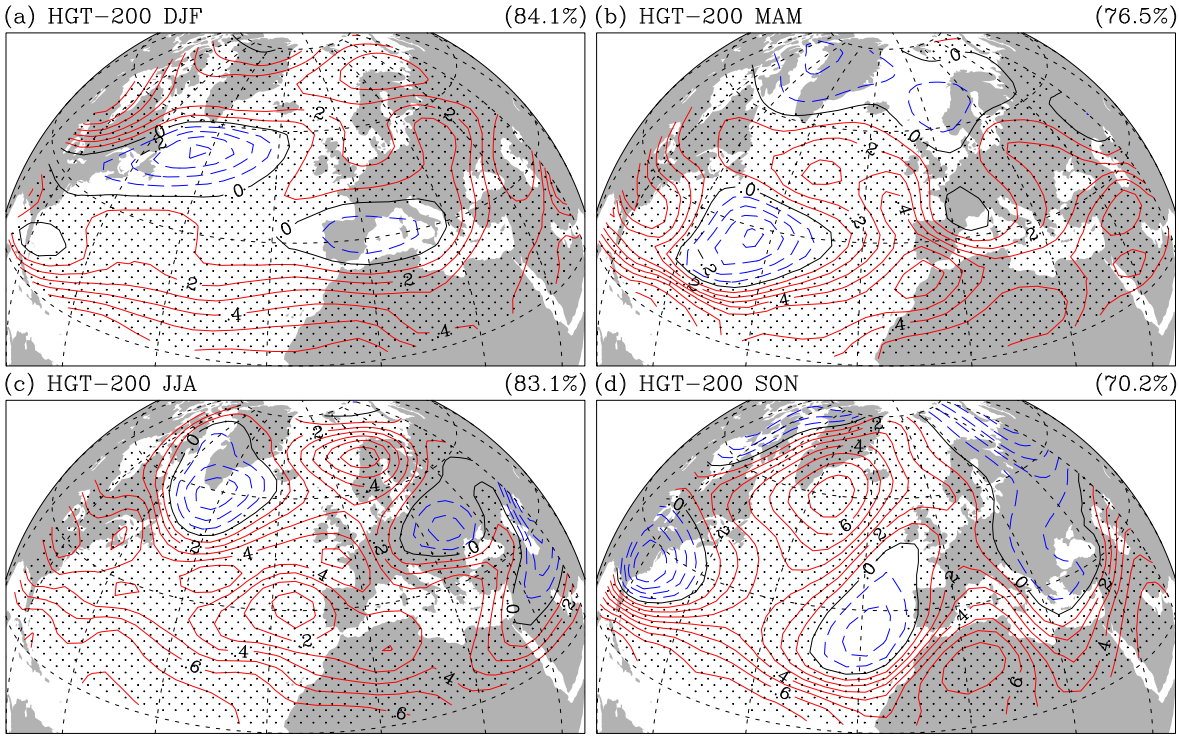

Figure 3.21: As in 3.16 but for HGT-200.

with significant ($\alpha = 0.01$) positive correlations are stippled. The percentage of the total EA area with significant positive correlations is printed on the right-hand top of each panel of the figures.

The percentages of figures 3.16 to 3.21 are high, meaning that the model simulates well the observed variability of all tested variables. TMP-SFC has the highest percentages due to the fact that surface temperatures in the ocean are observed sea level pressures. PRECIP has the lowest percentages. As already said in section 3.1.3, where the model climatology of PRECIP was tested, PRECIP is a difficult variable to simulate due to its highly irregular spatial and temporal behaviour. An interesting result is that winter is the better modelled season for all variables but TMP-SFC, for which it has, nevertheless, a percentage of 89.1%.

3.4.2 EOF patterns of MSLP

In this section we test the MUGCM skill to simulate atmospheric interannual variability over the EA region. To achieve this we applied an Empirical Orthogonal Function (EOF) analysis to both MUGCM and NCEP MSLP seasonal mean anomalies in the four standard seasons. In appendix A, the mathematical formulation of EOF analysis, or PCA, is presented. The EOF analysis of a time series of fields yield pairs of spatial patterns and associated time series. The pattern is usually referred to as EOF pattern and the associated time series is called Principal Component (PC) (see appendix A for the detailed nomenclature). Here, we focus only on the EOF patterns, which, when physical meaningful, are also named *teleconnections* [Wallace and Gutzler, 1981; Horel, 1981; Barnston and Livezey, 1987; Blackmon et al., 1984].

Remember that the MUGCM data set consists of 29 simulations of monthly means from January 1979 to February 1996. This results in 29 simulations of 17 DJF, MAM, JJA and SON seasonal means (see section 2.3, table 2.1). The MUGCM's grid over the EA region consists of 27 latitudes and 41 longitudes. There are, thus, $M = 27 \times 41 = 1107$ spatial locations and $N = 29 \times 17 = 493$ observations. The MUGCM data matrix, $\mathcal{X}$ (appendix A), for each MUGCM season, is built putting the (1107×17) matrix of each simulation, one after another, in the row direction and by increasing k (simulation number, see appendix A), yielding a (1107×493) data matrix.

The NCEP data set consists of monthly means from January 1959 to December 2001. This results in 51 DJF and 52 MAM, JJA and SON seasonal means. The NCEP grid over the EA region consists of 24 latitudes and 61 longitudes. Then we have $M = 25 \times 61 = 1525$ spatial locations and $N = 51$ observations for the DJF season and $N = 52$ observations for the MAM, JJA and SON seasons. Thus, the size of the NCEP data matrix, $\mathcal{X}$, is (1525×51) for DJF and (1525×52) for MAM, JJA and SON (see section 2.3, table 2.1).

The computation of MUGCM and NCEP EOFs is performed using the appropriate data matrix $\mathcal{X}$. These EOFs represent the total interannual variability in the modelled

and observed atmosphere over the EA region[7]. The first four EOFs of MUGCM and NCEP MSLP seasonal mean anomalies for the DJF, MAM, JJA and SON seasons are plotted in figures 3.22 to 3.25, respectively. Note that we removed the annual-mean cycle from the data prior to EOF computation, that is, the season climatology was removed from each seasonal mean value. The EOF loadings were not rescaled after computation and, thus, are dimensionless unit length vectors. The percentage of total variance represented by each EOF is indicated on the top right-hand corner of each panel. The sum of the percentages of total variance represented by the first four EOFs is 74.2(82.9), 72.5(72.8), 60.6(66.4) and 66.2(66.6)% for MUGCM (NCEP) in winter, spring, summer and autumn, respectively.

A quantitative measure of the similarity between the MUGCM and NCEP EOF patterns is supplied by table 3.4, where cross-correlations between modelled and observed EOFs are presented. We performed cross-correlations instead of simple correlations between equal ranked EOFs, because the same teleconnection pattern, that is, a large-scale mode of variability [Wallace and Gutzler, 1981], may be captured by one EOF of the NCEP data set and by a different EOF of the MUGCM data set. This is even more likely to occur when two modes of variability represent near values of the fraction of total variance. In this case the patterns may switch positions from one data set to the other. The main reason behind this behaviour is the difference in the number of observations of each data set and the different time periods to which they refer (see section 2.3, table 2.1).

Because one teleconnection pattern may be captured by different order MUGCM and NCEP EOFs, as explained in the previous paragraph, we used, in table 3.4, the bold face type to print the correlations computed from EOFs that represent the same teleconnection pattern. Normally these correlations are the highest in the line or column they belong, but not necessarily.

[7]It will be shown in sections 4.3 and 5.1 that the modelled EOFs obtained using the data matrix $\mathcal{X}$ built as explained above, are biased estimators of the Principal Vectors of total variability. Nevertheless, this bias is not important for the present purposes.

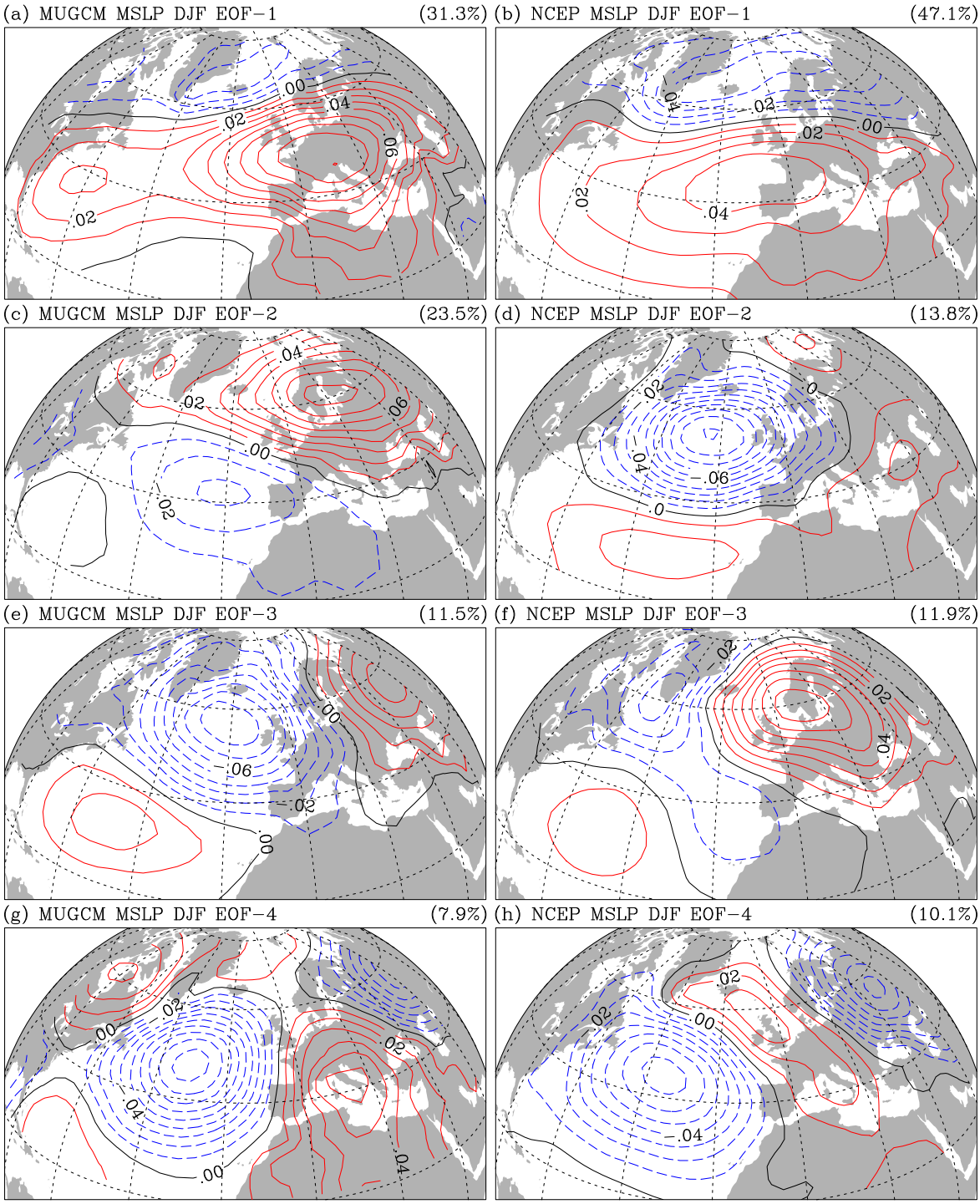

Figure 3.22: EOFs of MSLP DJF seasonal mean anomalies over the EA region. From MUGCM data: (a) EOF-1, (c) EOF-2, (e) EOF-3 and (g) EOF-4. From NCEP data: (b) EOF-1, (d) EOF-2, (f) EOF-3 and (h) EOF-4. Negative, zero and positive loadings plotted with dashed blue , solid black and solid red isopleths, respectively. EOFs are unit length and dimensionless. Grid lines represented every 20°.

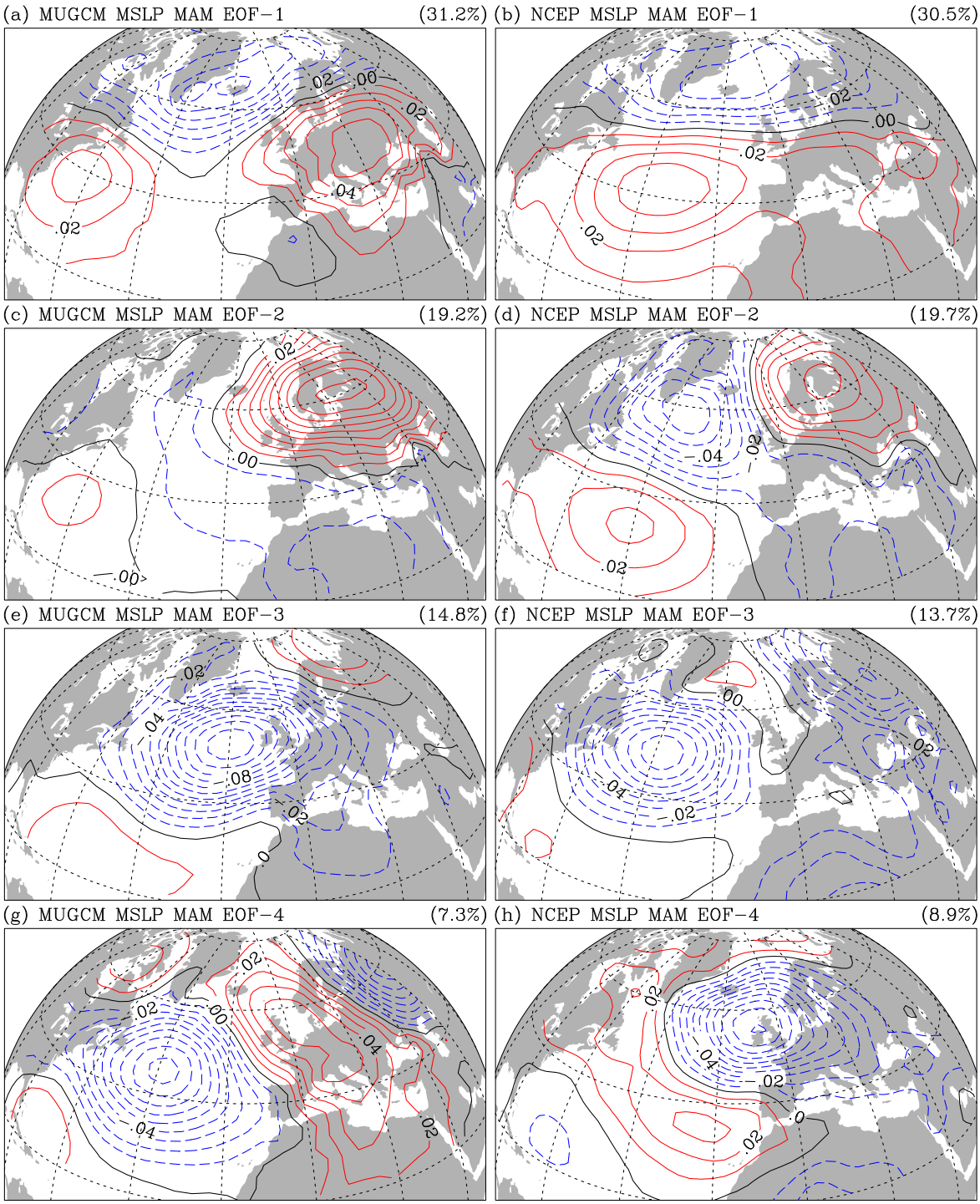

Figure 3.23: As in figure 3.22, but for MAM.

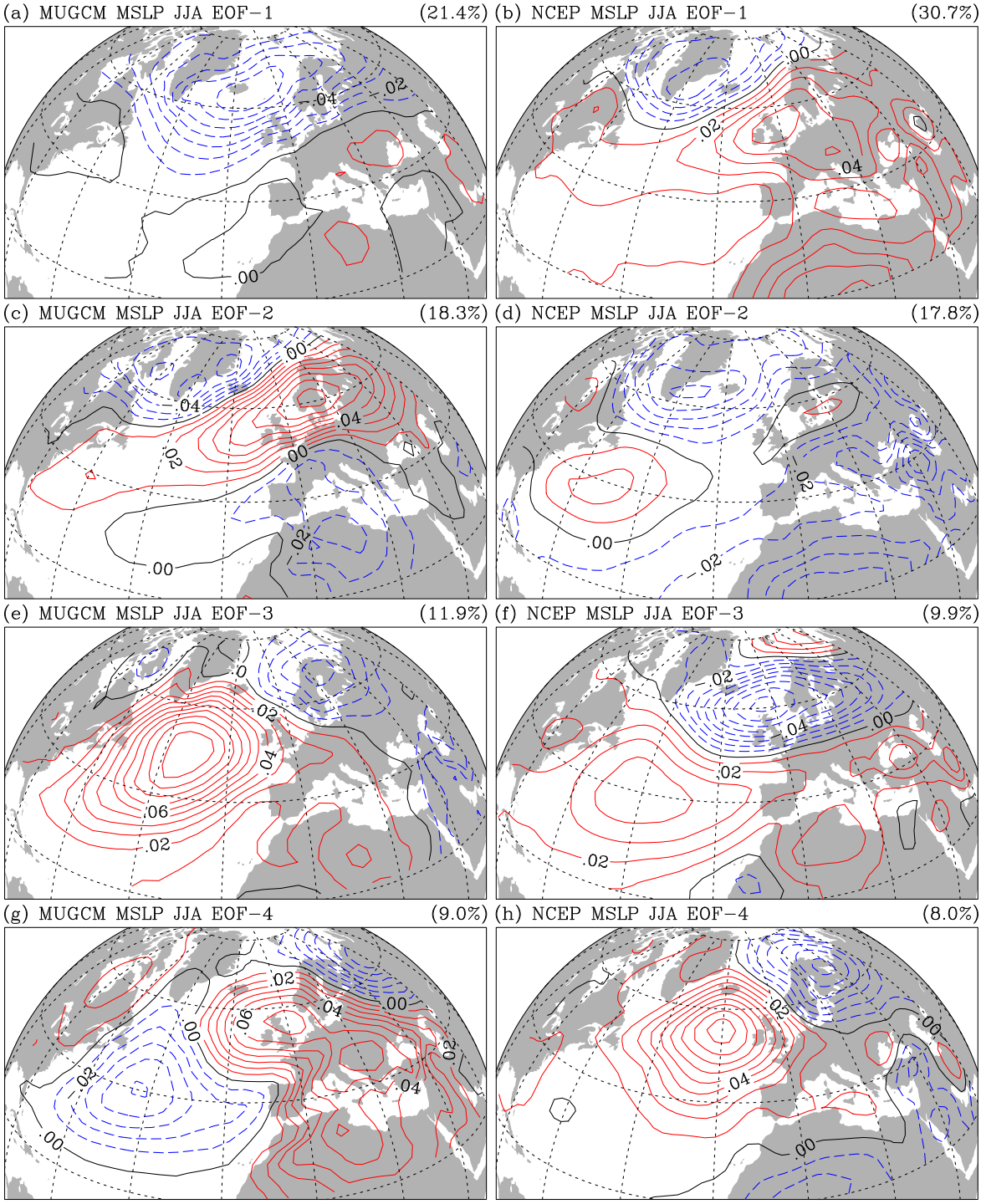

Figure 3.24: As in figure 3.22, but for JJA.

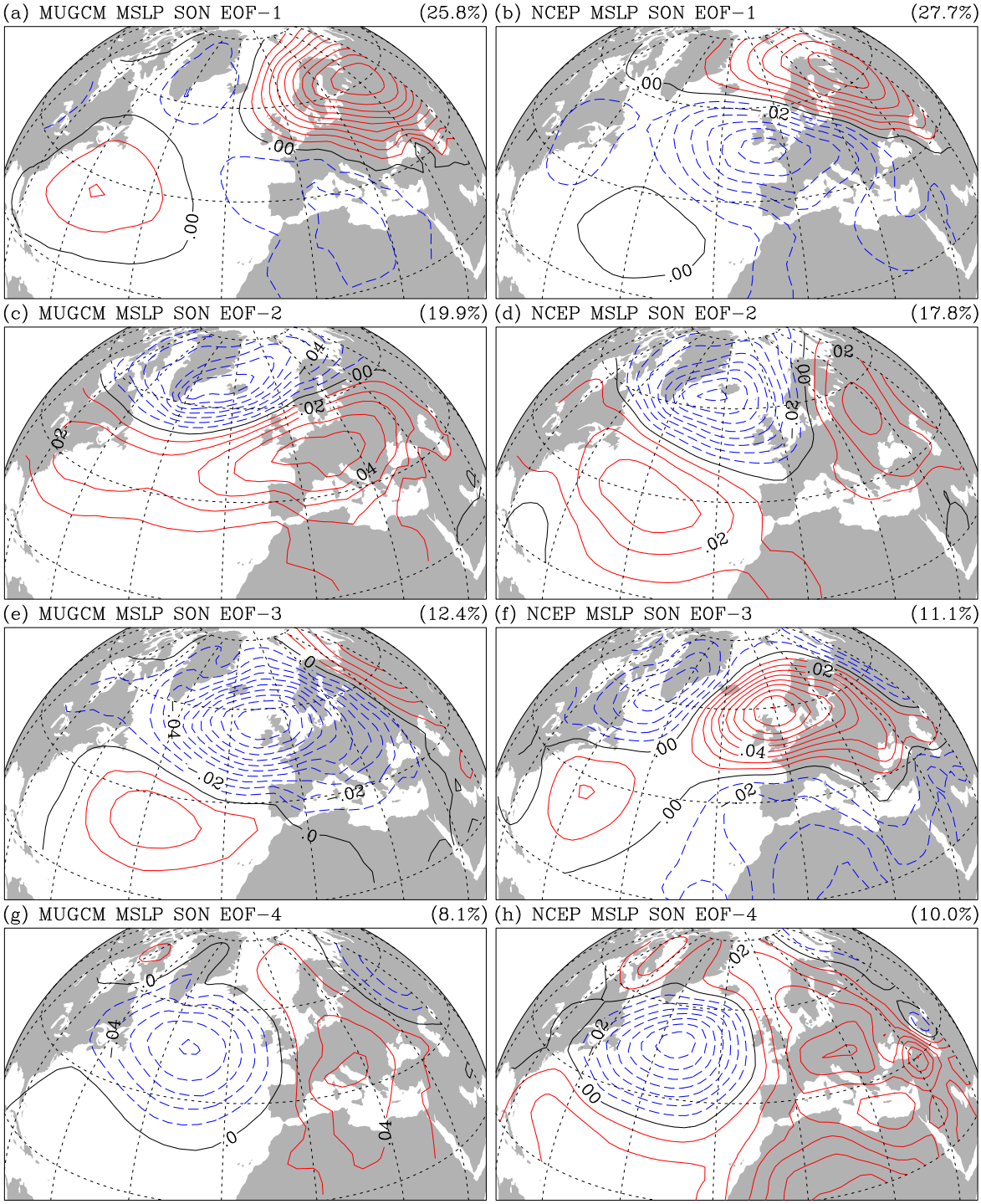

Figure 3.25: As in figure 3.22, but for SON.

Table 3.4: Spatial cross-correlations between MUGCM and NCEP MSLP seasonal EOFs over the EA region.

DJF EOFs	NCEP1 (47.1)	NCEP-2 (13.8)	NCEP-3 (11.9)	NCEP-4 (10.1)
MUGCM-1 (31.3)	**0.75**	-0.20	0.44	-0.02
MUGCM-2 (23.5)	-0.70	0.01	**0.61**	-0.05
MUGCM-3 (11.5)	0.08	**0.80**	0.28	-0.51
MUGCM-4 (7.9)	0.00	0.48	0.00	**0.73**
MAM EOFs	NCEP1 (30.5)	NCEP-2 (19.7)	NCEP-3 (13.7)	NCEP-4 (8.9)
MUGCM-1 (31.2)	**0.78**	0.32	0.00	0.23
MUGCM-2 (19.2)	-0.41	**0.71**	-0.24	0.38
MUGCM-3 (14.8)	-0.27	0.49	0.18	**0.61**
MUGCM-4 (7.3)	-0.24	-0.12	**0.72**	-0.49
JJA EOFs	NCEP1 (30.7)	NCEP-2 (17.8)	NCEP-3 (9.9)	NCEP-4 (8.0)
MUGCM-1 (21.4)	0.54	0.32	-0.60	0.08
MUGCM-2 (18.3)	0.67	0.50	-0.39	-0.09
MUGCM-3 (11.9)	0.08	0.34	0.30	0.78
MUGCM-4 (9.0)	0.42	-0.20	-0.33	0.28
SON EOFs	NCEP1 (27.7)	NCEP-2 (17.8)	NCEP-3 (11.1)	NCEP-4 (10.0)
MUGCM-1 (25.8)	0.73	0.29	0.47	-0.01
MUGCM-2 (19.9)	-0.53	0.74	0.29	0.08
MUGCM-3 (12.4)	0.45	**0.55**	-0.61	0.01
MUGCM-4 (8.1)	0.05	0.01	0.12	**0.90**

Winter is the season with the highest percentage of total variance captured by the first four EOFs: 74.2 and 82.9% for MUGCM and NCEP, respectively. This is mainly due to the fact that the first EOF (figure 3.22 (a,b)) manages to represent 31.3 and 47.1% of total MUGCM and NCEP variance, respectively. The spatial pattern responsible for this high variance is a north-south dipole which, for NCEP data, represents well the North Atlantic Oscillation in winter [Walker and Bliss, 1932; van Loon and Rogers, 1978; Wallace and Gutzler, 1981; Barnston and Livezey, 1987]: a meridional see-saw feature in the atmospheric mass between the Icelandic Low near 65°N and a broad east-west belt, centred near 40°N, extending from the east coast of the United States to the Mediterranean. The modelled pattern has also these characteristics but the highest loadings of the southern centre of action are shifted to the east by about 15°. The spatial correlation with the observed pattern is 0.75.

In the second NCEP EOF pattern (figure 3.22 (d)), representing 13.8% of NCEP total variance, we recognise the East Atlantic pattern which is characterised by a major centre of action, centred west of Great Britain and spanning the entire North Atlantic ocean from east to west, with a weaker centre of opposite sign south of it. This is the pattern reported by Barnston and Livezey [1987]. Its structure resembles the NAO shifted southward, however, its southern centre of action is dynamically linked to the tropical Atlantic. The *eastern Atlantic* pattern of Wallace and Gutzler [1981] has a third centre of action near the Black Sea, which also slightly appears in the NCEP pattern. The modelled East Atlantic pattern is captured by EOF-3 (figure 3.22 (e)), which accounts for 11.5% of total model variance and has a correlation of 0.80 with NCEP EOF-2. The northern and southern centres of action are well simulated, but another centre of action, in phase with the southern one, appears to the east of the northern one, making this pattern also similar to NCEP EOF-4, with which it has a correlation of -0.51.

NCEP EOF-3 (figure 3.22 (f)) explains 11.9% of total variance and shows the Scandinavia pattern, the EURASIA I pattern of Barnston and Livezey [1987]. The Scandinavia pattern is usually described as a primary centre over Scandinavia and two centres of

opposite sign over the eastern Atlantic extending to western Europe, and over eastern Russia (this one falling out of our domain). The modelled pattern appears in EOF-2, representing 23.5% of the MUGCM total variance. Its correlation with NCEP EOF-2 is only 0.61 because the "eastern Atlantic - western Europe" centre in the NCEP EOF is actually a "western Atlantic" centre extended to western Europe. Despite this low correlation, the primary circulation centre over Scandinavia is well simulated.

Finally, the East Atlantic/Western Russia pattern, the EURASIA II of Barnston and Livezey [1987], is identified in both NCEP and MUGCM EOF-4 (figure 3.22 (g,h)). These EOFs, representing 7.9% and 10.1% of MUGCM and NCEP total variance, respectively, show three of the four centres of the pattern which has a wave-train structure with anomaly centres over the Atlantic, western Europe, north eastern Siberia and northern China (out of the EA domain). The modelled pattern is very similar to the observed one (with a spatial correlation of 0.73) but the simulated european centre, which should cross Europe from the North Sea to the Mediterranean, is displaced southeastward.

In spring, the NAO pattern is still found in the first EOF of both MUGCM and NCEP data. Although the correlation between the two patterns increases from 0.75 in winter to 0.78 in spring, the southern belt of positive anomalies (in the NAO's positive phase) is now broken in two centres of equal sign located in western Atlantic and Europe. The Scandinavia pattern is now captured by the same EOF in both modelled and observed data, the second EOF, with a correlation of 0.71. The East Atlantic pattern is still found in EOF-3 of modelled data but changed to EOF-4 in observed data. The correlation between the two is 0.61. The East Atlantic/Western Russia pattern continues to be found in MUGCM EOF-4 but is now represented by NCEP EOF-3, with correlation of 0.72.

In summer, despite some high correlations, none of the four discussed teleconnection patterns are found.

In autumn, the East Atlantic pattern is recognised in MUGCM EOF-3 and NCEP EOF-2, with a correlation of 055, and the East Atlantic/Western Russia pattern is found in EOF-4 of both MUGCM and NCEP, with a correlation of 0.90. Apart from this, MUGCM EOF-2 resembles the NAO pattern and NCEP EOF-3 resembles the Scandinavia pattern.

3.5 Summary and conclusions

The validation of MUGCM's climatology and variability was presented in this chapter using the reanalysis from NCEP. The model climatology was assessed using: MSLP, TMP-SFC, PRECIP, HGT-850, HGT-500 and HGT-200 seasonal means in winter, spring, summer and autumn; vertical structures of zonal-mean zonal circulation and temperature in winter and summer; upper tropospheric zonal wind, in particular the jet streams, in winter and summer; planetary standing waves and eddy components (departures from zonal symmetry) in winter; and meridional and zonal mass overturning circulations in the Pacific and Atlantic (Hadley, Ferrel and Walker circulation) in winter.

MUGCM and NCEP seasonal climatologies for MSLP, TMP-SFC, PRECIP, HGT-850, HGT-500 and HGT-200 are significantly different, at a significance level of 0.001, at each grid point of the EA region. However, the patterns of these climatologies are in a fairly agreement, when compared visually and by spatial correlation. Overall, the model is able to reproduce the main features of the observed seasonal climatology, but some important biases were detected. The pressure cells over the Atlantic ocean are overestimated in winter, resulting in the overestimation of the meridional MSLP gradient between the Azores High and the Icelandic Low. Accordingly, the low-tropospheric circulation in the EA sector was also found to be overestimated by the model, as well as surface temperature and precipitation.

The modelled Pacific jet stream is biased 10° (30°) downstream in winter (summer), is displaced 10° to the north in both seasons, and is 15 m s^{-1} stronger in winter. The

Atlantic jet stream is biased 20° downstream (10° upstream) in winter (summer), is shifted 10° (5°) to the north in winter (summer), and is 10 m s^{-1} stronger in summer.

The modelled meridional mass circulation is overestimated. In particular, the thermally direct Hadley cells are too intense, which may have important consequences on the extratropical atmospheric response to tropical SST forcing. The Pacific jet stream overestimation may be a consequence of the intensified Pacific Hadley cells. It was also shown that the Walker cell in the equatorial Pacific and Atlantic are also overestimated by the model. The overestimation of the tropical atmospheric response to SST climatology indicates that its response to SST anomaly variability may also be overestimated. In conjuction with the overestimation of the mean state of the Pacific jet stream, this may lead to an overestimation of the extratropical atmospheric response to tropical Pacific SST anomalies, in particular, a stronger ENSO forcing in the NPA and EA regions.

Modelled and observed variability was tested using two different methodologies. Firstly, the correlation at each grid point of the MUGCM's grid in the EA sector, between the MUGCM-AMIP II ensemble mean and the NCEP MSLP, TMP-SFC, PRECIP, HGT-850, HGT-500 and HGT-200, was determined for each of the four standard seasons. Secondly, the spatial patterns of variability of MUGCM and NCEP seasonal mean anomalies of MSLP, in the EA region, obtained by PCA, were compared.

The correlations between modelled and observed grid point time series are significantly greater than zero, at a significance level of 0.01, in almost the entire area of the EA region.

PCA of seasonal mean MSLP in the EA region captured the major observed modes of MSLP anomaly variability in the EA region, namely, the North Atlantic Oscillation, the East Atlantic mode, the Scandinavia or EURASIA I mode, and the East Atlantic/Western Russia or EURASIA II mode. These modes are better captured by the model in winter and spring.

The NAO modelled pattern has the characteristic meridional see-saw in the atmospheric mass between the Icelandic Low and a broad east-west belt, centred near 40°N, extending from the east coast of the United States to the Mediterranean. However, the southern belt of NAO pattern has its highest loadings shifted to the east by about 15° in winter, and is broken in two centres of equal sign located in western Atlantic and Europe in spring.

Chapter 4

Internal and Forced Variability

The AMIP II experiment performed with the MUGCM allows the separation of total atmospheric variability of into a component due to atmospheric internal dynamics only, called *internal* or *natural variability* , and a component due to the variability of the lower-boundary conditions (SST and SIC) referred as *external* or *forced variability*.

In this chapter we describe the model of the ANOVA that better suits the layout of AMIP seasonal mean data for the purpose of this research: the *random-effects* model [Scheffé, 1959]. Statistics to estimate the partition of total seasonal mean variance into a boundary forced component and an internal component are presented. We also present an hypothesis test to determine if the variability due to boundary forcing is statistically greater than zero. Two test statistics are supplied, one of which is the popular *potential predictability*.

The MUGCM seasonal mean internal and forced variabilities are inferred applying the ANOVA to mean MSLP seasonal means in winter (DJF), spring (MAM), summer (JJA) and autumn (SON), in the EA sector. Winter is found to be the season with higher internal variability whereas spring is the season with higher forced variability, followed by winter, in the EA region. Accordingly, the percentage of total EA area with significant potential predictability is found to be higher in spring followed by winter.

The time series associated with the winter and spring modes of total variability of MSLP anomalies the EA region, obtained in section 3.4.2, are also subjected to the ANOVA. The North Atlantic Oscillation is found to more externally forced in spring than in winter.

Since the NAO is the major mode of atmospheric variability in the EA region, the ANOVA is also applied to the NAO index in twelve consecutive and overlapping 3-month seasons. Significant forcing is found in all seasons, except for autumn and October to November, the stronger forcing found again in spring.

4.1 Introduction

Lorenz [1963, 1965] demonstrated that, due to the inherent nature of instability and nonlinearity, atmospheric flows with only slightly different initial states will depart from each other and evolve eventually to flows that are just randomly related. Because of this, daily weather variations, which are due primarily to the internal dynamics of the atmosphere, cannot be predicted in detail more than 2-3 weeks in advance [Lorenz, 1982; Chen, 1989]. These short timescale fluctuations induce unpredictable variability in interannual variability of seasonal mean quantities [Leith, 1973], usually referred as *climate noise*, *natural* or *internal variability*. However, slowly varying *external* boundary conditions, such as anomalies of SST and SIC can cause predictable variations in seasonal mean quantities, referred as *external* or *forced variability*. It is clearly important to be able to asses where on the globe atmospheric variations are sufficiently affected by oceanic forcing to enable practical seasonal prediction. This requires the estimation, at each grid point of a global grid, of atmospheric *potential predictability* (*potential* indicates that this also depends in predictions of anomalous oceanic forcing, such as SST anomalies), usually defined as the signal-to-noise ratio (forced to internal variance ratio) or the forced to total (internal plus forced) variance ratio.

The estimation of potential predictability has been a topic of ongoing research in the climate community. One approach is to use the observational record, as proposed by

Madden [1976] (see also Shukla [1983]). In this method, internal variability is defined as the variability resulting from the variance and autocorrelation associated with daily weather fluctuations, and total variability is determined as the actual interannual variance of monthly or seasonal means. This method has the advantage that can be based purely on observational data. It also includes secondary sources of predictability, not just that due to the ocean surface. However, the use of the observational record alone may also be considered a disadvantage since it represents only one realisation of the atmospheric evolution.

Atmospheric General Circulation Models (AGCMs) have also been used to estimate potential predictability. Different experimental approaches with AGCMs have been conducted. One approach [Chervin and Schneider, 1976] consists of performing an ensemble of AGCM integrations, all forced by the SST climatology of one calendar month. Usually the January climatology is chosen because it is when internal variability is higher in the Northern Hemisphere. The ensemble variance then provides an estimate of the internal variance. The total interannual variability is estimated performing an integration of the same AGCM forced by realistic interannual variations of SST.

Another approach [Chervin, 1986] consists of performing an ensemble of AGCM integrations, all forced by the climatological annual cycle of SST (ACYC simulations, see chapter 2). The ensemble variance for each month provides an estimate of the internal variance for that calendar month. The average of the twelve calendar month variances then provides an estimate of internal variability. Again, the total interannual variability is estimated performing an integration of the same AGCM forced by realistic interannual variations of SST.

More recently, potential predictability has been measured using an ensemble of climate simulations all forced by the same observed interannually varying SSTs but started from different initial conditions, such as the AMIP and AMIP II experiments. Since these simulations provide temporally and spatially complete climate data, they are also valuable in other areas of climate research. For predictability studies, the philosophy is

that sensitivity to initial atmospheric conditions can be used to quantify the random component of interannual variability, whereas the relative similarity (or lack of it) between ensemble members can be used to quantify the potential predictable component of variance. The output of such ensemble has been analysed differently to asses potential predictability. Dix and Hunt [1995] performed an AMIP experiment (1979-1988) with three runs. The authors did not estimated potential predictability using estimates of internal, forced or total variabilities. Instead, they used monthly correlations between the individual simulations. The grid points with higher correlation are then considered to have higher potential predictability.

Stern and Miyakoda [1995] used a 9-member AMIP ensemble to estimate winter (DJF) and summer (JJA) potential predictability, which they preferred to term *reproducibility*. The authors defined the reproducibility index as the variability among the ensemble (internal variability) normalised by the climatological seasonal variability (total seasonal variability).

Kumar et al. [1996] used potential predictability to compare the suitability of two different AGCMs to make seasonal predictions. They performed an AMIP-like experiment (1982-1983) with 9 simulations with the two models, and determined maps of forced variability, which they defined as the variance of the ensemble mean around the climatological mean. After confirming that the internal variability was similar in the two models, they then used the maps of forced variability as a measure of potential predictability.

Nowadays, ANOVA is considered the best method to estimate potential predictability from an AMIP-like ensemble. This method was used by Harzallah and Sadourny [1995] (with a slightly variation of the method), Rowell et al. [1995]; Zwiers [1996]; Davis et al. [1997]; Rowell [1998] and Wang and Zwiers [1999].

4.2 The ANOVA model for AMIP seasonal mean data

In an AMIP experiment, an AGCM is initialised with a state of the atmosphere, randomly chosen from a previous control run, and forced by a time series of lower-boundary states of SST and SIC fields. This procedure is repeated several times, using, for each simulation, another randomly chosen initial state of the atmosphere and the same boundary states of SST and SIC fields. In other words, for each boundary forcing, we get an ensemble of responses of the atmosphere to that forcing. For seasonal mean data, we get an ensemble of K simulations of the response of the atmosphere to the boundary forcing in year $p = 1, ..., P$. Let x_{pk} be the seasonal mean variable resulting from the kth simulation in year p of our experiment. Since the simulations are initialised randomly, x_{pk} is actually one realisation of the random variable X_{pk}.

Now, since de atmosphere forgets its initial state very quickly [Lorenz, 1963], the effect of selecting different initial conditions is basically to select independent realisations of the simulated climate's path through its phase space. Thus, the K simulations are independent. Furthermore, since the seasonal means can be considered independent from one year to the next, the P ensembles are also independent. Therefore, the random variables X_{pk}, where $p = 1, ..., P$ and $k = 1, ..., K$, are $P + PK$ mutually independent random variables.

In summary, we can treat the AMIP seasonal mean data as independent random samples of length K of P populations that differ one from another by the boundary condition that forced it:

$$X_{p1}, X_{p2}, ..., X_{pK}, \quad p = 1, ..., P.$$

This type of data is suitable to be analysed by the *one-way* ANOVA model:

$$X_{pk} = \mu + A_p + E_{pk}$$

where (i) μ is the *overall mean response* which is the mean of the responses to all P boundary forcings, (ii) A_p is the *treatment effect* which is the deviation of the response to the boundary forcing in year p from the overall mean response, μ, and (iii) E_{pk} is the *error effect* which is a random deviation from $\mu + A_p$, is independent of the boundary forcing and is due only to internal dynamics. In the atmospheric sciences, the terms *treatment* and *error* are usually replaced by *forced* and *internal*, respectively.

The issue to discuss now is if we should consider the forced effects A_p, as constants or as a random variables, since this will determine if the *one-way* ANOVA model is a *fixed-effects* model or a *random-effects* model, respectively.

Indeed, the two options fits our data, depending on the point of view. By one side, the SST-SIC boundary in an particular year was not chosen at random, since it is the boundary observed in that year. Thus, since the SST-SIC boundaries are fixed, the responses of the atmosphere, A_p, $p = 1, ..., P$, are constants. On the other side, since the observed SSTs and SICs are realisations of random variables (they could have been different), we may consider that the choice of the SST-SIC boundaries was (indirectly) random. By this reasoning we may consider A_p, $p = 1, ..., P$, as random variables.

We conclude then that both ANOVA models fit our data and, thus, another criterium must be used in order to choose between one of them. This criterium relates to the purpose of the experiment. Murteira [1990] wrote "se os níveis do factor constituem um conjunto por assim dizer exaustivo, isto é, se o investigador tem intenção de confinar a análise a esse conjunto e não pretende raciocinar em relação a níveis que poderiam ter sido explicitados mas que não o foram, o modelo I é em geral mais adequado. Se os nveis constituem uma amostra de um conjunto mais vasto de níveis em relação ao qual o investigador tem o propósito de alargar as conclusões, então o modelo II é em regra mais recomendável"[1].

[1]Here, "níveis do factor" are the treatments and models I and II are the fixed-effects and random-effects models, respectively.

In our AMIP experiment, the P populations from which the P samples were drawn are part of a larger (infinite) set of populations and our purpose is to use the ANOVA to make inferences about this general set of populations, not only about the P populations. That is, we are not only interested in the atmospheric response to the 1979-1996 SST-SIC boundaries, but in the atmospheric response to the oceanic forcing in general. By this idea, we conclude that the appropriate model to use in this research is the *random-effects* model:

$$X_{pk} = \mu + A_p + E_{pk} \tag{4.1a}$$

$$A_p \sim \mathcal{N}(0, \sigma_F^2) \tag{4.1b}$$

$$E_{pk} \sim \mathcal{N}(0, \sigma_I^2) \tag{4.1c}$$

where μ is a constant, A_p and E_{pk} are $P + PK$ mutually independent random variables, $p = 1, ..., P$, $k = 1, ..., K$, $\mathcal{N}$ represents the Gaussian probability distribution, and σ_F^2 and σ_I^2 are the forced and internal variances, respectively.

Although we dedicated some attention in choosing between the *fixed-effects* and the *random-effects* model, we must stress out that main results of the analysis are insensitive to this choice. The test of the SST-SIC effects in the *fixed-effects* is done testing the null hypothesis that the constants A_p are all zero. In the *random-effects* model we test the null hypothesis that the random variables A_p all have zero variance. Although different, these tests are equivalent and yield the same results. The reason for this is that, in the *random-effects*, the random variables A_p have zero mean and thus the null hypothesis of the two tests are equivalent. Consequently, the same test statistic is used in both tests.

The only difference between the results of the two models is that, in the *random-effects* model, the variability of the forced effects is quantified by its variance, the *forced* variance σ_F^2, whereas in the *fixed-effects* the variance statistic cannot be used because A_p are constants, not random variables. Castanheira [2000], who used the *fixed-effects* model to analyse its AMIP simulations, overcame this problem by defining

forced variance as

$$\sigma_F^2 = \frac{1}{P-1}\sum_{p=1}^{P}(A_p - \mu)^2,$$

whose unbiased estimator is the same of that of the *random-effects* model.

Finally, we note that other authors [Rowell et al., 1995; Davis et al., 1997; Rowell, 1998] also considered the *random-effects* model the best model of the *One-way* ANOVA to analyse their AMIP-like data.

4.3 Partition of variability into forced and internal components

In this section we present the guidelines to obtain the partition of the variability in *random-effects* model of the *one-way* ANOVA. The complete derivation of the formulae can be be found in any statistics book that covers the subject such as Murteira [1990], Milton and Arnold [1995] and, of course, the more complete Scheffé [1959]. It can also be found in von Storch and Zwiers [1999] which is a statistics book oriented to climate science.

The *total sum of squares*,

$$SS_T = \sum_{p=1}^{P}\sum_{k=1}^{K}(X_{pk} - \overline{X}_{\bullet\bullet})^2, \tag{4.2}$$

where the dot notation indicates averaging over the missing subscript, can be decomposed into a *treatment* or *forced sum of squares*, SS_F, and a *error* or *internal sum of squares*, SS_I,

$$SS_T = SS_F + SS_I,$$

where

$$SS_F = K\sum_{p=1}^{P}(\overline{X}_{p\bullet} - \overline{X}_{\bullet\bullet})^2 \tag{4.3}$$

and

$$SS_I = \sum_{p=1}^{P}\sum_{k=1}^{K}(X_{pk} - \overline{X}_{p\bullet})^2. \tag{4.4}$$

The *forced sum of squares*, SS_F, is taken over P deviations that sum to zero, $\sum_{p=1}^{P}(\overline{X}_{p\bullet}-\overline{X}_{\bullet\bullet})=0$, thus it has $P-1$ degrees of freedom. The *internal sum of squares*, SS_I, is taken over PK deviations such that deviations within a particular ensemble must sum to zero, $\sum_{k=1}^{K}(X_{pk}-\overline{X}_{p\bullet})=0$. That is, SS_I is taken over deviations that are subject to P constraints. Consequently, SS_I has $P(K-1)$ degrees of freedom. The SS_T is summed over PK deviations that sum to zero, $SS_T=\sum_{p=1}^{P}\sum_{k=1}^{K}(X_{pk}-\overline{X}_{\bullet\bullet})=0$, and therefore it has $PK-1$ degrees of freedom. Note that the degrees of freedom SS_T is equal to the sum of the degrees of freedom of SS_F and SS_I, as should be.

By dividing a sum of squares, SS, by the corresponding degrees of freedom, we obtain a statistic called the *mean square*, MS. Thus, the *treatment* or *forced mean square* is given by

$$MS_F=\frac{K}{P-1}\sum_{p=1}^{P}(\overline{X}_{p\bullet}-\overline{X}_{\bullet\bullet})^2 \tag{4.5}$$

and the *error* or *internal mean square* is obtained by

$$MS_I=\frac{1}{P(K-1)}\sum_{p=1}^{P}\sum_{k=1}^{K}(X_{pk}-\overline{X}_{p\bullet})^2. \tag{4.6}$$

Substituting model 4.1 in equations 4.5 and 4.6, and taking the expectations of MS_F and MS_I, noting that $\mathcal{E}\big[1/(P-1)\sum_{p=1}^{P}(A_p-\mu)^2\big]=\sigma_F^2$, $\mathcal{E}\big[1/(P-1)\sum_{p=1}^{P}(\overline{E}_{p\bullet}-\overline{E}_{\bullet\bullet})^2\big]=\sigma_I^2/K$, and $\mathcal{E}\big[1/(K-1)\sum_{k=1}^{K}(E_{pk}-\overline{E}_{p\bullet})^2\big]=\sigma_I^2$, we obtain

$$\mathcal{E}\big[MS_F\big]=K\sigma_F^2+\sigma_I^2 \tag{4.7}$$

and

$$\mathcal{E}\big[MS_I\big]=\sigma_I^2. \tag{4.8}$$

From equations 4.7 and 4.8 we obtain unbiased estimators of the forced variance

$$\hat{\sigma}_F^2=\frac{MS_F-MS_I}{K}, \tag{4.9}$$

and internal variance

$$\hat{\sigma}_I^2=MS_I. \tag{4.10}$$

It is also common in research works with ensembles to use the so called *ensemble-mean variance*. The ensemble-means are also a random variables since are the average (over the ensemble) of random variables:

$$\overline{X}_{p\bullet} = \frac{1}{K}\sum_{k=1}^{K} X_{pk}.$$

Substituting model 4.1 into this equation, we have

$$\overline{X}_{p\bullet} = \mu + A_p + \overline{E}_{p\bullet},$$

where $A_p \sim \mathcal{N}(0, \sigma_F^2)$ and $\overline{E}_{pk} \sim \mathcal{N}(0, \sigma_I^2/K)$ are independent random variables. Thus, the variance of the random variables $\overline{X}_{p\bullet}$, the ensemble-mean variance, is

$$\sigma_{EM}^2 = \sigma_F^2 + \frac{\sigma_I^2}{K}, \tag{4.11}$$

whose unbiased estimator is

$$\hat{\sigma}_{EM}^2 = \hat{\sigma}_F^2 + \frac{\hat{\sigma}_I^2}{K}, \tag{4.12}$$

which, substituting equation 4.7 into equation 4.11, or using equations 4.9 and 4.10, becomes

$$\hat{\sigma}_{EM}^2 = \frac{MS_F}{K}. \tag{4.13}$$

Equation 4.11 shows that if we had an infinite ensemble, the forced variability would be obtained by the ensemble-mean variability. It also shows how the ensemble mean is contaminated by internal variability, demonstrating the importance of the size of the ensemble when the ensemble mean is used to represent the forced response of the atmosphere.

Finally, the total variability

$$\sigma_T^2 = \sigma_F^2 + \sigma_I^2, \tag{4.14}$$

can be estimated without bias by

$$\hat{\sigma}_T^2 = \hat{\sigma}_F^2 + \hat{\sigma}_I^2, \tag{4.15}$$

Considering X_{pk} of model 4.1 the seasonal mean MSLP (anomalies or not) at each grid point of the EA sector, with $P = 17$ years and $K = 29$ simulations, we computed equations 4.5, 4.6, 4.9 and 4.10 to estimate the forced and internal variabilities. The results are plotted in figure 4.1.

Note that, to use anomalies (season climatology removed) or the original data, has no effect on the results of the ANOVA. The reason for this is found in equation 4.2 where the operation $X_{pk} - \overline{X}_{\bullet\bullet}$ is precisely the removal of the season climatology from the data. This term has the same value whether the season climatology has been previously removed or not. Thus, since all the ANOVA formulae originates from this equation, the ANOVA results are insensitive to the climatology of the data.

Alternatively, we may simply notice that the climatology, $\overline{X}_{\bullet\bullet}$, has only impact on the overall mean response, $\mu = \mathcal{E}[\overline{X}_{\bullet\bullet}]$ of model 4.1, and therefore it has no effect on the variance components.

As expected, internal variability is higher in the extratropics because of the great eddy activity in this region. In the tropics, internal variability is negligible in all seasons. Winter is clearly the season with highest internal variability, whereas the lowest values of internal variability are found in summer.

In average, winter and spring are the seasons with higher forced variability. However, at high latitudes, spring presents higher forced variability than in winter, whereas in the longitudinal belt at midlatitudes extending from Florida to central Europe, forced variability is higher in winter than in spring.

Figure 4.1 shows that, in general, the regions of maximum forced variability also have high internal variability. We thus need to test, at every grid point, if the forced variability make a statistically significant contribution to the seasonal mean MSLP variability, in order to localise the regions of significant SST-SIC forcing.

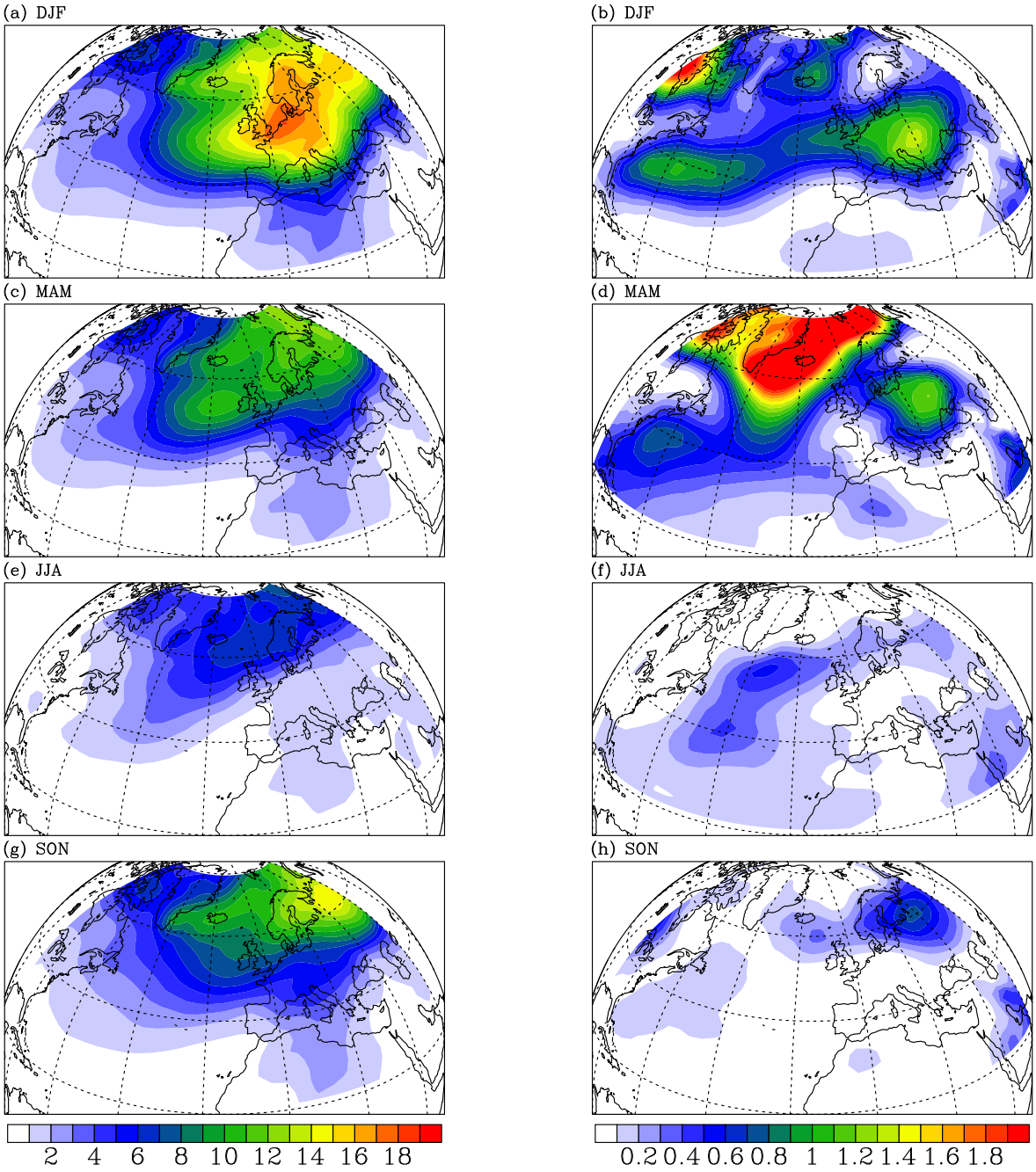

Figure 4.1: Internal (left) and forced (right) variance of MSLP seasonal mean anomalies, obtained by a Random-Effects One-Way ANOVA applied to the MUGCM ensemble, for (a,b) DJF, (c,d) MAM, (e,f) JJA, and (g,h) SON. Units in mb^2. Grid lines represented every 20^o.

4.4 Testing the SST-SIC forcing. Potential predictability

Once the total variability of the seasonal mean data has been partitioned into a SST-SIC forced and an internal component, the next step is to test if the total variability has a statistically significant contribution from the SST-SIC variability. That is, we want to perform the following test

$$H_0 : \sigma_F^2 = 0 \quad vs \quad H_1 : \sigma_F^2 > 0. \tag{4.16}$$

From equations 4.7 and 4.8 we see that, when H_0 is true, we expect that MS_I and MS_F to be close in value since both of them estimate σ_I^2. When H_0 is not true, we expect MS_F to be larger than MS_I. This suggests the ratio MS_F/MS_I as a logical test statistic. If H_0 is true, its value is expected to be close to 1, otherwise it is expected to be larger than 1. Under H_0 we have that $X_{pk} \sim \mathcal{N}(\mu, \sigma_I^2)$ and it can be easily shown [Murteira, 1990; Milton and Arnold, 1995; Scheffé, 1959] that $SS_F/\sigma_I^2 \sim \chi^2_{P-1}$ and $SS_I/\sigma_I^2 \sim \chi^2_{P(K-1)}$. where χ^2 represents the Qui-squared distribution. Since we also have that SS_F and SS_I are independent, then

$$F = \frac{MS_F}{MS_I} = \frac{\frac{SS_F/\sigma_I^2}{P-1}}{\frac{SS_I/\sigma_I^2}{P(K-1)}} \sim \mathcal{F}_{P-1,P(K-1)} \tag{4.17}$$

if H_0 is true, where $\mathcal{F}$ is the Fisher distribution. Also note that, using equations 4.10 and 4.13, the test statistic 4.17 can be expressed as

$$F = \frac{K\hat{\sigma}_{EM}^2}{\hat{\sigma}_I^2} \sim \mathcal{F}_{P-1,P(K-1)} \tag{4.18}$$

under H_0. Finally, we have that the criterion for rejecting H_0 at the α significance level is

$$F > \mathcal{F}_{P-1,P(K-1),\alpha} \tag{4.19}$$

An alternative test statistic to F which is very popular in the atmospheric sciences is the so called *potential predictability*, defined as the fraction of the variability due to SST forcing:

$$R = \frac{\hat{\sigma}_F^2}{\hat{\sigma}_T^2}. \tag{4.20}$$

After equation 4.18 and using the equations 4.12 and 4.15, we have that

$$F = 1 + K(R^{-1} - 1)^{-1}. \tag{4.21}$$

Then, using the test rule 4.19 and equation 4.21, the criterion for rejecting H_0 at the α significance level can ve expressed as

$$R > \left(1 + K\left(\mathcal{F}_{P-1,P(K-1),\alpha} - 1\right)^{-1}\right)^{-1}. \tag{4.22}$$

The fraction of total seasonal MUGCM MSLP variance due to SST-SIC forcing, R, that is, the potential predictability, is calculated using equation 4.20. The results for each season are presented in figure 4.2. To test if the SST-SIC forced variance is significantly greater that zero at the 0.05 significance level, we performed the test 4.16 using the test rule 4.22 with $K = 29$, $P = 17$ and $\alpha = 0.05$. Rejection of the null hypothesis, which occurs for $R > 0.0225$, implies significant SST-SIC forcing. At the grid points where the null hypothesis was not rejected, the value of R was not plotted. Thus, white regions in figure 4.2 are not significantly forced by the lower boundary ($R \leq 0.0225$).

Overall, potential predictability in the EA region is low (blue areas) but significant (not white areas). The percentage of the total area of the EA that has significant potential predictability is 90.4%, 91.1%, 85.5% and 63.7% in winter, spring, summer and autumn, respectively. These percentages are shown on the top right-hand corner of the panels of figure 4.2.

All seasons share high values of R in the tropical latitudes except in Africa. Note also the high values of R in the Arabian Peninsula, where it exceeds 40% in winter and in spring, probably because of the direct influence of the adjacent Indian Ocean.

In winter, the highest values of R are found in central and western part of the 20°N - 40°N domain, and and eastern Canada. Europe has low, but significant, values. The southern Iberian Peninsula, northwestern Africa and Scandinavia have no significant forced variability. In spring, the highest values of R are found across the Atlantic

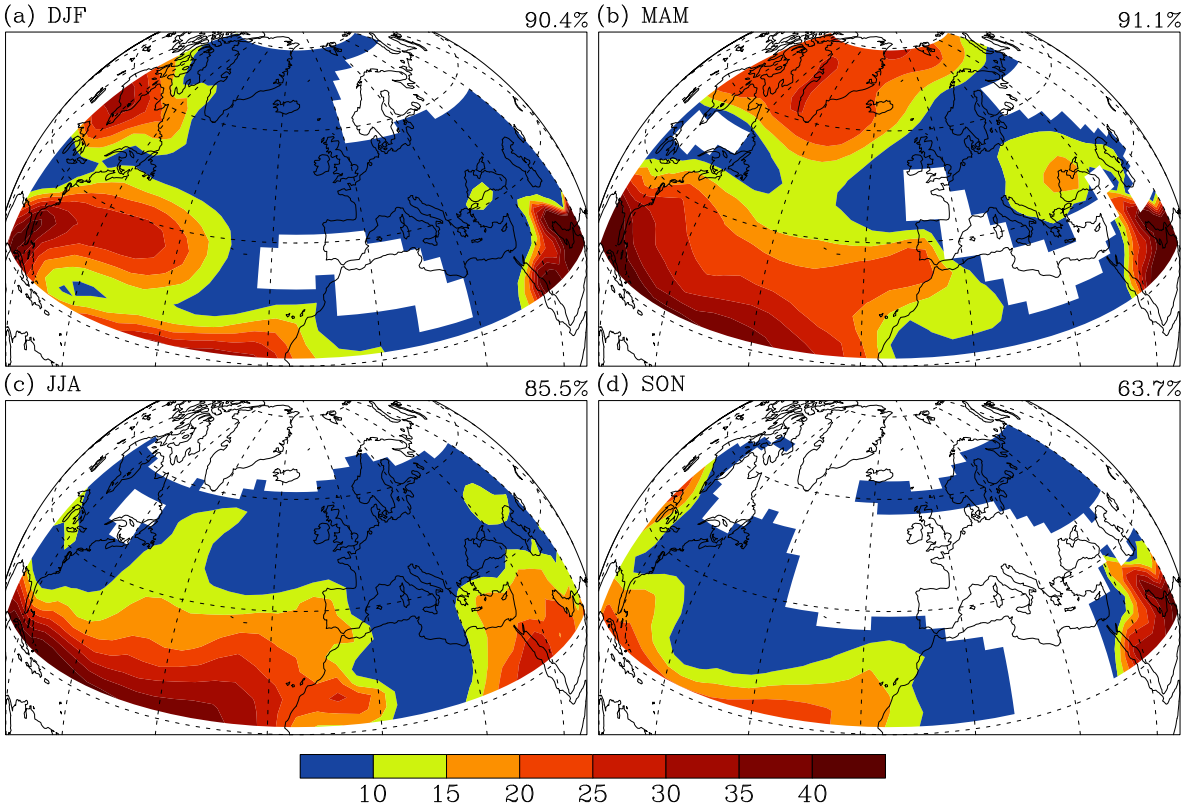

Figure 4.2: Percentage of total variance of seasonal mean MSLP anomalies due to SST-SIC forcing (potential predictability), obtained by a Random-Effects One-Way ANOVA applied to the MUGCM ensemble, for (a) DJF, (b) MAM, (c) JJA and (d) SON. White areas show where SST-SIC forcing is not significantly greater than zero at the 0.05 significance level ($R \leq 2.25\%$, see hypothesis test 4.16. The percentage of the EA area with significant SST-SIC forcing is shown on the top right-hand corner of each panel. Grid lines represented every 20º.

basin between 20°N and 40°N, reaching Portugal, and at high latitudes, matching large regions of the NAO centres of action. On the other hand, the eastern and western half parts of Spain and France, respectively, have insignificant values of R. In summer, the Atlantic basin south of 40°N have the highest values of R, but almost everywhere north of 60°N, except Scandinavia, present insignificant R values. In autumn, large areas over the middle Atlantic and Europe have no significant SST-SIC forcing.

In section 3.4.2 we performed an EOF analysis to the seasonal mean anomalies of MUGCM MSLP for the four standard seasons. We found that the primary teleconnection patterns in the EA region (North Atlantic Oscillation, Scandinavia or EURASIA I, East Atlantic, and East Atlantic/West Russia or EURASIA II [Wallace and Gutzler, 1981; Barnston and Livezey, 1987]) were better identified as the first four EOFs in winter and spring. Since these EOFs represent uncorrelated modes of the total variability in the EA sector[2], we applied the ANOVA to the principal components (PCs) of these EOFs, in winter and spring, to see if the teleconnection patterns they represent are significantly forced by SST-SIC variations. The calculations are performed considering X_{pk} of model 4.1, the PC of each variability mode, and performing the hypothesis test 4.16 using potential predictability (equation 4.20) as the test statistic and the test rule 4.22.

The results, presented in table 4.1 (first value in each entry of the table), show that all patterns of variability, except EOF-2 in winter (winter SCAND), have a statistical significant contribution from the lower-boundary variability (values greater than 2.25%). As expected from figure 4.2, where spring stands out as the season with the highest SST-SIC forcing over almost the entire EA area, the NAO, Scandinavia and East Atlantic/Western Russia modes have more external forcing in spring than in winter. For the East Atlantic mode, the opposite occurs just because the lower-latitude centre of action, where the oceanic forcing is higher, have higher loadings in winter than in spring.

[2]Actually these EOFs are biased estimators of the EOFs of total variability (see sections 4.3 and 5.1. Nevertheless, this bias is not important for the present purposes.

Table 4.1: Percentage of total variance of the PCs of winter (DJF) and spring (MAM) seasonal mean MSLP anomalies over the EA region, due to SST-SIC forcing, obtained by a Random-Effects One-Way ANOVA (first value of each season/PC entry). All values significantly exceed zero at the 0.05 significance level ($R > 2.25\%$). The second value of each entry represents the percentage of the total MSLP EA variance represented by the PC. The third value represents the percentage of the total MSLP EA variance due to the SST-SIC forced variability of the PC.

Season	PC-1 (NAO)			PC-2 (SCAND)			PC-3 (EA)			PC-4 (EA-WR)		
DJF	9.4	31.3	3.0	1.1	23.5	0.3	9.2	11.5	1.1	10.0	7.9	0.8
MAM	30.0	31.2	9.4	3.1	19.2	0.6	6.1	14.8	0.9	9.9	7.3	0.7

NAO - North Atlantic Oscillation; SCAND - Scandinavia;
EA - East Atlantic; EA-WR - East Atlantic/Western Russia

For completeness, we also present, in the second value in each season/PC entry of table 4.1, the percentage of the total seasonal mean MSLP variability in the EA region that is represented by the corresponding mode in that season (see figures 3.22 and 3.23). The third value in each entry shows the contribution of the forced variability of the same mode in that season to the total MSLP variability in the EA region. For example, 30.0% of the NAO (EOF-1) variability in spring is SST-SIC forced. However, since the NAO itself only represents 31.2% of the total EA variability in spring, this value represents only 9.4% of the total variability. These low values are in agreement with the low values of figure 4.2 (blue areas) and stand out as a first and indication of the difficulty in determining oceanic forcing signals in the EA region.

The fraction of the modelled NAO variance due to SST-SIC forcing in winter (9.4%) and in spring (30.0%) should be interpreted with caution. We must realise that these values are computed from modelled PCs whose associated EOF patterns are biased with respect to their NCEP counterparts (see discussion in section 3.4.2). For this reason, and also because the NAO is the major mode of atmospheric variability in the EA region in winter and spring (see figures 3.22 and 3.23), we also applied the

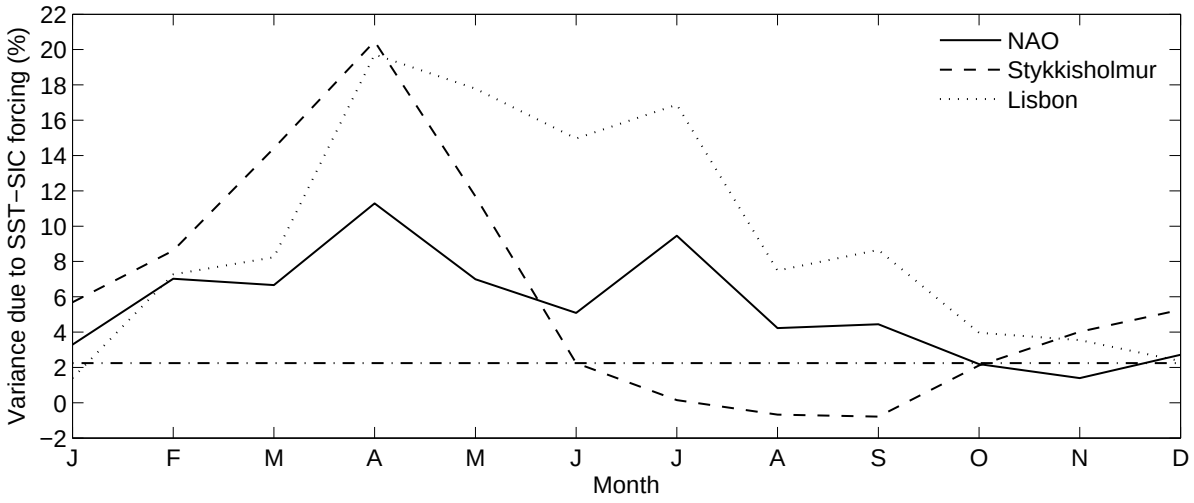

Figure 4.3: Percentage of total variance of the NAO (solid line), Stykkisholmur (dashed line) and Lisbon (dotted line) indices due to SST-SIC forcing (potential predictability, R), obtained by a Random-Effects One-Way ANOVA, for 3-month running mean calendar seasons. The dashed-dotted horizontal line indicates the value above which the percentage exceeds zero at the 0.05% significance level (R = 2.25%). Seasons identified by its middle month.

ANOVA to a station-based NAO index. This index is usually formed by subtracting the normalised time series of MSLP at Stykkisholmur, Iceland (65.1^{o}N, 22.7^{o}W) from that at Lisbon, Portugal (38.7^{o}N, 9.1^{o}W). Here we used the MUGCM grid points nearest to Stykkisholmur and Lisbon: (65.96N, 22.50W) and (39.13^{o}N, 7.5^{o}W), respectively. The analysis was performed to twelve 3-month seasons. The seasonal means were computed by a 3-month running mean.

Figure 4.3 shows the percentage of the seasonal mean NAO index variance due to SST-SIC forcing (solid curve). The horizontal dashed-dotted line indicates the value above which the percentage significantly exceeds zero at the 0.05 ($R = 2.25\%$) significance level, and seasons are identified by its middle month. This figure shows that the NAO is forced by the lower boundary fluctuations in all seasons except for SON (autumn) and OND, the highest forcing found in MAM, in agreement with figure 4.2. Note that negative values have no physical meaning arising simply from sampling variability (see equations 4.20 and 4.9).

The same analysis was performed to the time series of seasonal MSLP at Stykkisholmur and at Lisbon used to build the NAO index. The purpose of this analysis is to see if one of the poles, Stykkisholmur or Lisbon, contributes more to the external forcing of the NAO index than the other. It is clear from figure 4.3 that the NAO index curve follows that of Lisbon. The correlation of the NAO index curve with the Lisbon (Stykkisholmur) index curve is 0.84 (0.56). This suggest that NAO is forced mainly through its southern pole, specially from late spring to late summer, when the percentages of the Lisbon index curve are much larger than the ones of the Stykkisholmur index curve.

4.5 Summary and conclusions

In section 4.2 we discussed the nature of the simulated seasonal mean data and found that can be analysed by the *one-way* ANOVA. We argued that the model of the *one-way* ANOVA that better suits the nature of the data and the purpose of the research is the *random-effects* model. In the framework of this model, we showed how to partition the total variability into forced and internal components. Unbiased estimators of these variances and an hypothesis test to determine the significance of the forced variance were presented. The test was supplied with two different test statistics: the traditional one and the so called *potential predictability* which is the fraction of the total variability that is due only to SST-SIC forcing.

In section 4.3 the *random-effects one-way* ANOVA was applied to the MUGCM seasonal mean MSLP data over the EA region for the all standard seasons: winter (DJF), spring (MAM), summer (JJA) and autumn (SON). Maps of the estimated internal and forced variances were presented (figure 4.1) and it was found that winter is the season with higher internal variability and that winter and spring are the seasons with higher forced variability. The maps for the percentage of total variance due to boundary forcing were also presented (figure 4.2), where only the values significantly, at the 0.05 significance level, greater that zero were plotted. It was found that seasonal mean MSLP is significantly SST-SIC forced in almost the entire area of the EA, except in

SON, the higher forcing occurring in winter and spring.

In section 3.4.2, using the EOF technique, we represented the modelled seasonal mean anomaly MSLP variability in the EA sector by orthogonal patterns of variability and identified the NAO as the major mode in winter and spring (figures 3.22(a) and 3.23(a)), representing 31.3% and 31.2% of total variance, respectively.

In section 4.4 we computed the percentage of the seasonal mean variance of the four modes of variability, identified by the EOF analysis, due to SST-SIC forcing, in winter and spring, and found that all modes are significantly forced in both seasons (table 4.1). The NAO was found to be the mode with strongest forcing in spring (30.0%).

The ANOVA was also performed to the NAO index in overlapping 3-month mean seasons (figure 4.3) and it was found that the NAO is significantly forced in all seasons except in SON (autumn) and OND, the higher forcing found again in MAM.

In summary, the MUGCM MSLP variability in the EA region, in general, and its major mode of variability (the NAO), in particular, have a weak but significant SST-SIC contribution, the forcing being strong in winter and specially in spring. These results agree with the seasonal predictability results found by Davis et al. [1997] when analysing an ensemble of MSLP simulated by the Hadley Centre atmospheric climate model HADAM1.

Chapter 5

Global Ocean Forcing

In chapter 4 an ANOVA was performed to the 29-member ensemble of MSLP seasonal means generated by the MUGCM. The ANOVA revealed significant boundary layer forcing in the EA region, specially in winter and spring. This forcing was found in MSLP raw data and also in its major mode of total variability, the NAO. With this knowledge at hand, the logical next step is to determine the spatial pattern of the MSLP anomaly variability in the EA that is solely originated by this SST-SIC forcing.

In this chapter, the first and second leading modes of MSLP variability in the EA region forced by SST and SIC variability are estimated using three methods: PCA of the ensemble mean (standard PCA), Singular Value Decomposition Analysis (SVDA) proposed by Ward and Navarra [1997] and the *optimal detection* algorithm introduced by Venzke et al. [1999] (optimal PCA). The analysis is done on seasonal mean MSLP anomaly data in winter and spring. The spring season is chosen because it is when the global SST-SIC forcing on seasonal MSLP variability in the EA sector is higher, as consistently demonstrated in section 4.4. Winter is also chosen not only because it is the second standard season with higher boundary forcing, but also because it is the season with higher seasonal mean MSLP total variability in the EA region (internal plus forced, see figure 4.1). Furthermore, winter and spring are the seasons when the teleconnection patterns of the NAO, Scandinavia, East Atlantic or Eurasia I, and East Atlantic-Western Russia or Eurasia II [Wallace and Gutzler, 1981; Horel, 1981;

Blackmon et al., 1984; Barnston and Livezey, 1987], are better identified (section 3.4.2).

Due to the large ensemble size (29 members), the first and second forced modes of variability (spatial patterns and associated time series) obtained by SVDA are almost identical to those obtained by standard PCA. The reader may skip section 5.2, where the SVDA modes are determined, since, as they are almost equal to the standard PCA modes, they will not be used in subsequent analyses. Also due to the size of the ensemble, the standard PCA leading mode is also almost identical to the leading mode of optimal PCA, but the second modes obtained by the two methods are different.

In order to localise the regions of the global ocean where SST anomaly variability is responsible for the detected MSLP forced modes in the EA region, SST anomalies at each grid point of the globe are regressed onto the standard and optimal PCs. From the resulting global regression patterns, three localised statistical significant regression patterns outstand: (i) the regression pattern in the equatorial Pacific, found by the regression of the SST anomaly field onto standard PC-1 and optimal PC-1, in winter and spring, resembling the SST anomaly pattern of the ENSO; (ii) the regression pattern in the tropical Atlantic, found by the regression of the SST anomaly field onto the optimal PC-1 in spring, resembling the pattern of the Atlantic Interhemispheric SST anomaly Gradient (AISG); and (iii) the regression pattern in the north Atlantic basin, found by the regression of the SST anomaly field onto the optimal PC-2 in winter, resembling the SST anomaly pattern of the North Atlantic SST anomaly Tripole (SST Tripole).

Further analysis comparing the patterns and time series associated to the three mentioned SST modes with the regression patterns and the forced PCs, give evidence that (i) the ENSO forces the standard and optimal PC-1, in winter and in spring, (ii) the AISG also forces the optimal PC-1 in spring, and (iii) the SST Tripole forces the optimal PC-2 in winter.

Optimal detection PCA proves to be efficient in extracting SST-forced signals from an ensemble of data with high internal variability, such as the MUGCM AMIP II ensemble

of MSLP in the EA sector. Without optimal detection, the forcing from the two Atlantic SST modes would not be detected. Furthermore, an ANOVA-based test shows that the leading optimal detected response is not only significant but also consistent, in both seasons, and comparison with observations give reliability to the optimal PC-1 in winter and spring and also to the optimal PC-2 in spring.

5.1 Forced response by standard PCA

When an ensemble of GOGA simulations are available, such as our AMIP II simulations, the response pattern of some atmospheric variable to the global SST-SIC forcing is usually determined performing an EOF analysis of the ensemble mean of that variable. The reason to use the ensemble mean lies in the multivariate version of equation 4.11,

$$\mathcal{E}\left[\mathbf{S}_{EM}\right] = \mathbf{\Sigma}_F + \frac{1}{K}\mathbf{\Sigma}_I, \tag{5.1}$$

where $\mathcal{E}$ is the mathematical expectation operator, $\mathbf{S}_{EM}$ is the ensemble mean sample covariance matrix, and $\mathbf{\Sigma}_F$ and $\mathbf{\Sigma_I}$ are the *true* or populational forced and internal covariance matrices, respectively.

Equation 5.1 shows that, in the limit of an infinite ensemble, the forced variability would be given by the ensemble mean variability, $\mathbf{S}_{EM} \rightarrow \mathbf{\Sigma}_F$ as $K \rightarrow \infty$. So this is why the EOFs of the ensemble mean (normalised eigenvectors of $\mathbf{S}_{EM}$) are usually used to estimate the forced PVs (normalised eigenvectors of $\mathbf{\Sigma}_F$), and the usual practice is to perform the maximum economically possible number of simulations to achieve a better estimation.

For an uncorrelated in space internal variability, $\mathbf{\Sigma}_I = \sigma_I^2\mathbf{I}$, where $\mathbf{I}$ is the unitary matrix, the eigenvalues of $\mathbf{S}_{EM}$ would be σ_I^2 times the eigenvalues of $\mathbf{\Sigma}_F$ but its normalised eigenvectors would be the same. However, internal variability is not uncorrelated in space. Worse still, the directions defined by the higher ranked ensemble mean EOFs, where the variance of the ensemble mean PCs is higher and where we expect to find the

forced response, are also the directions where the internal variability is higher. Before we demonstrate this, let us first define the notation of the variables at work.

Let X_{mpk} be a random seasonal mean atmospheric variable at spatial location m ($m = 1, ..., M$) in year p ($p = 1, ..., P = 17$) of the kth ($k = 1, ..., K = 29$) simulation. Note that this is the same notation used in the Analysis of Variance of chapter 4 with an extra subscript, m, to indicate the spatial location. In appendix A, where the mathematical formalism of Principal Component Analysis was presented, X_{mn} represents the nth ($n = 1, ..., N$) observation of a random seasonal atmospheric variable at spatial location m. The correspondence between X_{mn} of appendix A and X_{mpk} is that $n = (k-1)P+p$, as we did in the EOF analysis of the total variability of seasonal MSLP in section 3.4.2.

Each random variable X_{mpk} can be expressed as the sum of the ensemble mean in year p and the departure of the kth ensemble member from this mean:

$$X_{mpk} = \overline{X}_{mp\bullet} + X'_{mpk}, \tag{5.2}$$

where the dot notation represents, as before, averaging over the missing subscript.

Let $\mathcal{X}_{EM}$ be a $(M \times P)$ random data matrix whose (m, p) entry is $\overline{X}_{mp\bullet}$ and $\mathcal{X}_{I,k}$ be a $(M \times P)$ random data matrix whose (m, p) entry is X'_{mpk}. That is, $\mathcal{X}_{EM}$ is the data matrix of the ensemble mean and $\mathcal{X}_{I,k}$ is the data matrix of the departures of the kth simulation from the ensemble mean. Also let $\mathcal{X}_I = \left[\mathcal{X}_{I,1}|...|\mathcal{X}_{I,k}|...|\mathcal{X}_{I,K}\right]$ be a $(M \times PK)$ random data matrix of the departures of all simulations.

The sample covariance matrices of $\mathcal{X}_{EM}$, $\mathcal{X}_{I,k}$ and $\mathcal{X}_I$ are, respectively:

$$\mathbf{S}_{EM} = \frac{1}{P-1}\mathcal{X}_{EM}\mathcal{X}_{EM}^{\mathrm{T}}, \tag{5.3}$$

$$\mathbf{S}_{I,k} = \frac{1}{P(K-1)}\mathcal{X}_{I,k}\mathcal{X}_{I,k}^{\mathrm{T}}, \tag{5.4}$$

$$\mathbf{S}_I = \frac{1}{P(K-1)}\mathcal{X}_I\mathcal{X}_I^{\mathrm{T}}. \tag{5.5}$$

Note that equation 6.7 is a multivariate version of equation 4.10 and is an unbiased estimator of $\mathbf{\Sigma}_I$. Note also that

$$\mathbf{S}_I = \sum_{k=1}^{K} \mathbf{S}_{I,k}. \tag{5.6}$$

The ensemble mean sample covariance matrix, $\mathbf{S}_{EM}$, is $(M \times M)$ symmetric matrix and therefore admits M eigenvectors. However, as seen in appendix A, since $\mathcal{X}_{EM}$ is a $(M \times P)$ matrix with $P < M$, only the first P eigenvalues of $\mathbf{S}_{EM}$ are non-zero. Thus, the EOFs of $\mathcal{X}_{EM}$ are the first P normalised eigenvectors of $\mathbf{S}_{EM}$. Let them to be denoted by $\mathbf{e}_{EM,r}$, with $r = 1, ..., P$, and assembled as columns in the $(M \times P)$ matrix $\mathbf{E}_{EM}$. The associated principal components are, as seen in equation A.9

$$\mathcal{Y} = \mathbf{E}_{EM}^{\mathrm{T}} \mathcal{X}_{EM}, \tag{5.7}$$

whose covariance matrix is, as also seen in equations A.5 and A.6,

$$\mathbf{S}_Y = \mathbf{E}_{EM}^{\mathrm{T}} \mathbf{S}_{EM} \mathbf{E}_{EM} = \mathbf{\Lambda}_{EM} \tag{5.8}$$

where $\mathbf{\Lambda}_{EM}$ a $(P \times P)$ diagonal matrix whose elements are the non-zero eigenvalues of $\mathbf{S}_{EM}$.

The first two leading modes for the MUGCM seasonal DJF and MAM MSLP are presented in figure 5.1. In winter, the first and second mode account for 51.7% and 15.2% of the ensemble mean variance, respectively, and in spring they represent 69.1% and 12.5%, respectively.

Figure 5.2 (a,b) presents, for winter and spring, the first PC of the ensemble mean, $\mathbf{e}_{EM,1}^{\mathrm{T}} \mathcal{X}_{EM}$, and the projections of the individual ensemble members $k = 1, ..., K = 29$ onto the first EOF of the ensemble mean, $\mathbf{e}_{EM,1}^{\mathrm{T}} \mathcal{X}_{I,k}$. Figure 5.2 (c,d) shows analogous plots but for the second ensemble mean EOF.

The spread of the projections of the individual ensemble members onto the first two leading EOFs of the ensemble mean show that internal variability is high in the space spanned by these two directions. In fact, as stated above, the higher ranked ensemble

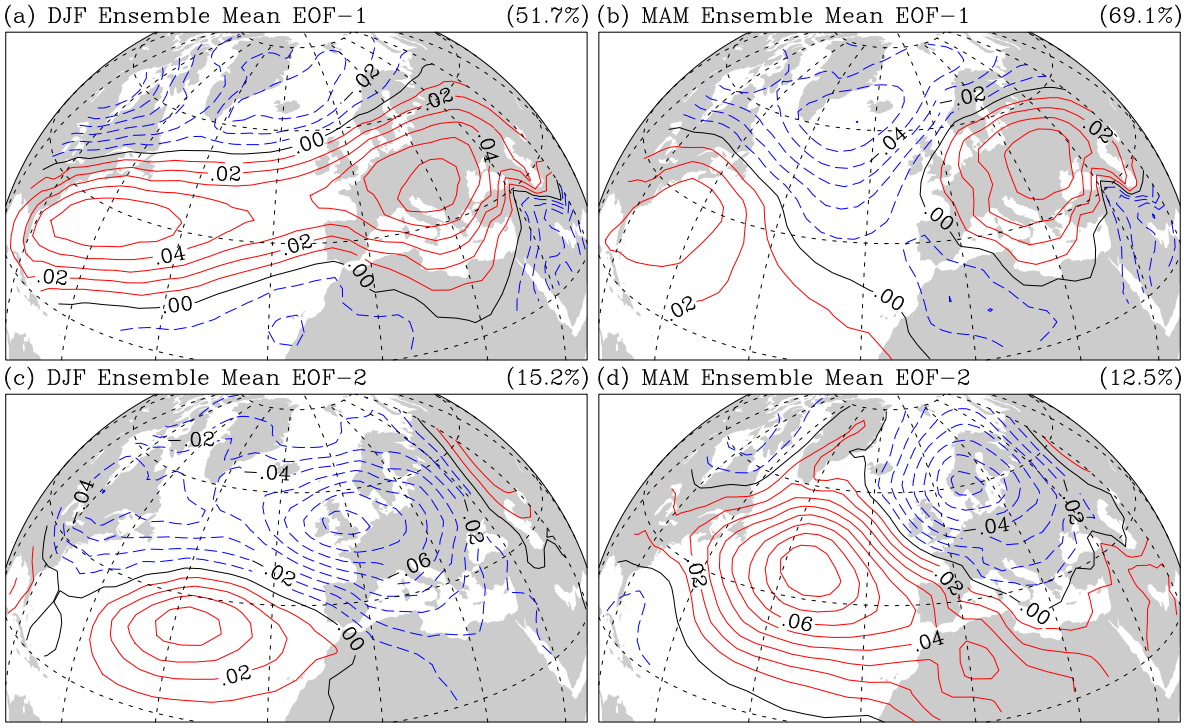

Figure 5.1: First two leading standard EOFs of MUGCM MSLP anomalies in (a,c) DJF and (b,d) MAM. Negative, zero and positive loadings plotted with dashed blue, solid black and solid red lines. The percentage of total variance explained by each EOF is printed on the upper right-hand corner of each panel. Grid lines represented every 20°.

mean EOFs, where the associated PCs have higher variance and where we expect to find the forced response, define directions where the internal variability is also high. This feature can be confirmed by figure 5.2 (e,f) which shows the variances of the first 15 ensemble mean PCs (red line), $\mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{e}_{EM,r} = \lambda_{EM,r}$ and the variances of the projections of the $k = 1, ..., K = 29$ individual ensemble members onto the first 15 EOFs of the ensemble mean (black lines) $\mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{S}_{I,k}\mathbf{e}_{EM,r}$, where $r = 1, ..., 15$. Also shown in figure 5.2 (e,f, red dotted line) is the average of these 29 variances:

$$\frac{1}{K}\sum_{k=1}^{K}\mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{S}_{I,k}\mathbf{e}_{EM,r} = \frac{1}{K}\mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{S}_{I}\mathbf{e}_{EM,r},$$

which, according to equation 5.1, is an estimation of the variance, in the directions of $\mathbf{e}_{EM,r}$, that we would obtain in the absence of any true forced response.

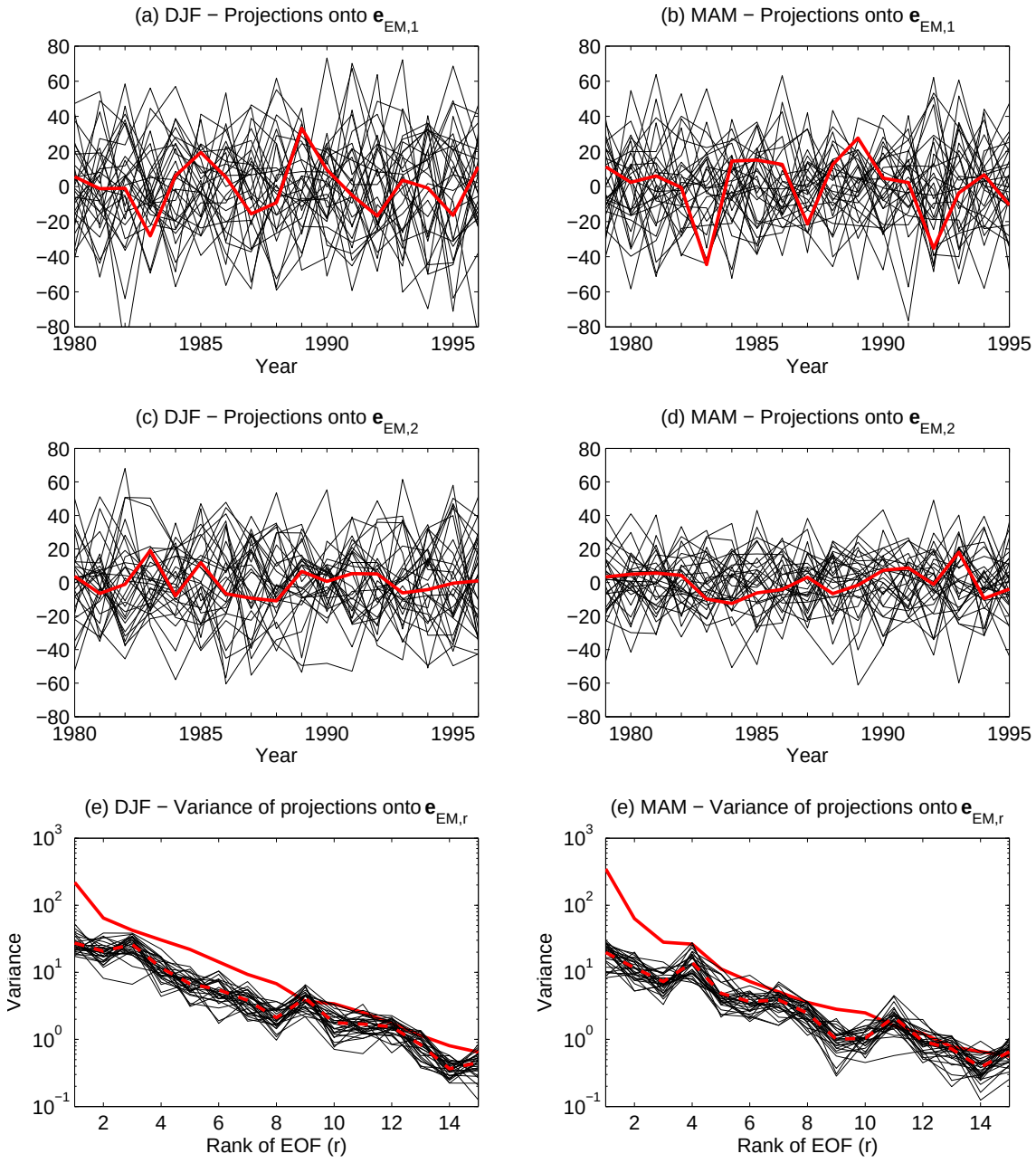

Figure 5.2: Results of PCA of DJF and MAM seasonal mean MSLP ensemble mean in the EA region. (a,b) Projection of the ensemble mean (red line) and of the individual ensemble members (black lines) onto EOF-1. (c,d) As in (a,b) but for EOF-2. (e,f) Variance of the projections of the ensemble mean (red solid line), of the individual ensemble members (black lines) and their mean (red dashed line) onto the first 15 EOFs.

The purpose of comparing $\mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{e}_{EM,r}$ with $(1/K)\mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{S}_{I}\mathbf{e}_{EM,r}$ is to determine if exists a forced response in the state space direction defined by $\mathbf{e}_{EM,r}$. Or, if the pattern $\mathbf{e}_{EM,r}$ is a good estimator of a true forced pattern. This question is formulated by an ANOVA test analogous to the ANOVA test 4.16:

$$H_0 : \mathbf{e}_{EM,r}^{\mathrm{T}}\ \mathbf{\Sigma}_F\ \mathbf{e}_{EM,r} = 0 \quad vs \quad H_1 : \mathbf{e}_{EM,r}^{\mathrm{T}}\ \mathbf{\Sigma}_F\ \mathbf{e}_{EM,r} > 0. \tag{5.9}$$

The test statistic is, analogously to the test statistic 4.18,

$$F = K\frac{\mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{e}_{EM,r}}{\mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{S}_{I}\mathbf{e}_{EM,r}} \sim \mathcal{F}_{P-1,P(K-1)} \tag{5.10}$$

under H_0. Since $\mathcal{E}[\mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{e}_{EM,r}] = \mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{\Sigma}_{I}\mathbf{e}_{EM,r}$ if H_o is true and $\mathcal{E}[\mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{S}_{I}\mathbf{e}_{EM,r}] = \mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{\Sigma}_{I}\mathbf{e}_{EM,r}$ then the expected value of the F is 1 if H_o is true. If H_o is not true, F is expected to have a value greater than 1. Thus, the criterion for rejecting H_o with a significance level of α is

$$F > \mathcal{F}_{P-1,P(K-1),\alpha}. \tag{5.11}$$

Table 5.1 presents the results of the test 5.9 applied to the first and second EOFs of the MUGCM ensemble mean of winter and spring MSLP seasonal means. Since all values of the F statistic exceed the cutoff value $\mathcal{F}_{16,17\times29,0.05} = 1.66$, the null hypothesis is rejected, meaning that the forced response in both patterns and in both seasons significantly exceed zero at the 5% significance level.

Venzke et al. [1999] defines a forced response that significantly exceeds zero at the 5% significance level ($F > \mathcal{F}_{0.05}$) as a *detectable* response (at the 95% level) and recall that, since the F statistic increases linearly with the ensemble size, "it is always possible to *detect* a non-zero but arbitrarily weak forced response given a sufficiently large ensemble". They then define a *consistent* response, at the 95% level, as a response whose statistic F/K exceeds the cutoff value $\mathcal{F}_{0.05}$. The physical difference between a *detectable* response and a *consistent* response on a pattern $\mathbf{e}_{EM,r}$, at the 95% level, is that, while in the former we have a 95% chance of finding a positive correlation between $\mathbf{e}_{EM,r}^{\mathrm{T}}\mathcal{X}_{EM}$ and the true forced time series of that pattern, in the later we also have a 95% chance of finding a positive correlation between $\mathbf{e}_{EM,r}^{\mathrm{T}}\mathcal{X}_{I,k}$ and the true forced time series associated with $\mathbf{e}_{EM,r}$.

Table 5.1: Variances of the projections of the ensemble mean and of individual ensemble members onto the first two leading EOFs of the MUGCM ensemble mean of DJF and MAM MSLP seasonal means. Values of the F statistic greater than 1.66 (bold) indicate a *detectable* response, that is, a forced response that significantly exceed zero at the 0.05 significance level.

r	$\mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{e}_{EM,r}$	$\mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{S}_{I}\mathbf{e}_{EM,r}$	F	F/K
DJF EOF 1	218.94	794.28	**7.99**	0.28
MAM EOF 1	347.98	573.04	**17.61**	0.61
DJF EOF 2	64.49	599.69	**3.12**	0.11
MAM EOF 2	63.11	338.08	**5.41**	0.19

Under the definitions of Venzke et al. [1999], it can be seen in table 5.1 that we *detected* a forced response on EOF-1 and EOF-2 of the MUGCM ensemble mean of winter and spring MSLP seasonal means, but these responses are not *consistent*. Obviously, these forced responses (figure 5.2 (a-d)) failed the test of consistency because of the large values of $\mathbf{e}_{EM,1}^{\mathrm{T}}\mathbf{S}_{I}\mathbf{e}_{EM,1}^{\mathrm{T}}$ and $\mathbf{e}_{EM,2}^{\mathrm{T}}\mathbf{S}_{I}\mathbf{e}_{EM,e}^{\mathrm{T}}$ compared to those of $\mathbf{e}_{1}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{e}_{1}^{\mathrm{T}}$ and $\mathbf{e}_{2}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{e}_{2}^{\mathrm{T}}$, respectively (figure 5.2 (e,f) and table 5.1). In section 5.3 we present and apply the method proposed by Venzke et al. [1999] that also estimates the forced patterns using the ensemble mean but maximizes the signal-to-noise ratio $(\mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{e}_{EM,r}^{\mathrm{T}})/(\mathbf{e}_{EM,r}^{\mathrm{T}}\mathbf{S}_{I}\mathbf{e}_{EM,r}^{\mathrm{T}})$. The purpose is, of course, to *detect* forced signals and see if they are strong enough to be considered *consistent*.

5.2 Forced response by SVDA

An alternative method to PCA of the ensemble mean to obtain forced modes of variability was proposed by Ward and Navarra [1997]. As referred at the beginning of this chapter, the first and second modes obtained by this method are almost equal to those obtained by standard PCA, and because of this these results will not be used in subsequent analyses. The reader may confirm the equality of the results of both methods by comparing figures 5.1 and 5.3, and by inspection of figure 5.4, and then move directly to section 5.3.

The method proposed by Ward and Navarra [1997] is based on the Singular Value Decomposition Analysis (SVDA) of Bretherton et al. [1992]. The general SVDA of Bretherton et al. [1992] consists of performing the singular value decomposition of the cross-covariance matrix of the time series of two data fields[1]. The resulting leading pair of patterns is such that the covariance between the time series associated to each pattern is maximum. The leading pair of modes (patterns and time series) is, thus, a pair of modes of maximum covariability. Bretherton et al. [1992] and Wallace et al. [1992] compared the method with other methods such as Principal Component Analysis with fields combined (CPCA) and Canonical Correlation (CCA), among others. The method was also compared with CCA by Cherry [1996]. An early example of the application of the method can be found in Wallace et al. [1992] who performed a SVDA to wintertime sea surface temperature anomalies in the North Pacific and 500 mb height anomalies in the northern hemisphere.

Using the notation of equation 6.6, let X_{mpk} represent a random seasonal mean atmospheric anomaly at spatial location m ($m = 1, ..., M$) in year p ($p = 1, ..., P$) of the kth ($K = 1, ..., K$) simulation. Let $\mathcal{X}_k$ be a $(M \times P)$ random matrix whose (m, p) entry is X_{mpk}, that is, it holds the data of the kth simulation. The first step of the method proposed by Ward and Navarra [1997] consists of building two matrices, $\mathcal{X}_{L/2}$ and $\mathcal{X}_{R/2}$, where $\mathcal{X}_{L/2}$ is built putting $\mathcal{X}_k$ matrices one after another in the row direction, and $\mathcal{X}_{R/2}$ is built in the same way but using $\mathcal{X}'_k$ matrices (the L and R subscripts stand for *left* and *right*, respectively). Each pair of indices (k, k') is a combination formed by selecting two diferent elements, at a time, from the set of numbers $\{1, 2, \ldots, K\}$. Thus, the combinations are

$$(1,2); (1,3); \ldots; (1,K); \quad (2,3); (2,4); \ldots; (2,K); \quad (K-1,K), \tag{5.12}$$

[1]Bretherton et al. [1992] used the abbreviation SVD to denote both the general matrix operation and the method they proposed. Here, we will differentiate them using the abbreviation SVDA to denote the method and *singular value decomposition* to refer to the matrix operation.

and the matrices $\mathcal{X}_{L/2}$ and $\mathcal{X}_{R/2}$ are then formed as follows:

$$\mathcal{X}_{L/2} = \Big[\overbrace{\mathcal{X}_1|\ldots|\mathcal{X}_1}^{K-1} \quad | \quad \overbrace{\mathcal{X}_2|\ldots|\mathcal{X}_2}^{K-2} \quad | \quad \ldots \quad | \quad \mathcal{X}_{K-1}\Big] \tag{5.13}$$

$$\mathcal{X}_{R/2} = \Big[\mathcal{X}_2|\ldots|\mathcal{X}_K \quad | \quad \mathcal{X}_3|\ldots|\mathcal{X}_K \quad | \quad \ldots \quad | \quad \mathcal{X}_K\Big]. \tag{5.14}$$

The second step of the method consists of building other two matrices from $\mathcal{X}_{L/2}$ and $\mathcal{X}_{R/2}$ as follows,

$$\mathcal{X}_L = \Big[\mathcal{X}_{L/2} \quad | \quad \mathcal{X}_{R/2}\Big] \tag{5.15}$$

$$\mathcal{X}_R = \Big[\mathcal{X}_{R/2} \quad | \quad \mathcal{X}_{L/2}\Big]. \tag{5.16}$$

It can be seen straightforwardly from the series of combinations 5.12 that there are $(K-1)+(K-2)+\ldots+1$ combinations. We can, alternatively, determine this number recurring to a theorem of *Combinatorial Analysis* (e.g., Apostol [1969], pg. 481; Milton and Arnold [1995], pg. 14) that states that the number of distinct subsets of r elements that may be formed from a set of $K \geq r$ distinct elements, denoted by ${}_KC_r$, is given by

$${}_KC_r = \binom{K}{r} = \frac{K!}{r!(K-r)!}.$$

For $r = 2$, the number of combinations is

$${}_KC_2 = \binom{K}{2} = \frac{K(K-1)}{2},$$

which is equal to $(K-1)+(K-2)+\ldots+1$.

In the third step of the method, the cross-covariance matrix between $\mathcal{X}_L$ and $\mathcal{X}_R$ is determined,

$$\mathbf{S}_{LR} = \frac{1}{(K-1)(KP-1)}\mathcal{X}_L\mathcal{X}_R^{\mathrm{T}}, \tag{5.17}$$

which is a M-dimensioned square matrix. Note that $\mathcal{X}_L$ and $\mathcal{X}_R$ are formed with X_{mpk} anomalies relative to the global mean, that is, for each station m, we have

$$\sum_{k=1}^{K}\sum_{p=1}^{P} X_{mpk} = 0,$$

so, each one of the KP observations in station m can be expressed in terms of the rest $KP-1$ observations. Note also that $\mathbf{S}_{LR}$ is taken over ${}_2K C_2 P$ products between X_{mpk} anomalies. Thus, there are

$$\frac{{}_2K C_2 P}{KP}(KP-1) = (K-1)(KP-1)$$

degrees of freedom, which are used in equation 6.4.

Finally, the singular value decomposition of $\mathbf{S}_{LR}$ is performed,

$$\mathbf{S}_{LR} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}, \tag{5.18}$$

where $\mathbf{U}$ and $\mathbf{V}$ are $(M \times M)$ matrices whose columns form an orthonormal set of vectors called the *left* and *right singular vectors* of $\mathbf{S}_{LR}$, respectively, and $\mathbf{\Lambda}$ is a $(M \times M)$ diagonal matrix whose elements are called the *singular values* of $\mathbf{S}_{LR}$. Now note that, the constructs 5.15 and 5.16 of $\mathcal{X}_l$ and $\mathcal{X}_r$ assure that $\mathcal{X}_L \mathcal{X}_R^{\mathrm{T}} = (\mathcal{X}_L \mathcal{X}_R^{\mathrm{T}})^{\mathrm{T}}$ and, thus, $\mathbf{S}_{LR}$ is symmetric. Consequently, its singular right vectors are equal to its singular left vectors which are the seeked forced patterns, $\mathbf{E} = \mathbf{U} = \mathbf{V}$. The fraction of total variance represented by the mth spatial pattern is given by $\lambda_m / \sum_{m'=1}^{M} \lambda_m$, where λ_m is the mth singular value of $\mathbf{S}_{LR}$, that is, the value of the (m,m) entry in Λ.

The method was applied to MUGCM MSLP seasonal anomalies in winter and spring. Figure 5.3 shows the two leading spatial patterns for each season. These patterns are equal to the forced patterns obtained by standard PCA of the ensemble mean (section 5.1, figure 5.1).

Projecting the individual ensemble members, $\mathcal{X}_k, \quad k = 1, \ldots, K$, onto the forced patterns, we obtain

$$\mathcal{Y}_k = \mathbf{E}^{\mathrm{T}} \mathcal{X}_k, \qquad k = 1, \ldots K, \tag{5.19}$$

which are $(M \times P)$ matrices whose lines are the time series associated with the forced patterns.

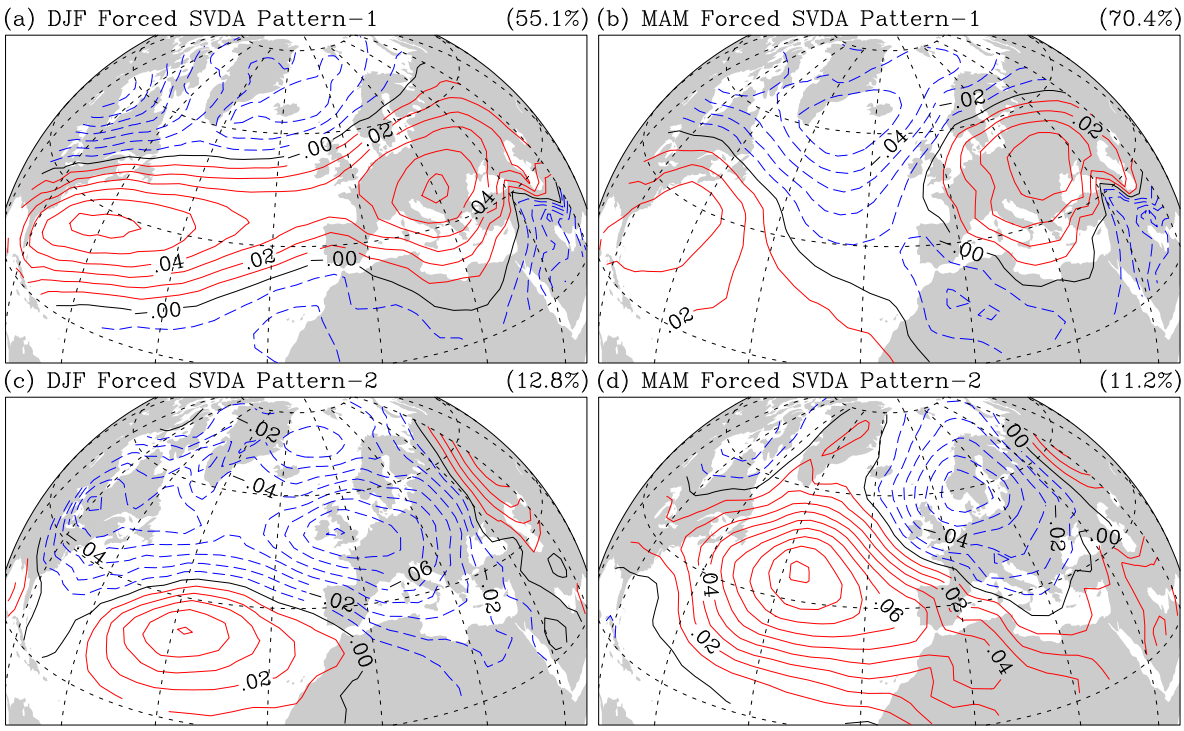

Figure 5.3: First two leading SVDA spatial patterns of MUGCM MSLP seasonal mean anomalies in (a,c) DJF and (b,d) MAM. Negative, zero and positive loadings plotted with dashed blue, solid black and solid red lines. The percentage of total variance explained by each SVDA mode is printed on the upper right-hand corner of each panel. Grid lines represented every 20°.

Figure 5.4 shows, for winter and spring, the time series associated to the two leading forced patterns obtained by standard PCA and SVDA. The time series associated to the standard PCA forced patterns (dashed lines) are the principal components of the ensemble mean of MSLP anomalies. The time series associated to the SVDA patterns are the ensemble mean of the projections of individual ensemble members of MSLP anomalies onto these patterns, that is, the first two lines of the $(M \times P)$ matrix

$$\mathcal{Y}_{EM} = \frac{1}{K} \sum_{k=1}^{K} \mathcal{Y}_k, \tag{5.20}$$

where the matrices $\mathcal{Y}_k, \quad k = 1, \ldots, K$, are given by equation 5.19.

The agreement between the patterns and between the time series (correlation higher then 0.98), obtained by the two methods, means that SVDA has brought no further

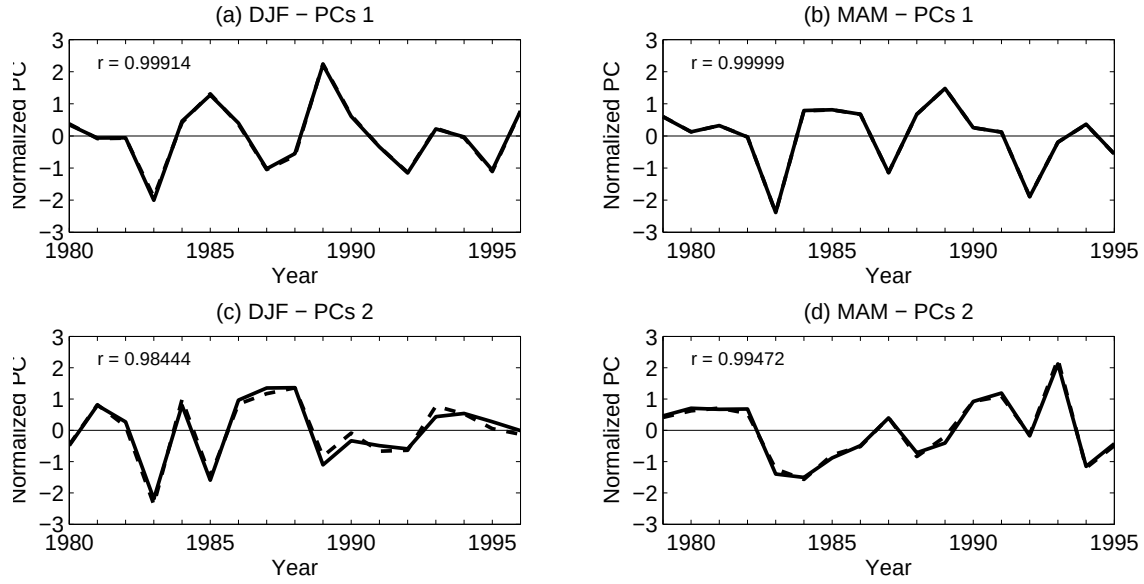

Figure 5.4: Normalised forced time series of winter and spring MUGCM MSLP in the EA region obtained by standard PCA of the ensemble mean (dashed lines) and by SVDA (solid lines): (a) DJF PC-1, (b) MAM PC-1, (c) DJF PC-2 and (d) MAM PC-2. The correlation between the two time series is printed on the upper left-hand corner of each panel.

information than the standard PCA, in this case. For an infinite ensemble size, the contribution from internal variability to the ensemble mean tends to zero (equation 5.1), and the forced modes of standard PCA become equivalent to those of SVDA [Ward and Navarra, 1997]. It seems, thus, that for our ensemble with 29 members the use of SVDA is redundant.

5.3 Forced response by Optimal Detection PCA

As we have already mentioned, equation 5.1 shows that if the internal variability had the form $\mathbf{\Sigma}_I = \sigma_I^2 \mathbf{I}$ (uncorrelated in space) then the expectation of the eigenvectors of the sample covariance of the ensemble mean, $\mathbf{S}_{EM}$, would have the direction of the eigenvectors of $\mathbf{\Sigma}_F$. Further, if $\sigma^2 = 1$ then the expected eigenvalues of $\mathbf{S}_{EM}$ would be equal the to eigenvalues of $\mathbf{\Sigma}_F$, that is, the variance represented by the PCs associated with the expected eigenvectors of $\mathbf{S}_{EM}$ would be the same as the variance of the true forced PCs.

Prewhitening operator

Following this idea, the so called *optimal detection* algorithm, proposed by Venzke et al. [1999][2], seeks to find an operator, $\mathcal{W}$ such that, when applied to the ensemble mean, $\mathcal{W}^{\mathrm{T}}\mathcal{X}_{EM}$, implies, using equation 5.1, that

$$\mathcal{E}\left[\mathcal{W}^{\mathrm{T}}\mathbf{S}_{EM}\mathcal{W}\right] = \mathcal{W}^{\mathrm{T}}\mathbf{\Sigma}_F\mathcal{W} + \mathbf{I}. \tag{5.21}$$

Such operator must then satisfy equation

$$\mathcal{W}^{\mathrm{T}}\mathbf{\Sigma}_I\mathcal{W} = K\mathbf{I}. \tag{5.22}$$

Since $\mathbf{\Sigma}_I$ is unknown, $\mathcal{W}$ has to be estimated from the sample covariance matrix of the internal variability, through the equation

$$\mathbf{W}^{\mathrm{T}}\mathbf{S}_I\mathbf{W} = K\mathbf{I}, \tag{5.23}$$

where $\mathbf{W}$ is an estimator of $\mathcal{W}$, called the *prewhitening* operator. Now, noting that

$$\mathbf{S}_I = \mathbf{E}_I\mathbf{\Lambda}_I\mathbf{E}_I^{\mathrm{T}}, \tag{5.24}$$

where the columns of $\mathbf{E}_I$ and the diagonal of $\mathbf{\Lambda}_I$ contain the eigenvectors and eigenvalues, respectively, of $\mathbf{S}_I$,[3] equation 5.23 may be written as

$$\mathbf{W}^{\mathrm{T}}\mathbf{E}_I\mathbf{\Lambda}_I\mathbf{E}_I^{\mathrm{T}}\mathbf{W} = K\mathbf{I},$$

equation from which it is easy to see, using $\mathbf{E}_I^{\mathrm{T}}\mathbf{E}_I = \mathbf{I}$, that the operator $\mathbf{W}$ that satisfies it is equal to $K^{1/2}\mathbf{E}_I\mathbf{\Lambda}_I^{-1/2}$.

Venzke et al. [1999] point out that, since the variance of the lowest ranked EOFs of internal variability is generally underestimated, we will only have

$$\mathbf{W}^{\mathrm{T}}\mathbf{\Sigma}_I\mathbf{W} \simeq K\mathbf{I}, \tag{5.25}$$

[2]Actually, this algorithm was previously introduced by Allen and Smith [1997] to make the Singular Spectrum Analysis able to separate signal from "coloured" noise, and is based on the fingerprinting algorithm of Hasselmann [1979].

[3]Note that since $\mathbf{S}_I = [P(K-1)]^{-1}\mathcal{X}_I\mathcal{X}_I^{\mathrm{T}}$, where $\mathcal{X}_I$ is a $(M \times KP)$ matrix, only the first $\min(M, KP)$ diagonal elements (eigenvalues) of Λ_I will be non-zero and thus only the first $\min(M, KP)$ columns (eigenvectors) of $\mathbf{E}_I$ are associated with non-zero eigenvalues.

if only the EOFs with realistic variance, the highest ranked EOFs, are retained in the computation of the prewhitening operator. Thus the prewhitening operator must be computed by

$$\mathbf{W} = K^{1/2}\mathbf{E}_I^{(Q)}(\mathbf{\Lambda}_I^{(Q)})^{-1/2}, \tag{5.26}$$

where $\mathbf{E}_I^{(Q)}$ is a $(M \times Q)$ matrix whose columns are the first Q columns of $\mathbf{E}_I$ (the Q highest ranked eigenvectors of $\mathbf{S}_I$) and $\mathbf{\Lambda}_I^{(Q)}$ is a $(Q \times Q)$ diagonal matrix whose elements are the first Q elements of $\mathbf{\Lambda}_I$ (the Q highest eigenvalues of $\mathbf{S}_I$). The prewhitening operator $\mathbf{W}$ is thus a $(M \times Q)$ matrix and it can be easily verified that satisfies equation 5.23 whatever the number of retained internal EOFs, Q.

The truncation level, Q, must then be as high as possible in order to achieve a better estimation of the forced EOFs (see latter discussion associated to equation 5.41) but low enough to avoid sampling problems. Chang et al. [2000] propose the following rule of thumb to determine the maximum number of well sampled EOFs: the minimum number of samples per EOF must be greater than five. Since, in our experiment, we have $17 \times 29 = 493$ samples, than the truncation level must not exceed 98 EOFs.

The criterion proposed by Venzke et al. [1999] is that Q should be the number which the cumulative average ratio between the variance of the projection of the ensemble mean onto the EOFs of internal variability and the variance of the PCs of internal variability multiplied by a factor of $1/K = 1/29$,

$$\frac{K}{Q}\sum_{r=1}^{Q}\frac{\mathbf{e}_{I,r}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{e}_{I,r}}{\mathbf{e}_{I,r}^{\mathrm{T}}\mathbf{S}_I\mathbf{e}_{I,r}} \tag{5.27}$$

becomes stable.

Figure 5.5 (a,b) presents the values of $\mathbf{e}_{I,r}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{e}_{I,r}$ and $(1/K)\mathbf{e}_{I,r}^{\mathrm{T}}\mathbf{S}_I\mathbf{e}_{I,r}$, and figure 5.5 (c,d) shows the cumulative average ratio of these variances, for $r = 1, ..., 100$, applied to the MUGCM MSLP winter and spring seasonal mean in the EA sector. Inspection of figure 5.5 (c,d) suggests the choice of $Q = 30$ for both seasons.

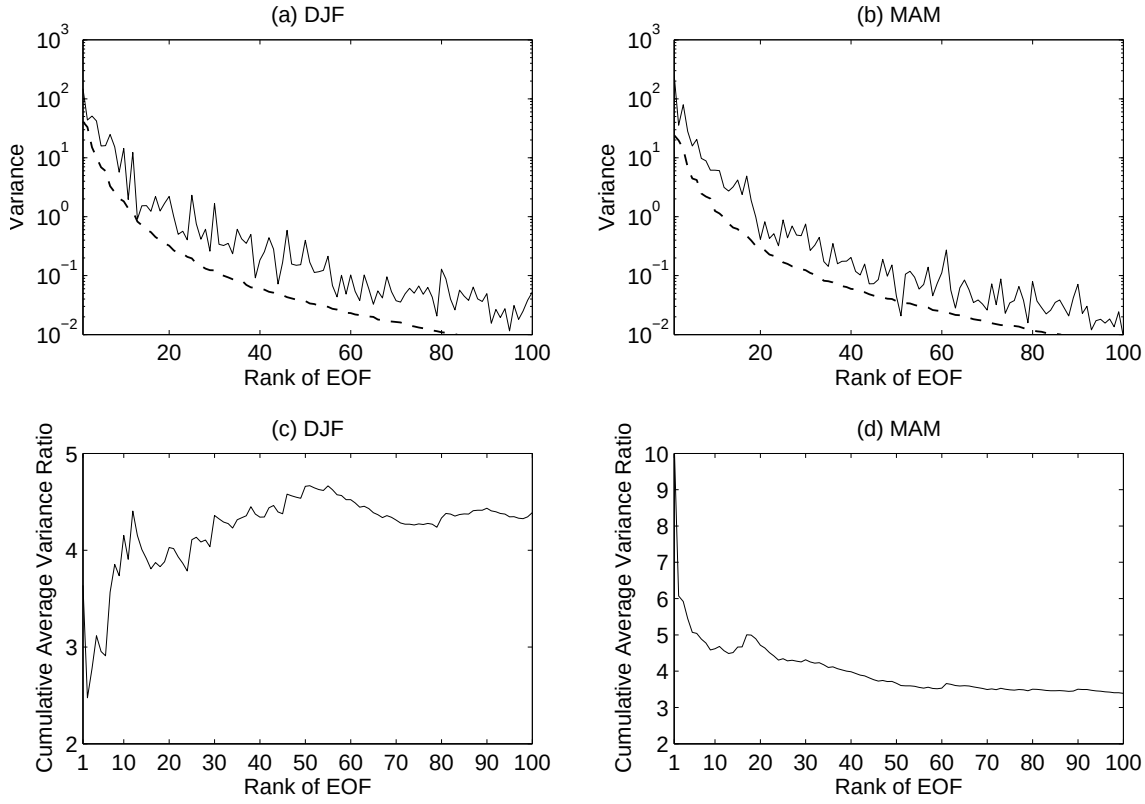

Figure 5.5: (a,b) Variance of the projection of the MUGCM ensemble mean of DJF and MAM seasonal means of MSLP onto the EOFs of internal variability (solid line) and variance of internal variability PCs multiplied by a factor of 1/29 (dashed line). (c,d) Cumulative average ratio of the variances plotted in (a,b).

Optimal Filters and Optimised Forced Principal Components

Once the value of Q has been chosen and the associated prewhitening operator determined, the next step is to apply the prewhitening transformation to the sample covariance matrix of the ensemble mean,

$$\mathbf{S}'_{EM} = \mathbf{W}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{W}, \tag{5.28}$$

and to perform a singular value decomposition of the $(Q \times Q)$ prewhitened sample covariance matrix,

$$\mathbf{S}'_{EM} = \mathbf{E}'_{EM}\mathbf{\Lambda}'_{EM}\mathbf{E}'^{\mathrm{T}}_{EM}, \tag{5.29}$$

to obtain the eigenvectors of $\mathbf{S}'_{EM}$ which are, as can be seen by equations 5.21 and 5.28, estimators of the eigenvectors of $\mathcal{W}^{\mathrm{T}}\mathbf{\Sigma}_F\mathcal{W}$. By the same reasoning, the eigenvalues of

$\mathbf{S'}_{EM}$,

$$\mathbf{\Lambda'}_{EM} = \mathbf{E'}^{\mathrm{T}}_{EM}\mathbf{S'}_{EM}\mathbf{E'}_{EM}, \tag{5.30}$$

are estimators of the eigenvalues of $\mathcal{W}^{\mathrm{T}}\mathbf{\Sigma_F}\mathcal{W}$.

Note that the covariance matrix of equation 5.28 is the sample covariance matrix of the prewhitened ensemble mean $\mathcal{X'}_{EM} = \mathbf{W}^{\mathrm{T}}\mathcal{X}_{EM}$, and the eigenvectors of $\mathbf{S'}_{EM}$, $\mathbf{E'}_{EM}$, are the EOFs of this transformed ensemble mean. This implies that the vectors $\mathbf{E'}_{EM}$ define the directions that maximise the variance of the transformed ensemble mean. Since the transformed internal variability, $\mathbf{W}^{\mathrm{T}}\mathbf{S}_i\mathbf{W}$, has equal variance in all directions (equation 5.23), than $\mathbf{E'}_{EM}$ are the set of vectors that maximise the ratio of the transformed signal variance to the transformed noise variance,

$$K\frac{\mathbf{E'}^{\mathrm{T}}_{EM}\mathbf{S'}_{EM}\mathbf{E'}_{EM}}{\mathbf{E'}^{\mathrm{T}}_{EM}\mathbf{W}^{\mathrm{T}}\mathbf{S}_I\mathbf{W}\mathbf{E'}_{EM}} = K\frac{\mathbf{\Lambda'}_{EM}}{K\mathbf{I}} = \mathbf{\Lambda'}_{EM}, \tag{5.31}$$

where we used equations 5.23 and 5.30. Using equation 5.28 and defining

$$\mathbf{F} = \mathbf{W}\mathbf{E'}_{EM}, \tag{5.32}$$

equation 5.31 becomes

$$K\frac{\mathbf{F}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{F}}{\mathbf{F}^{\mathrm{T}}\mathbf{S}_I\mathbf{F}} = \mathbf{\Lambda'}_{EM}, \tag{5.33}$$

which means that the set of vectors $\mathbf{F}$ maximise the ratio of the ensemble mean variance to internal variance. The columns of the $(M \times Q)$ matrix $\mathbf{F}$ are patterns with large weights where the signal-to-noise ratio is high and, thus, are used to project the ensemble mean to obtain the signal-to-noise maximised time-varying forced response:

$$\mathcal{Y}_{s/n} = \mathbf{F}^{\mathrm{T}}\mathcal{X}_{EM}. \tag{5.34}$$

For this reason, these patterns are called also *optimal filter* patterns and the forced time series, the lines of the $(Q \times P)$ matrix $\mathcal{Y}_{s/n}$, are called *optimal* forced principal components.

The covariance matrix of $\mathcal{Y}_{s/n}$ is

$$\begin{aligned} \mathbf{S}_{Y_{s/n}} &= \mathbf{F}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{F} && (5.35)\\ &= \mathbf{E'}^{\mathrm{T}}_{EM}\mathbf{S'}_{EM}\mathbf{E'}_{EM} && (5.36)\\ &= \mathbf{\Lambda'}_{EM}, && (5.37) \end{aligned}$$

where we used equations 5.34, 5.3, 5.32, 5.28 and 5.30. Note that equation 5.36 means that, since $\mathbf{E'}$ are the eigenvectors of $\mathbf{S'}_{EM}$, the *optimised* forced PCs, $\mathcal{Y}_{s/n}$, are the principal components of the prewhitened ensemble mean $\mathcal{X'}_{EM} = \mathbf{W}^{\mathrm{T}}\mathcal{X}_{EM}$. In fact, substituting equation 5.32 into equation 5.34 we get $\mathcal{Y}_{s/n} = \mathbf{E'}^{\mathrm{T}}_{EM}\mathbf{W}^{\mathrm{T}}\mathcal{X}_{EM} = \mathbf{E'}^{\mathrm{T}}_{EM}\mathcal{X'}_{EM}$, which is the projection of the prewhitened ensemble mean onto its EOFs.

Equations 5.33, 5.35 and 5.37 provide together an important property of the optimised principal components, $\mathcal{Y}_{s/n}$: in the directions of the optimal filter patterns, $\mathbf{F}$, the variances of $\mathcal{Y}_{s/n}$ (diagonal elements of $\mathbf{\Lambda'}_{EM}$) are equal to the corresponding signal-to-noise ratios. This is why they are also called the *signal-to-noise maximised* forced PCs.

Figure 5.6 (a,b) presents, for winter and spring, the projection of the ensemble mean onto the first optimal filter pattern, $\mathbf{f}^{\mathrm{T}}_{EM,1}\mathcal{X}_{EM}$, i.e., the first optimised forced PC, and the projections of the K individual ensemble members onto this filter pattern, $\mathbf{f}^{\mathrm{T}}_{EM,1}\mathcal{X}_{I,k}, \quad k = 1, ..., K = 29$. Figure 5.6 (c,d) shows analogous plots but for the second optimal filter pattern.

Figure 5.6 (a-d) shows that the spread of the projection of the individual ensemble members onto the first two leading optimal filter patterns has been successfully reduced, specially for the first mode, when compared with the spread of their projections onto the first two leading standard EOFs of the ensemble mean (Figure 5.2 (a-d)).

Figure 5.6 (e,f) shows the variance of the projection of the ensemble mean onto the 15 highest ranked optimal filter pattern (the 15 highest ranked optimised forced PCs),

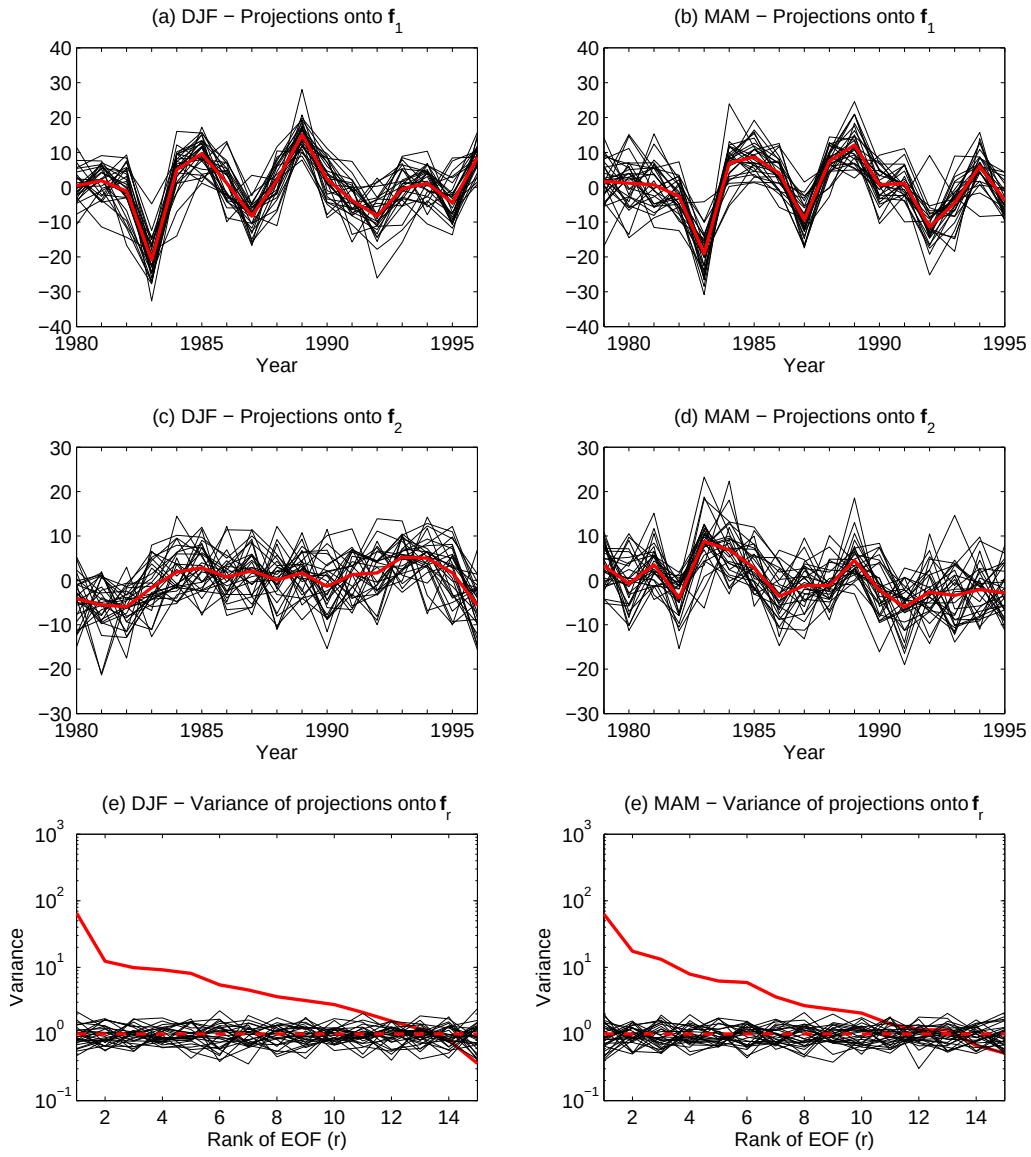

Figure 5.6: Results of optimal detection PCA of DJF and MAM seasonal mean MSLP ensemble mean in the EA region. (a,b) Projection of the ensemble mean (red line) and of the individual ensemble members (black lines) onto the first optimal filter pattern. (c,d) As in (a,b) but for the second optimal filter pattern. (e,f) Variance of the projections of the ensemble mean (red solid line), of the individual ensemble members (black lines) and their mean (red dashed line) onto the first 15 optimal filter patterns.

$\mathbf{f}_{EM,r}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{f}_{EM,r} = \lambda'_{EM,r}$, where $r = 1, ..., 15$. Note again that, since $(1/K)\mathbf{f}_{EM,r}^{\mathrm{T}}\mathbf{S}_I\mathbf{f}_{EM,r} = 1$, these variances are actually, signal-to-noise ratios.

Also shown figure 5.2 (e,f) is the variance of the projections of the individual ensemble members onto these optimal filter patterns, $\mathbf{f}_{EM,r}^{\mathrm{T}}\mathbf{S}_{I,k}\mathbf{f}_{EM,r}$, with $r = 1, ..., 15$, and the average of these 29 variances, $(1/K)\sum_{k=1}^{K}\mathbf{f}_{EM,r}^{\mathrm{T}}\mathbf{S}_{I,k}\mathbf{f}_{EM,r} = (1/K)\mathbf{f}_{EM,r}^{\mathrm{T}}\mathbf{S}_{I}\mathbf{f}_{EM,r} = 1$, which, according to equation 5.1 is an estimation of the variance, in the directions of $\mathbf{f}_{EM,r}$, that we would obtain in the absence of any true forced response.

Analogously to what was done in section 5.1 to test if the ensemble mean PCs represented statistically significant forced responses, the significance of the response $y_{s/n,r} = \mathbf{f}_{EM,r}^{\mathrm{T}}\mathcal{X}_{EM}$ (the rth line of $\mathcal{Y}_{s/n}$) is tested with the hypothesis test

$$H_0 : \mathbf{f}_{EM,r}^{\mathrm{T}}\ \mathbf{\Sigma}_F\ \mathbf{f}_{EM,r} = 0 \quad vs \quad H_1 : \mathbf{f}_{EM,r}^{\mathrm{T}}\ \mathbf{\Sigma}_F\ \mathbf{f}_{EM,r} > 0, \tag{5.38}$$

using the test statistic

$$F = K\frac{\mathbf{f}_{EM,r}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{f}_{EM,r}}{\mathbf{f}_{EM,r}^{\mathrm{T}}\mathbf{S}_{I}\mathbf{f}_{EM,r}} \sim \mathcal{F}_{P-1,P(K-1)}, \tag{5.39}$$

under H_0, and the test criterion for rejecting H_o with a significance level of α,

$$F > \mathcal{F}_{P-1,P(K-1),\alpha}. \tag{5.40}$$

Table 5.2 presents the results of the test 5.38 applied to the first and second PCs obtained by the optimal detection PCA of the MUGCM winter and spring MSLP in the EA region. Since all values of the F statistic exceed the cutoff value $\mathcal{F}_{16,17\times29,0.05} = 1.66$, the null hypothesis is rejected, meaning that the forced response in both patterns and in both seasons significantly exceed zero at the 5% significance level. Using the definition of Venzke et al. [1999] we say that all responses are *detectable*, as also verified for the responses obtained by the standard PCA (table 5.1).

The values of the F/K statistic (table 5.2, last column) indicates that the dominant forced response is *consistent* in both seasons ($F/K > \mathcal{F}_{16,17\times29,0.05} = 1.66$). This fact represents an important improvement achieved by the optimal detection algorithm. Remember that the dominant forced response obtained by the standard PCA was detectable but not consistent (table 5.1). The responses associated with the second optimal filter have values of F/K four (in winter) and three (in spring) times greater

Table 5.2: Variances of the projections of the ensemble mean and of individual ensemble members onto the first two leading optimal filter patterns of the MUGCM MSLP in DJF and MAM. Values of the F statistic greater than 1.66 (bold) indicate a *detectable* response, that is, a forced response that significantly exceed zero at the 0.05 significance level. Values of F/K greater than 1.66 (bold) indicates a *consistent* response.

r	$\mathbf{f}_{EM,r}^{\mathrm{T}}\mathbf{S}_{EM}\mathbf{f}_{EM,r}$	$\mathbf{f}_{EM,r}^{\mathrm{T}}\mathbf{S}_{I}\mathbf{f}_{EM,r}$	F	F/K
DJF Optimal Filter 1	65.26	29	**65.26**	**2.25**
MAM Optimal Filter 1	62.99	29	**62.99**	**2.17**
DJF Optimal Filter 2	12.35	29	**12.35**	0.43
MAM Optimal Filter 2	17.43	29	**17.43**	0.06

than the corresponding values of the responses obtained by the conventional method. However, these values are not large enough to represent consistent responses.

Optimised Patterns of the Forced Response

So far, we determined the two time-varying forced responses of the MUGCM seasonal MSLP with higher signal-to-noise ratios and, using an ANOVA test, concluded that both are *detectable* in winter and spring, being the first also *consistent* in both seasons. We saw that these responses are the two highest ranked principal components of the prewhitened ensemble mean or, equivalently, are the projections of the MUGCM MSLP ensemble mean onto the two dominant optimal filter patterns. Although the optimal filter patterns, $\mathbf{F}$, provide a set of vectors to extract signal-to-noise maximised time-varying forced responses, they are not estimators of the eigenvectors of $\mathbf{\Sigma}_F$. Note, for instance, that they do not form an orthonormal basis in the Euclidean metric, since $\mathbf{F}\mathbf{F}^{\mathrm{T}} = \mathbf{W}\mathbf{W}^{\mathrm{T}} = K(\mathbf{S}_I^{(Q)})^{-1}$.

The sample spatial forced patterns may be obtained as follows. We have already seen that the eigenvectors and eigenvalues of $\mathbf{S'}_{EM}$ are estimators of the eigenvectors and eigenvalues of $\mathcal{W}^{\mathrm{T}}\mathbf{\Sigma}_F\mathcal{W}$. Thus, assuming that we have sufficient data that $\mathbf{S'}_{EM} \simeq$

$\mathcal{E}[\mathbf{S'}_{EM}]$ and $\mathbf{W} \simeq \mathcal{W}$, we have

$$\mathbf{W}^{\mathrm{T}}\mathbf{\Sigma}_F\mathbf{W} = \mathbf{E'}_{EM}\mathbf{\Lambda'}_{EM}\mathbf{E'}^{\mathrm{T}}_{EM}.$$

Multiplying on the left by $(\mathbf{W}^{\mathrm{T}})^{-1}$ and on the right by $\mathbf{W}^{-1}$ and rearranging we get

$$(\mathbf{W}\mathbf{W}^{-1})^{\mathrm{T}}\mathbf{\Sigma}_F(\mathbf{W}\mathbf{W}^{-1}) = [(\mathbf{W}^{-1})^{\mathrm{T}}\mathbf{E'}_{EM}]\mathbf{\Lambda'}_{EM}[(\mathbf{W}^{-1})^{\mathrm{T}}\mathbf{E'}_{EM}]^{\mathrm{T}}, \tag{5.41}$$

where

$$\mathbf{W}^{-1} = K^{-1/2}(\mathbf{\Lambda}_I^{(Q)})^{1/2}(\mathbf{E}_I^{(Q)})^{\mathrm{T}} \tag{5.42}$$

is the pseudo-inverse of $\mathbf{W}$.

Since $\mathbf{W}\mathbf{W}^{-1} = \mathbf{E}_I^{(Q)}(\mathbf{E}_I^{(Q)})^{\mathrm{T}}$, if all the EOFs of internal variability were used to construct the prewhitening operator ($Q = M$), than, by equation 5.24, the vectors $\mathbf{E}_I^{(M)} = \mathbf{E}_I$ would be an orthonormal basis and, consequently, the left hand side of equation 5.41 would be equal to $\mathbf{\Sigma}_F$. In this situation, equation 5.41 would imply that the vectors $(\mathbf{W}^{-1})^{\mathrm{T}}\mathbf{E'}_{EM}$ would be (estimators of) the eigenvectors of $\mathbf{\Sigma}_F$. This is the reason why we define the forced spatial patterns or *optimised* forced EOFs as the columns of the $M \times Q$ matrix

$$\mathbf{E}_F = (\mathbf{W}^{-1})^{\mathrm{T}}\mathbf{E'}_{EM}. \tag{5.43}$$

Since we must truncate the EOFs of internal variability in the construction of the prewhitening operator, for the reasons exposed above, the *optimised* forced EOFs, $\mathbf{E}_F$ are estimators of the eigenvectors of $\mathbf{E}_I^{(Q)}(\mathbf{E}_I^{(Q)})^{\mathrm{T}}\mathbf{\Sigma}_F\mathbf{E}_I^{(Q)}(\mathbf{E}_I^{(Q)})^{\mathrm{T}}$. That is, they provide an estimate of the subspace in which the true eigenvectors lie but do not provide an estimate of the true forced eigenvectors themselves, since $\mathbf{E}_F$ is not an orthonormal basis in the Euclidean metric [Venzke et al., 1999]: $\mathbf{E}_F\mathbf{E}_F^{\mathrm{T}} = (\mathbf{W}^{-1})^{\mathrm{T}}\mathbf{W}^{-1} = K^{-1}\mathbf{S}_I^{(Q)}$.

From equation 5.29, since $\mathbf{E'}_{EM}$ is orthonormal and $\mathbf{\Lambda'}_{EM}$ is invertible, we have $\mathbf{E'}_{EM} = \mathbf{S'}_{EM}\mathbf{E'}_{EM}\mathbf{\Lambda'}^{-1}_{EM}$. Substituting this equation in equation 5.45 using equations 5.28, 5.3, 5.32 and 5.34 we obtain

$$\begin{aligned}\mathbf{E}_F &= \frac{1}{P-1}\mathbf{W}\mathbf{W}^{-1}\mathcal{X}_{EM}\mathcal{Y}^{\mathrm{T}}_{s/m}\mathbf{\Lambda'}^{-1} \\ &= \frac{1}{P-1}\mathbf{E}_I^{(Q)}(\mathbf{E}_I^{(Q)})^{\mathrm{T}}\mathcal{X}_{EM}\mathcal{Y}^{\mathrm{T}}_{s/m}\mathbf{\Lambda'}^{-1},\end{aligned} \tag{5.44}$$

which, by analogy to equation A.10 is equivalent to regressing the filtered ensemble-mean $\mathbf{W}\mathbf{W}^{-1}\mathcal{X}_{EM} = \mathbf{E}_I^{(Q)}(\mathbf{E}_I^{(Q)})^{\mathrm{T}}\mathcal{X}_{EM}$ onto the optimised PCs.

Because of equation 5.44 Venzke et al. [1999] propose the following approximation to determine the forced patterns

$$\mathbf{E}_F \simeq \frac{1}{P-1}\mathcal{X}_{EM}\mathcal{Y}_{s/n}^{\mathrm{T}}\mathbf{\Lambda}'^{-1} \tag{5.45}$$

which, again by analogy to equation A.10, is equivalent to regressing the ensemble mean onto the optimised PCs. This approximation was also used by Chang et al. [2000] to determine the SST-forced patterns of wind stress and surface heat flux in the Tropical Atlantic.

As already mentioned, the vectors $\mathbf{E}_F$, given by the accurate computation (equation 5.43) or by the approximate computation (equation 5.45), are not orthogonal. Allen and Smith [1997] show how to build an orthonormal basis of signal-to-noise maximising vectors from the leading signal-to-noise maximising components of $\mathbf{E}'$. First we reconstruct a filtered transformed covariance matrix from $\mathbf{S}_{EM}$ using the $R < Q$ highest ranked signal-to-noise components of $\mathbf{E}'$

$$\mathbf{S}'^{(R)}_{EM} = \mathbf{E}'^{(R)}\mathbf{\Lambda}'^{(R)}(\mathbf{E}'^{(R)})^{\mathrm{T}}, \tag{5.46}$$

where $\mathbf{E}'^{(R)}$ is a $(Q \times R)$ matrix whose columns are the first R columns of $\mathbf{E}'$ and $\mathbf{\Lambda}_I^{(R)}$ is a $(R \times R)$ diagonal matrix whose elements are the first R elements of $\mathbf{\Lambda}_I$. Then, we back-transform to the coordinates of the original state space

$$\mathbf{S}^{(R)}_{EM} = (\mathbf{W}^{-1})^{\mathrm{T}}\mathbf{S}'^{(R)}_{EM}\mathbf{W}^{-1}, \tag{5.47}$$

and re-diagonalise

$$\mathbf{S}^{(R)}_{EM} = \mathbf{E_f}^{(R)}\mathbf{\Lambda_f}^{(R)}(\mathbf{E_f}^{(R)})^{\mathrm{T}}, \tag{5.48}$$

where $\mathbf{E_f}^{(R)}$ and $\mathbf{\Lambda_f}^{(R)}$ are $(M \times M)$ matrices. The first R column vectors of $\mathbf{E_f}^{(R)}$ are the R highest ranked signal-to-noise maximising EOFs and are estimators of the R true dominant forced EOFs.

We have computed the first and second highest ranked optimised (or signal-to-noise maximising) forced EOFs of the MUGCM MSLP seasonal means in winter and spring in the EA region, using the three methods presented above: (i) the accurate computation of the non-orthonormal vectors of Venzke et al. [1999] (equation 5.43), (ii) the approximate computation of the non-orthonormal vectors of Venzke et al. [1999] (equation 5.45), and (iii) the computation of the orthonormal vectors of Allen and Smith [1997] (equation 5.48) with $R = 2$. The patterns obtained by the three methods for both EOFs and seasons present negligible differences. Because of this, although the patterns shown in figure 5.7 were computed using the Allen and Smith [1997] procedure, we propose that they should be regarded as computed by equation 5.45 since its interpretation is easier: patterns obtained by regressing the ensemble mean MSLP onto the optimised PCs.

Figure 5.7 (a,b) shows the leading signal-to-noise maximising EOFs estimating the leading modes of true forced response of the MUGCM MSLP in winter and spring. Analogous plots for the second leading modes are presented in Figure 5.7 (c,d).

The dominant modes, in winter and spring, have a meridional dipole structure that projects on the dipole of the NAO. These patterns are very similar to the patterns obtained by standard PCA of the ensemble mean, figure 5.1 (a,b). The second EOFs of the ensemble mean (figure 5.1 (c,d)) also presents a dipole design but the axis of the dipole has a southwest-northeast direction. The second optimised forced EOF in spring also presents some resemblance with the corresponding ensemble mean EOF, but with a stronger southern pole. In winter, the dipole structure found in EOF-2 of the ensemble mean seem to disappear in the optimised EOF, whose pattern is dominated by a centre of action located over the western Atlantic between 40°N and 60°N.

5.4 Comparison with observations

In this section we compare the results obtained by the two methods exposed in the last two sections with the observations. Before doing this, let us first compare the results of the two methodologies with each other.

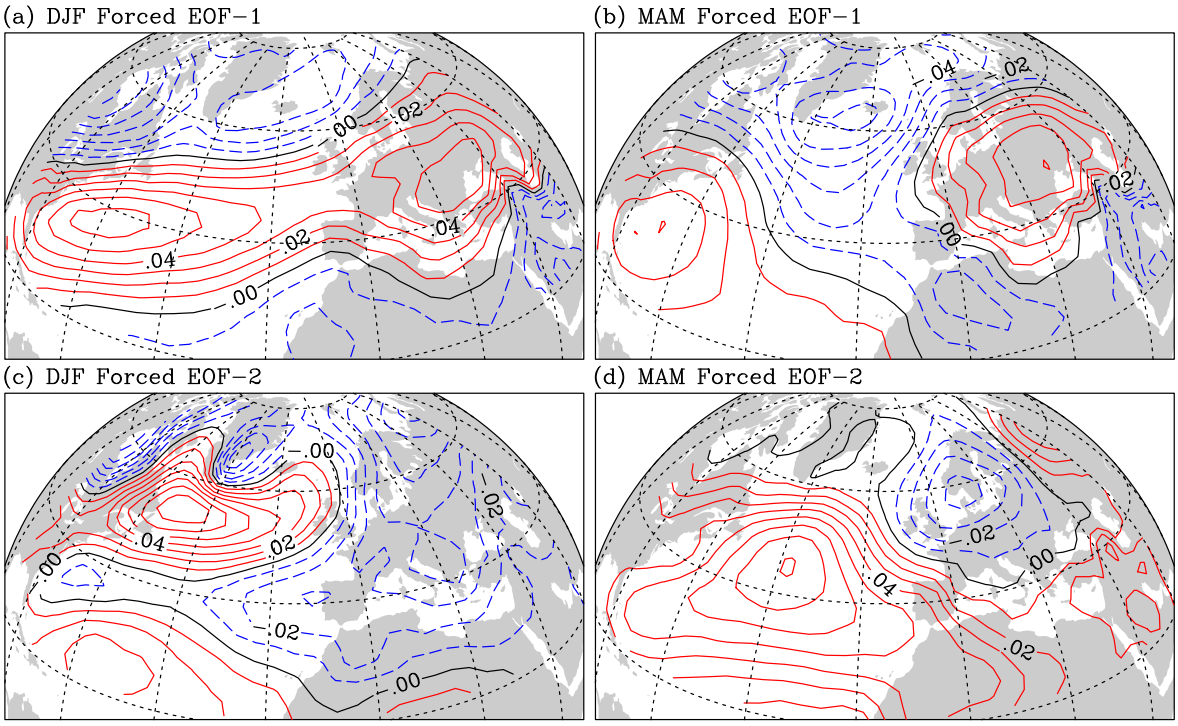

Figure 5.7: First two leading optimal EOFs of MUGCM MSLP seasonal mean anomalies in (a,c) DJF and (b,d) MAM. Negative, zero and positive loadings plotted with dashed blue, solid black and solid red lines. Grid lines represented every 20°.

As already noticed in the last section, the winter and spring MSLP forced patterns obtained by standard PCA of the ensemble mean (figure 5.1) and by optimal detection PCA (figure 5.7) are noticeable different only for the second EOF. The agreement between the first standard and optimal EOFs was somehow expected given the relatively large number (29) of integrations performed with the MUGCM.

Figure 5.8 compares the time-varying forced responses, normalised to unit variance, obtained by standard PCA in winter and spring (red lines of figure 5.2 (a-d)) and those obtained by optimal detection (red lines of figure 5.6 (a-d)). The leading responses obtained by the two methods are almost equal, in both seasons, with correlations higher that 0.9. By contrast, the responses associated with the second mode differ substantially. The correlation in winter is 0.1 and in spring is -0.49.

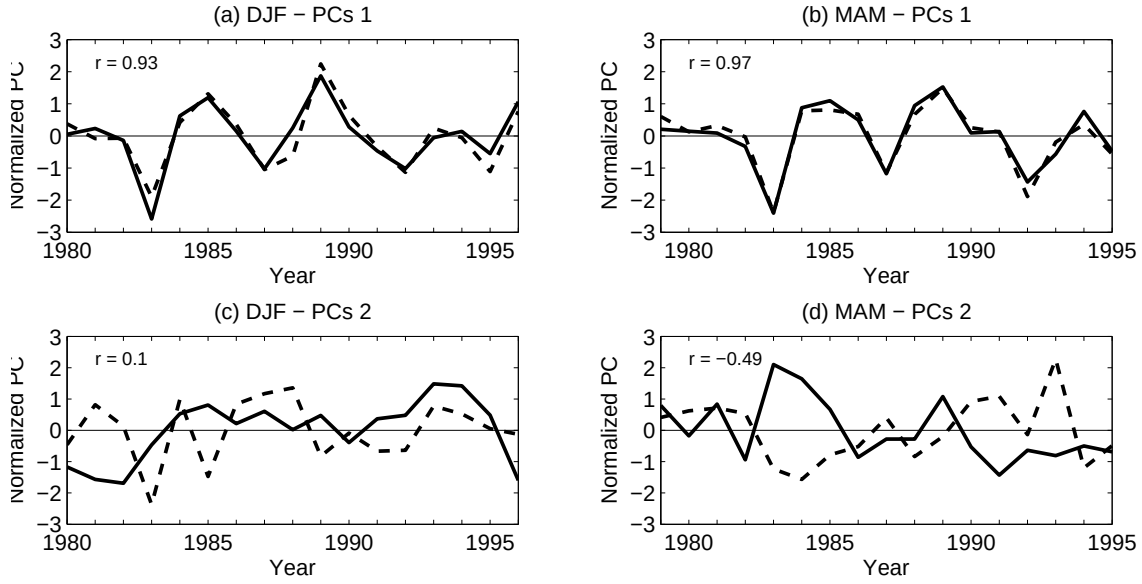

Figure 5.8: Normalised forced principal components of winter and spring MUGCM MSLP in the EA region obtained by standard PCA of the ensemble mean (dashed lines) and by optimal detection PCA (solid lines): (a) DJF PC-1, (b) MAM PC-1, (c) DJF PC-2 and (d) MAM PC-2. The correlation between the two time series is printed on the upper left-hand corner of each panel.

In summary, the results (spatial patterns and associated time series), for the MUGCM ensemble, obtained by conventional PCA of the ensemble mean and by the optimal detection algorithm are noticeable different only for the second mode.

In order to see if the determined time series of the forced response have any resemblance with fluctuations of the observed MSLP in the EA region, we project the NCEP seasonal mean MSLP anomalies, in winter and spring, onto the first two EOFs of the ensemble mean and onto the first two optimal filter patterns obtained by the optimal detection procedure. Before computing the projections, the NCEP MSLP was interpolated to the MUGCM grid in the EA region, and the time series were truncated to the years simulated by the MUGCM, that is, from DJF 1980 to DJF 1996 and from MAM 1979 to MAM 1995. The time series obtained by these projections are presented, along with the associated MUGCM time series, in figure 5.9.

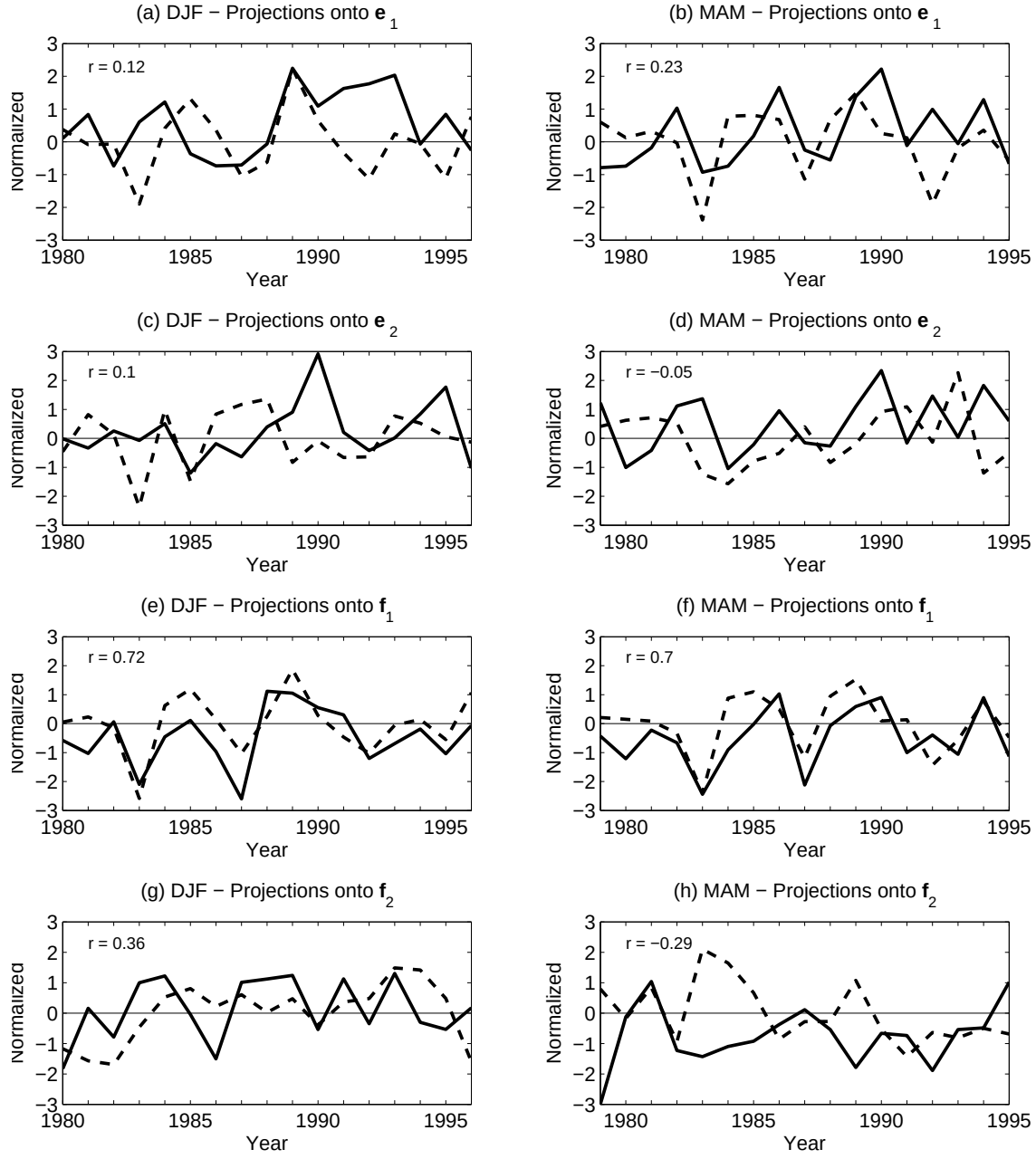

Figure 5.9: Normalised projections of MUGCM MSLP ensemble mean (dashed lines) and of NCEP MSLP (solid lines) onto: (a) DJF ensemble mean EOF-1, (b) MAM ensemble mean EOF-1, (c) DJF ensemble mean EOF-2 and (d) MAM ensemble mean EOF-2. (e) DJF optimal filter 1, (f) MAM optimal filter, (g) DJF optimal filter 2 and (h) MAM optimal filter 2. The correlation between the two time series is printed on the upper left-hand corner of each panel.

The four upper panels of figure 5.9 reveals that the principal components of the ensemble mean have little agreement with the projections of the observed data onto the associated forced patterns. The first optimal principal component, however, correlates very well (around 0.7 in both seasons) with the projection of the observed data onto the associated optimal filter pattern. The second optimal principal component in winter correlates better with the observations than its counterpart obtained by standard PCA, 0.4 and 0.1 respectively, but the same cannot be said for the spring season, when the modelled and observed time series appear out of phase.

In summary, the comparison with observations suggest that we may rely only on the first and second optimal response in winter and on the first optimal response in spring.

5.5 Localisation of the forcing oceanic regions

The analyses performed so far to study the forced response of the MSLP in the EA region to fluctuations in the SST-SIC field (the ANOVA in chapter 4 and the conventional and optimal detection PCA in sections 5.1 and 5.3) did not explicitly use this field. Consequently, the forced signals obtained by these analyses are known to be forced by the variability of the SST and SIC fields but the specific regions of these fields whose variability is responsible for the forced signals in the EA region are unknown. The purpose of this section is, therfore, to localise these regions and associated SST anomaly patterns responsible for the forced patterns obtained in the last sections.

The patterns of SST anomalies associated with (or responsible for) the time series of the forced response are determined by linear regressing the time series of seasonal SST anomalies at each gird point of the globe onto these time series, after being normalised to unit variance. The percentage of the SST anomaly variability explained by the regression is given by the squared correlation between the time series involved in the regression.

Figure 5.10 presents the patterns of SST anomalies associated with the first and second principal components of the MUGCM MSLP ensemble mean, in winter and spring. Figure 5.11 shows analogue plots, but for the optimal detected principal components. In figures 5.10 and 5.11 the regression coefficients are represented by contour lines and the percentage of the SST anomaly variability explained by the regression is represented by shaded contours.

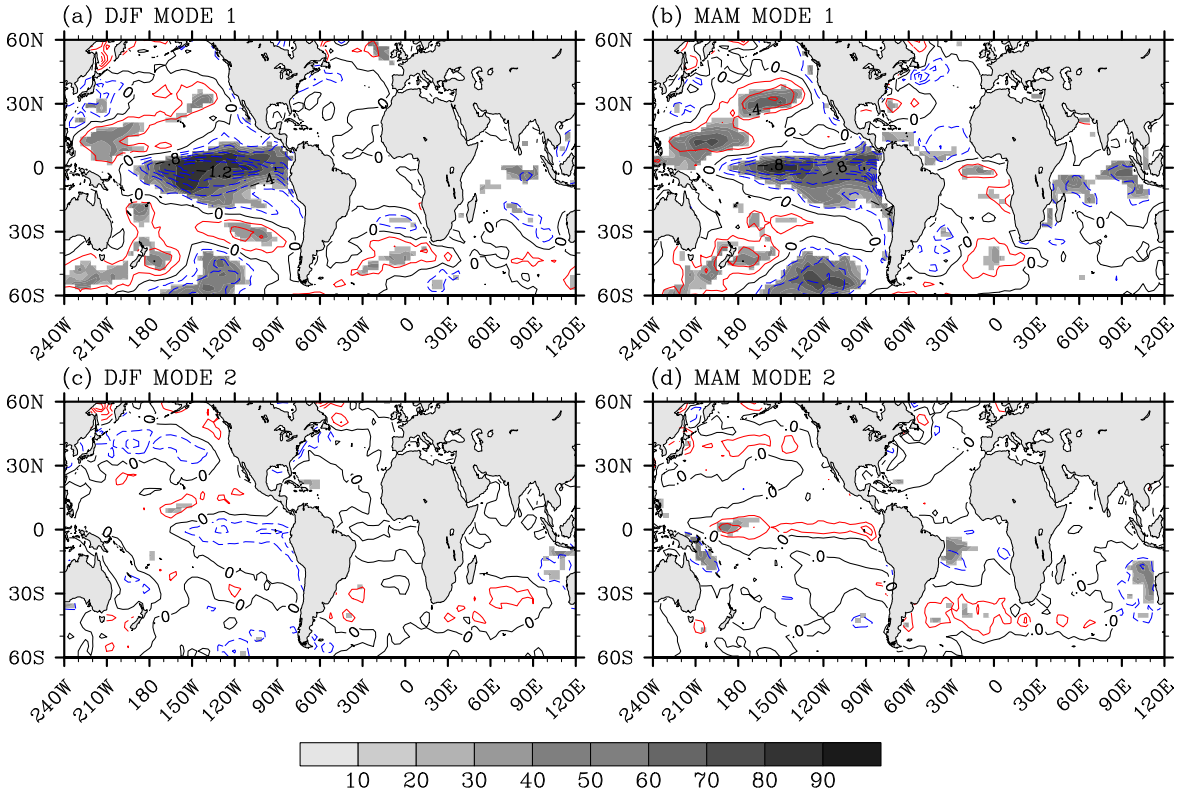

Figure 5.10: Regression coefficients (contour lines) and percentage of total variance explained (shaded contours) by the regression of SST anomalies onto the normalised standard forced (a) DJF PC-1, (b) MAM PC-1, (c) DJF PC-2, and (d) MAM PC-2, of the Euro-Atlantic MUGCM MSLP anomalies. White areas are where the regression is not significant at the 0.07 significance level. Contour interval is 0.2 K per standard deviation of the PC. Dashed blue, solid black and solid red lines for negative, zero and positive coefficients, respectively.

In winter, the first response obtained by both methods is mainly associated with SST anomalies in the equatorial Pacific (figures 5.10 (a) and 5.11 (a)). Large areas of

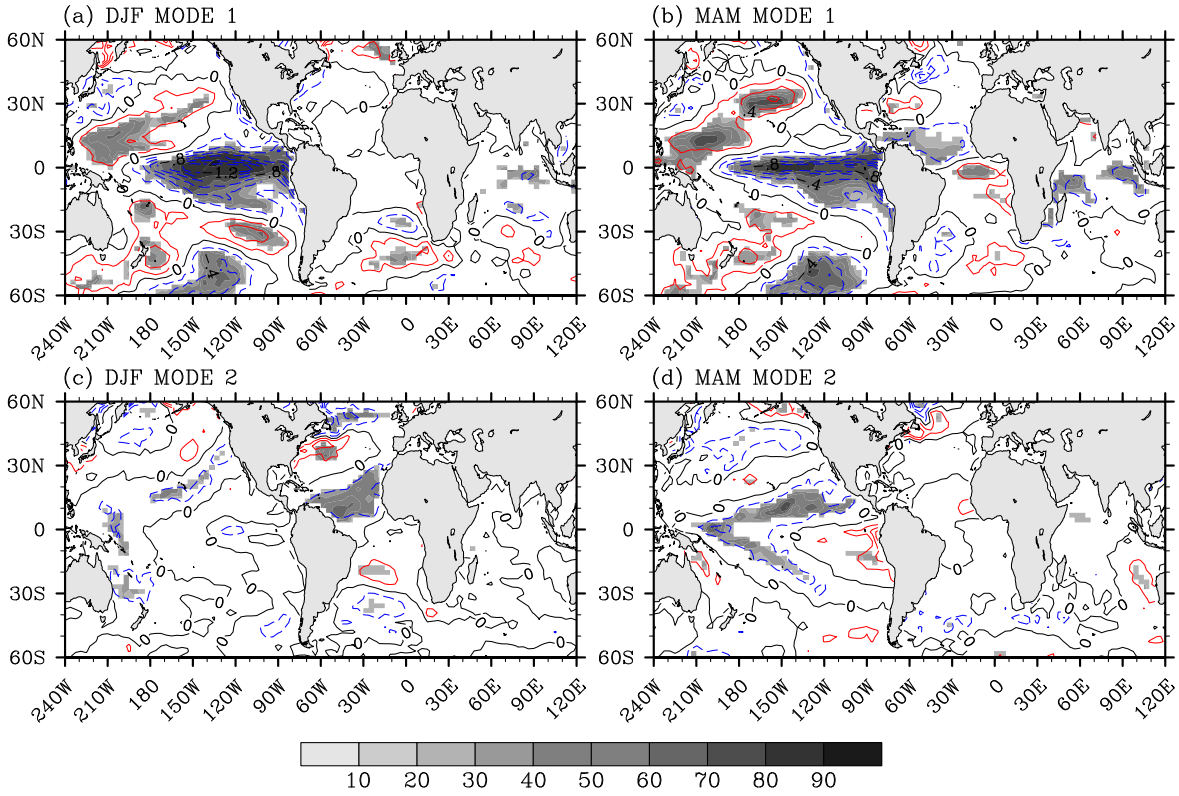

Figure 5.11: As figure 5.10 but for the regression of SST anomalies onto the normalised optimal forced PCs.

significant regression coefficients are also found in the extratropical Pacific.

In spring, the first response of the two methods (figures 5.10 (b) and 5.11 (b)) are, as in winter, associated with SST anomalies in the equatorial Pacific but SST significant anomalies in the tropical Atlantic and tropical Indian oceans are also found. In the first optimal forced mode, the SST regression pattern is similar to the one found for the standard forced mode, but the area of significant regression coefficients in the tropical Atlantic is much larger (figure e5.11 (b)).

The winter second standard response is not statistically significantly associated with any region in the globe (figure 5.10 (c)) but the optimal response is found to be associated with SST anomalies in the north Atlantic (figure 5.11 (b)).

The second standard mode in spring is associated with SST anomalies in the western tropical Atlantic off the Nordeste Brazil coast (figure 5.10 (d)) whereas the optimal mode is associated with SST anomalies in the tropical Pacific outside the equatorial belt (figure 5.11 (d)).

5.6 Modes of SST forcing

Three statistical significant regression patterns outstand from the figures 5.10 and 5.11:

- the regression pattern in the equatorial Pacific, found by the regression of the SST anomaly field onto standard PC-1 and optimal PC-1, in winter and spring, resembling the SST anomaly pattern of the El Niño-Southern Oscillation (ENSO);
- the regression pattern in the tropical Atlantic, found by the regression of the SST anomaly field onto the optimal PC-1 in spring, resembling the pattern of the Atlantic Interhemispheric SST anomaly Gradient mode (AISG).
- the regression pattern in the north Atlantic basin, found by the regression of the SST anomaly field onto the optimal PC-2 in winter, resembling the SST anomaly pattern of the North Atlantic SST Tripole mode (SST Tripole).

In order to test if these three SST modes force indeed the first and/or second detected forced modes of MSLP in the EA, as suggested by the item list above, we began by expressing each SST mode by a spatial pattern of variability and its associated time series (often called the mode index). This was done by performing a PCA of the SST anomalies in the oceanic regions where the SST mode occur. Then, for each item of the list above, the forced MSLP PC was compared with the time series of the forcing SST mode. We also compared the pattern of the regression of SST anomalies onto the forced MSLP PC with the pattern of the forcing SST mode.

Note that computing the SST indices using PCA, instead of area-weighted SST indices, provides also the spatial patterns of variability, associated with the indices, that can be compared with the regression patterns.

5.6.1 ENSO SST forcing

The ENSO SST mode is defined here as the first mode of a PCA of the SST anomalies in the equatorial Pacific belt (20°S-20°N, 120°E-70°W)[4]. This mode, representing 65.5% (55.8%) of the winter (spring) SST variability in the equatorial Pacific, is shown in figure 5.12. The pattern is plotted as coloured shadings in figure 5.12 (a,b) and its normalised PC (representing the ENSO index) is plotted in figure 5.12 (c,d) as red dashed lines.

Figure 5.12 (a,b) also shows the winter and spring patterns in the tropical Pacific of the regression of SST anomalies onto the normalised optimal forced PC-1 of the EA MSLP (the same of figure 5.11 (a,c)), as contour lines overlaid on the SST ENSO pattern. Figure 5.12 (c,d) shows, along with the ENSO index, the MSLP normalised optimal forced PC-1 (plotted as a continuous blue line, with opposite sign for easier comparison).

The agreement between the patterns and between the time series of figure 5.12 is remarkable. The correlation between the time series is -0.943 in winter and -0.944 in spring, with a significance level not higher than 0.001. This results shows that the SST anomalies associated with ENSO are responsible for the leading optimal forced PC of MUGCM MSLP in the EA region, which means that ENSO SST anomalies are the major SST-forcing of the EA MUGCM MSLP.

Now that the leading forced mode of MSLP anomalies in the EA region has been shown to be forced by ENSO SST anomalies, it is useful to better understand the relationship between the time series associated to this mode and the MSLP anomalies in the EA region. As mentioned in section 5.3, the optimal forced patterns (figure 5.7)

[4]This region is similar to the region (20°S-20°N, 120°E-60°W) used for the *Tropical Pacific SST EOF* defined by the Climate Analysis Branch (CAB) of the Physical Sciences Division (PSD), formerly the Climate Diagnostics Division (CDC), of the Earth System Research Laboratory (ESRL) of the american National Oceanic and Atmospheric Administration (NOAA) - www.cdc.noaa.gov/ClimateIndices/List/#Tropicaleof.

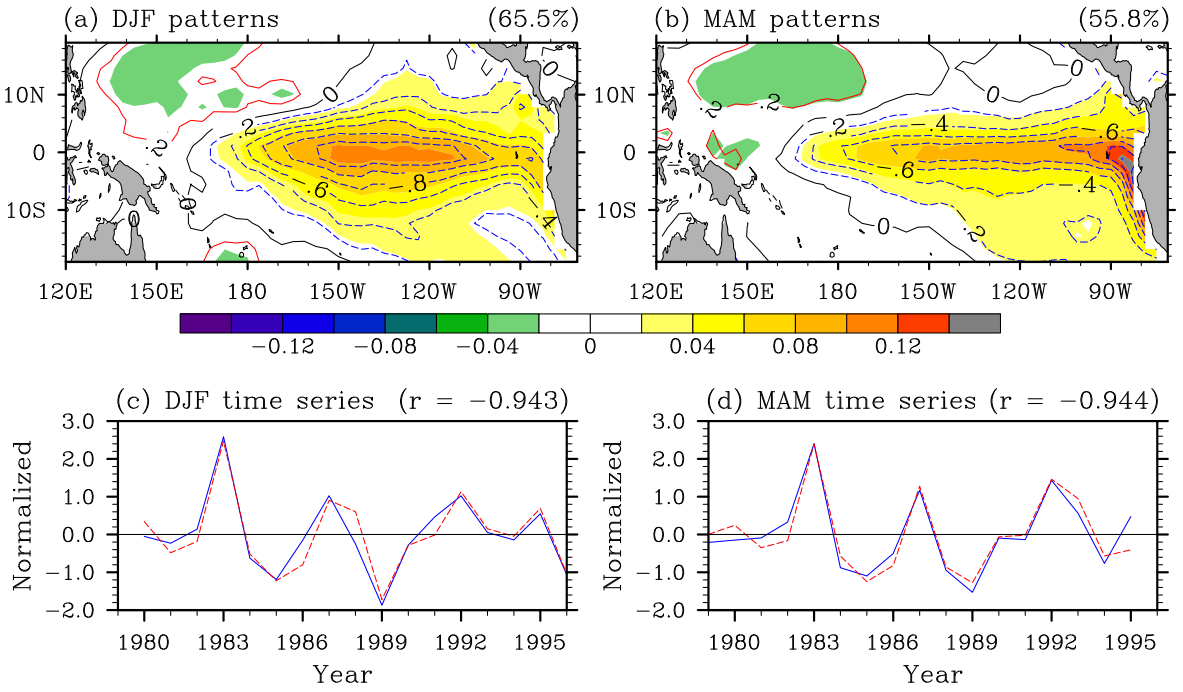

Figure 5.12: (a) Regression coefficients of SST anomalies onto the EA MSLP normalised optimal forced PC-1 (contour lines; contour interval is 0.2 K per standard deviation of the PC), dashed lines for negative coefficients) overlaid on the SST EOF-1 (coloured shading; representing the SST ENSO pattern), in DJF. (b) Same as (b) but for MAM. (c) EA MSLP normalised optimal forced PC-1 (blue line, plotted with opposite sign) and tropical Pacific normalised SST PC-1 (red dashed line; representing the ENSO index), in DJF. (d) Same as (c) but for MAM. Correlations significantly different from zero at a significance level not higher than 0.001. Values printed on the top right-hand corner of the panels represent the percentage of total variance explained by the ENSO mode in (a,b) and the correlations between the time series in (c,d).

may be interpreted as the regression patterns of the ensemble mean MSLP anomalies onto the corresponding optimised PCs. Figure 5.13 shows, for winter and spring, the pattern (contour isolines) obtained by linear regressing the ensemble mean of MSLP anomalies at each grid point of the EA region onto the *normalised to unit variance* optimal forced PC-1, and also the pattern (shaded contours) of the fraction of total MSLP variance at each grid point explained by the regression, given by the squared

correlation between the time series[5].

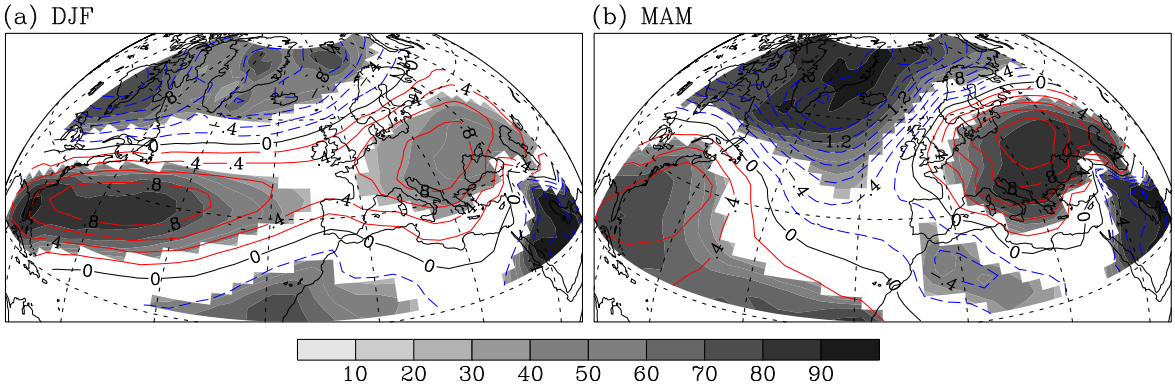

Figure 5.13: Regression coefficients (contour lines) and percentage of total variance explained (shaded contours) by the regression of the MUGCM ensemble mean of MSLP anomalies onto the normalised optimal forced (a) DJF PC-1 and (b) MAM PC-2 of the Euro-Atlantic MUGCM MSLP anomalies. White areas are where the regression is not significant at the 0.05 significance level. Contour interval is 0.2 mb per standard deviation of the PC. Dashed blue, solid black and solid red lines for negative, zero and positive coefficients, respectively. Grid lines represented every 20°.

From figures 5.12 and 5.13 it can be seen that, in winter and spring, positive SST anomalies in the tropical Pacific during a warm ENSO episode induces a pattern of MSLP anomalies in the EA region that projects on the negative phase of the leading EOF of the total variability MUGCM MSLP anomalies in the EA region (section 3.4.2, figures 3.22 (a) and 3.23 (a)) that were recognised to be similar to the observed NAO pattern (section 3.4.2, figures 3.22 (b) and 3.23 (b)), specially in winter. In both seasons, the regressions are statistical significant and explain a high fraction of total MSLP variance at those grid points that belong to the majors centres of action of the EOF.

[5]If the regressions were taken onto *non-normalised* PCs, the patterns would be very similar to the optimal forced patterns of figure 5.7. Note also that, since the fraction of total variance explained by the regression are squared correlations, they are insensitive to the normalisation.

Since the analysis is linear, the opposite also applies: negative SST anomalies in the tropical Pacific during a cold ENSO episode induces a pattern of MSLP anomalies in the EA region that projects on the positive phase of the leading EOF of total variability MUGCM MSLP anomalies in the EA region.

In summary, it was shown that SST anomalies associated with ENSO forces the NAO, in winter and spring. The linear analysis showed that the warm (cold) ENSO phase induces a negative (positive) phase of the NAO, specially in winter.

The same analysis was performed using the normalised standard forced PC-1, instead of the optimal forced PC-1, and almost identical results (not shown) were obtained. Note that the correlation between the standard and optimal leading PCs is higher that 0.93 (see figure 5.8) (a,b) and that they represent the same pattern of variability (compare figures 5.1 (a,b) with 5.7 (a,b)). The choice of using the optimal forced PC-1 instead of the standard forced PC-1 is justified by the fact that the optimal PC is more reliable, when compared to analogue time series obtained from the NCEP data, than the standard PC (see figure 5.9).

5.6.2 Atlantic Interhemispheric SST Gradient forcing

The Atlantic Interhemispheric SST Gradient (AISG) mode is defined here as the leading mode of a PCA of the SST anomalies in tropical Atlantic (20°S-20°N, 80°W-20°E)[6]. The AISG mode represents 48.3% of the spring SST variability in the tropical Atlantic and is presented in figure 5.14. The pattern is plotted in panel (a), as coloured shadings, in and its normalised PC (representing the MAM index) is plotted in panel (b) with a red dashed line.

Overlaid on the spring AISG pattern in figure 5.14 (a), we plotted, as contour lines, the pattern of regression coefficients of SST anomalies onto the normalised optimal PC-1 in spring. The normalised optimal forced PC-1 is presented, along with the AISG

[6]This region is similar to the region (21°S-32°N, 74°W-15°E) used for the Atlantic Meridional Mode (AMM) defined by the CAB/PSD/ESRL/NOAA - www.cdc.noaa.gov/ClimateIndices/List/#AMM.

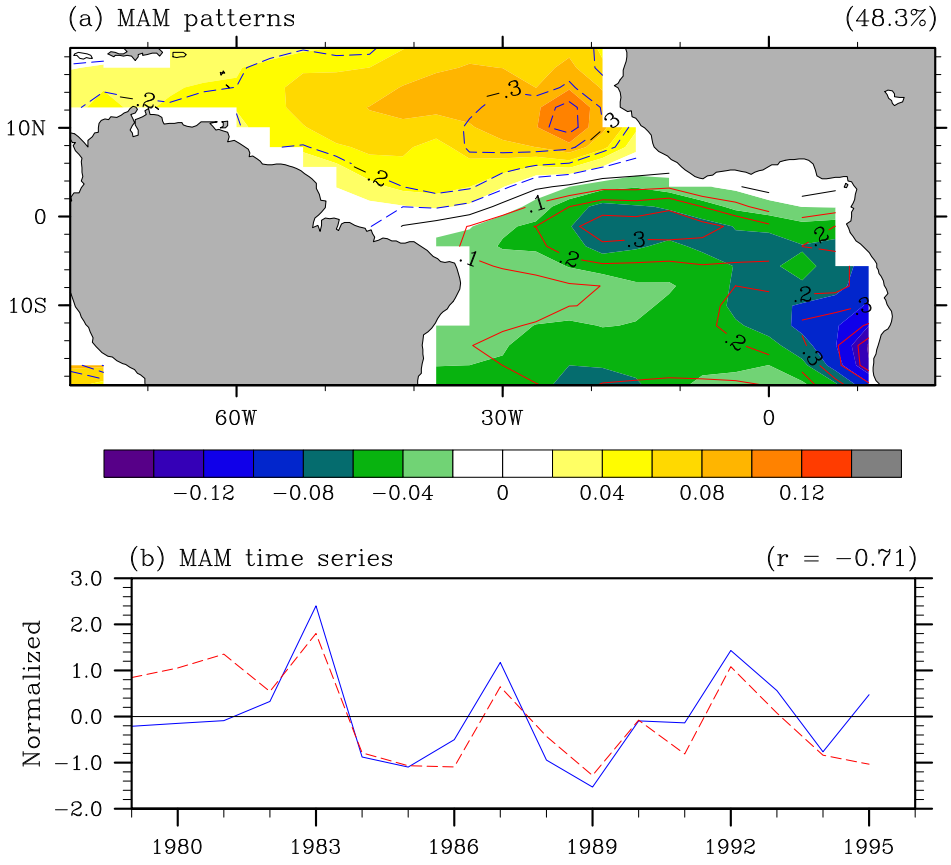

Figure 5.14: (a) Regression coefficients of SST anomalies onto the EA MSLP normalised optimal forced PC-1 (contour lines; contour interval is 0.1K, dashed for negative coefficients) overlaid on the SST EOF-1 (coloured shading; representing the AISG pattern in MAM. (b) EA MSLP normalised optimal forced PC-1 (blue line, plotted with opposite sign) and the normalised SST PC-1 (red dashed line; representing the AISG index) in MAM. Correlations significantly different from zero at a significance level not higher than 0.002. Values printed on the top right-hand corner of the panels represent the percentage of total variance explained by the SST Inter-Hemispheric mode in (a) and the correlation between the time series in (b).

index, in figure 5.14 (b) as a continuous blue line plotted with opposite sign for easier comparison.

The results from figure 5.14 confirms that the spring AISG forces the optimal forced PC-1 in spring. The correlation between the SST gradient index and the optimal forced PC-1 is -0.71, with a significant level not higher then 0.002.

Figures 5.14 and 5.13 (b) show that a positive northward SST anomaly gradient in the tropical Atlantic (positive AISG: positive SST anomalies in the northern hemisphere and negative anomalies in the southern hemisphere), induces a pattern of MSLP anomalies in the EA region that projects onto the negative phase of the leading EOF of the total variability MUGCM MSLP anomalies in the EA region in spring (section 3.4.2, figure 3.23 (a)) that, as mentioned in section 5.6.1, was recognised to be similar to the observed NAO pattern (section 3.4.2, figure 3.23 (b)).

Since the analysis is linear, the opposite can also applies: a negative northward SST anomaly gradient in the tropical Atlantic (negative AISG) induces a pattern of MSLP anomalies in the EA region that projects on the positive phase of the leading EOF of total variability MUGCM MSLP anomalies in the EA region in spring.

In summary, it was shown that SST anomalies associated with the AISG forces the NAO, in spring. The linear analysis showed that a positive (negative) AISG induces a negative (positive) phase of the NAO.

5.6.3 North Atlantic SST Tripole forcing

The North Atlantic SST anomaly Tripole mode is defined here as the leading mode of a PCA of the SST anomalies in north Atlantic basin (0°-60°N, 80°W-0°)[7]. This mode represents 30.4% of the winter SST variability in the north Atlantic and is presented in figure 5.15. The pattern is plotted in panel (a), as coloured shadings, in and its normalised PC (representing the Tripole index) is plotted in panel (b) with a red dashed line.

Overlaid on the winter SST Tripole pattern figure 5.15(a), we plot, as contour lines, the pattern of regression coefficients of SST anomalies onto the normalised optimal PC-2 in winter. The normalised optimal forced PC-2 is presented, along with SST Tripole

[7]This region is similar to the region (10°S-70°N, 80°W-0°) used for the Atlantic Meridional Mode (AMM) defined by the CAB/PSD/ESRL/NOAA - www.cdc.noaa.gov/ClimateIndices/List/#Atlantictripole

index, in figure 5.15(b).

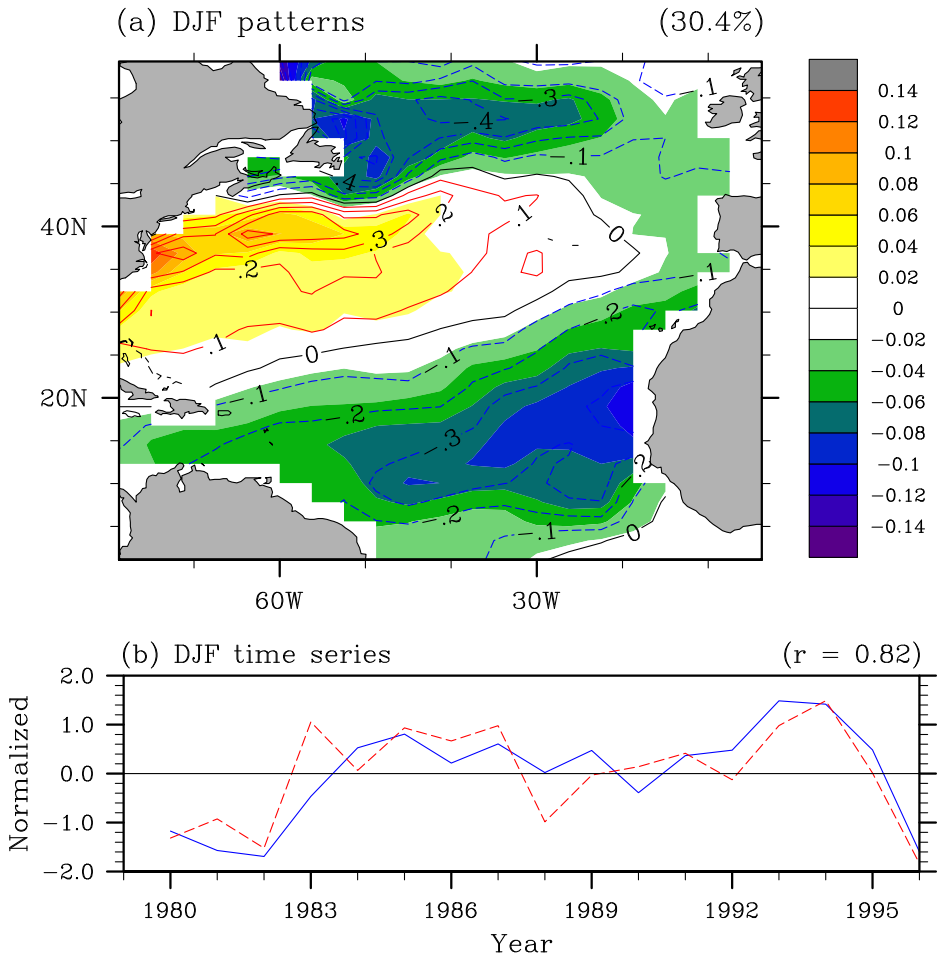

Figure 5.15: (a) Regression coefficients of SST anomalies onto the EA MSLP normalised optimal forced PC-2 (contour lines; contour interval is 0.1K, dashed for negative coefficients) overlaid on the SST EOF-1 (coloured shading; representing the North Atlantic SST Tripole pattern), in DJF. (b) EA MSLP normalised optimal forced PC-2 (blue line) and the SST PC-1 (red dashed line; representing the Tripole index) in DJF. Correlations significantly different from zero at a significance level not higher than 0.001. Values printed on the top right-hand corner of the panels represent the percentage of total variance explained variance by the SST Tripole mode in (a) and the correlation between the time series in (b).

The results from figure 5.15 confirms that the winter SST anomaly Tripole in the Atlantic forces the optimal PC-2 in winter. The correlation between the SST Tripole index and the optimal forced PC-2 is 0.82, with a significant level not higher then 0.001.

Figure 5.16 shows the pattern (contour isolines) obtained by linear regressing the ensemble mean of MSLP anomalies at each grid point of the EA region, in spring, onto the *normalised to unit variance* spring optimal forced PC-2, and also the pattern (shaded contours) of the fraction of total MSLP variance at each grid point explained by the regression, given by the squared correlation between the time series.

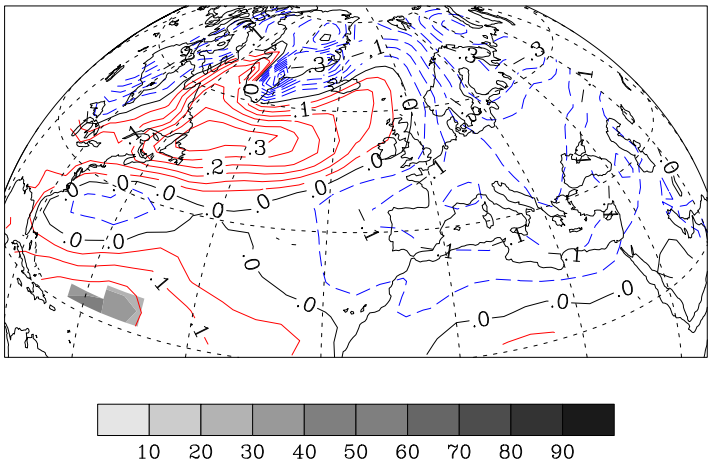

Figure 5.16: Regression coefficients (contour lines) and percentage of total variance explained (shaded contours) by the regression of the MUGCM ensemble mean of MSLP anomalies onto the normalised optimal forced PC-2 of the Euro-Atlantic MUGCM MSLP anomalies, in winter. White areas are where the regression is not significant at the 0.05 significance level. Contour interval is 0.05 mb per standard deviation of the PC. Dashed blue, solid black and solid red lines for negative, zero and positive coefficients, respectively. Grid lines represented every 20°.

As can be seen in 5.16 the regressions are not significant at the 0.05 significance level. This means that the forcing of the SST Tripole on the optimal PC-2 in winter is not strong enough to be induce a statistical significant forcing on the MUGCM MSLP anomalies at the grid points of the EA region.

5.7 Summary and conclusions

In sections 5.1 to 5.3 we determined, for winter and spring, estimates of the two leading forced variability modes of MSLP in the EA region, using three different methods: PCA of the ensemble-mean (standard PCA), SVDA of Ward and Navarra [1997] and the optimal detection method of Venzke et al. [1999] (optimal PCA). Due to the large ensemble size (29), the results of SVDA are equivalent to those of standard PCA, and, because of this, no further analysis was done with the SVDA results. So, only the standard and optimal PCA results are discussed here.

The standard and optimal patterns of these modes are presented in figures 5.1 and 5.7, respectively, and the associated time series (standard and optimal PCs) are shown, respectively, in figures 5.2 (a-d) and 5.6 (a-d) as red lines. The leading forced patterns obtained, in each season, by the the two methods are similar, certainly because of the large number (29) of the MUGCM's ensemble. For the second mode, the difference between the patterns yielded by the two methods become more evident, specially in winter.

Using an ANOVA based test, it was found that both standard PC-1 and PC-2 are statistically significant in both seasons (table 5.1). The same test performed onto the optimal PCs showed that both optimal PC-1 and PC-2 are also statistically significant in both seasons, and even statistically consistent (by the definition of Venzke et al. [1999]) in the case of the optimal PC-1 in both seasons.

In section 5.4, the standard PCs and optimal PCs, in winter and spring, were compared with each other (figure 5.8) and also compared with analogue signals obtained from the observations (figure 5.9). It was found that the two methodologies yield different results only for the second mode. Despite the high agreement between the standard PC1 and optimal PC1, only the optimal PC-1 was found to be very well correlated with the observations, in both seasons. The optimal PC-2 also shows some agreement with the observed counterpart but only in winter.

In section 5.5, the time series of seasonal SST anomalies, at each grid point of the globe, is regressed onto the normalised standard and optimal forced PCs (figures 5.10 and 5.11). From these global regression patterns, three localised statistical significant regression patterns outstand: (i) the regression pattern in the equatorial Pacific, found by the regression of the SST anomaly field onto standard PC-1 and optimal PC-1, in winter and spring, resembling the SST anomaly pattern of the ENSO; (ii) the regression pattern in the tropical Atlantic, found by the regression of the SST anomaly field onto the optimal PC-1 in spring, resembling the pattern of the AISG. (ii) the regression pattern in the north Atlantic basin, found by the regression of the SST anomaly field onto the optimal PC-2 in winter, resembling the SST Tripole.

The results of section 5.5, summarised in the last paragraph, motivated the analysis performed in section 5.6. The issue was to test if the three SST modes of variability (ENSO SST, AISG and SST Tripole) do force the MSLP variability in the EA region. This issue was addressed by (i) comparing the forcing SST pattern with the regression pattern of SST anomalies onto the forced optimal PC, (ii) comparing the forcing SST index with the optimal forced PC, and (iii) diagnosing the percentage of the MUGCM ensemble mean MSLP variance, at each grid point of the EA region, explained by the optimal PC.

In subsection 5.6.1 it was shown that ENSO SST pattern has a remarkable agreement with the regression pattern of SST anomalies onto the optimal forced PC-1, in both winter and spring (figure 5.12 (a,b)), and that the associated time series, the ENSO index and the optimal forced PC-1, have a statistical significant high correlation (0.94) in both seasons (figure 5.12 (c,d)). Furthermore, the regression pattern of the ensemble mean MSLP anomalies in the EA region onto the optimal forced PC-1 (5.13 (a,b)) projects well onto the positive phase of the NAO, specially in winter, and explain a high fraction of total MSLP variance in the centres of action of the NAO. Consequently, positive (negative) SST anomalies in the tropical Pacific during a warm (cold) ENSO episode induces a pattern of MSLP anomalies in the EA region that projects onto the negative (positive) phase of the NAO.

In subsection 5.6.2 it was shown that, in spring, the Atlantic Inter-Hemispheric SST anomaly Gradient pattern agrees very well with the regression pattern of SST anomalies onto the optimal forced PC-1 (figure 5.14 (a)), and that the associated time series, the AISG index and the optimal forced PC-1, have a statistical significant correlation of -0.71 (figure 5.14 (b)). Since the regression pattern of the MUGCM ensemble mean MSLP anomalies onto the optimal PC-1 in spring (5.16) projects onto the positive phase of the NAO, and explain a high fraction of total MSLP variance in the centres of action of the NAO, it was concluded a positive (negative) AISG induces a negative (positive) phase of the NAO in spring.

In subsection 5.6.3 it was shown that, in winter, the North Atlantic SST Tripole pattern is very similar to the regression pattern of SST anomalies onto the optimal forced PC-2 (figure 5.15 (a)), and that the associated time series, the Tripole index and the optimal forced PC-2, have a statistical significant correlation of 0.82 (figure 5.15 (b)). However, the regression pattern of the MUGCM ensemble mean MSLP anomalies onto the optimal PC-1 in winter (5.16) is no statistical significant. Consequently, the forcing of the SST Tripole on the optimal PC-2 in winter is not strong enough to be induce a statistical significant forcing on the MUGCM MSLP anomalies at the grid points of the EA region.

Combining the results of last sections, we conclude that the use of the optimal detection PCA brings no huge advantage, relative to standard PCA, in detecting the leading forced mode of MSLP in the EA region. The patterns are similar, the time series (PCs) are highly correlated and the regression patterns of the global SST anomalies onto the PCs are also similar: the regression patterns in the equatorial Pacific (associated to the ENSO SST pattern) and in the tropical Atlantic (associated to the AISG pattern). Nevertheless, the ANOVA shows that the leading optimal detected response is not only significant but also consistent, and the comparison with observations give reliability to the optimal detected forced response. Furthermore, while the regression pattern of SST anomalies onto the optimal PC-1 is statistical significant in the tropical Atlantic in winter, the regression pattern onto the standard PC-1 is not significant in this region

(or, more precisely, it is significant only in part of the pattern).

Chapter 6

Tropical Pacific Ocean Forcing

In chapter 5 it was shown that, the winter and spring major forced mode of MSLP variability in the EA region is forced by SST anomalies associated with the ENSO phenomenon. Furthermore, it was shown that the forced MSLP pattern projects very well on the NAO pattern in winter. These results, thus, suggest that the NAO is forced by the ENSO in the winter season, and motivate further investigation.

In this chapter, we address the forcing of the tropical Pacific SST anomalies on the MSLP variability in the EA sector by studying the sensitivity of the NAO to ENSO polarity and strength. We use the 29 17-year integrations of the MUGCM to estimate the Probability Density Functions of composites of NAO indices associated with warm and cold ENSO years.

The results show signals of the ENSO phases both in the mean strength of the NAO as well as in its internal variability. During the cold ENSO phase, the Probability Density Function (PDF) of the NAO index presents a small but positive mean value, whereas it is negative during the warm ENSO phase. Also the NAO variability associated with each ENSO phase shows a different behaviour: during the warm phase the PDF of the NAO index presents a larger variance and suggests a bimodality, whereas no bimodality is suggested in the cold phase.

Part of the study presented in this chapter is analogous to the study of Melo-Gonçalves et al. [2005]. Note, however, that Melo-Gonçalves et al. [2005] used HGT-850 seasonal anomalies instead of MSLP seasonal anomalies. The authors chose HGT-850 because it is a level at the top of the boundary layer and so more sensitive to large scale effects. They note, however, that this choice is not important for the identification of the NAO. In fact the NAO has been identified both at the mean sea level pressure (see figure 3.22 (b) or Glowienka-Hense [1990]) as well as at the 700 hPa geopotential field (e.g. Barnston and Livezey [1987]). Another difference between the work of Melo-Gonçalves et al. [2005] and this chapter is the definition of the Euro-Atlantic domain: (20^oN-85^oN, 60^oW-60^oE) instead of (20^oN-80^oN, 90^oW-60^oE) used here.

6.1 Introduction

The NAO is the most prominent atmospheric variability mode over the EA region. It is a mode identified long time ago in the observations [Walker and Bliss, 1932]. In the late 20th century, the simulation of the NAO using AGCMs forced by climatological surface boundaries gave a clear proof that the NAO appears as an internal mode of the atmospheric circulation (e.g., Barnett [1985]; Glowienka-Hense [1990]). The NAO is also well reproduced by the current coupled ocean-atmospheric models [Paeth et al., 1999] and, if the SST field is prescribed as observed during the last century, also part of the historic behaviour of a low-pass filtered NAO index can be reproduced [Rodwell et al., 1999; Latif et al., 2000].

An important finding of section 5.6.1 is that the dominant MSLP mode of forced variability in winter, in the EA region, has a meridional dipole structure that projects on the positive phase of the NAO (figure 5.7 (a)). Consequently, the forcing of the ENSO SST mode on the dominant forced PC suggests a negative correlation between the ENSO index and NAO index. The observed correlation between these indices is, however, close to zero, as shown in figure 6.1 where a standardised station-based NAO index and the standardised Niño 3.4 index are plotted.

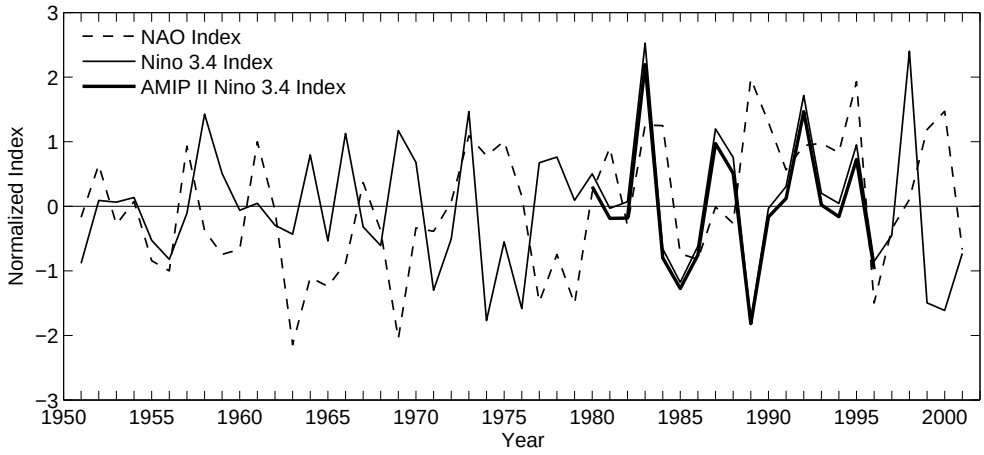

Figure 6.1: Station-based NAO index (dashed line), Niño 3.4 index from SST NCEP data (solid thin line) and Niño 3.4 index from SST AMIP II data (solid heavy line). All indices computed from DJF seasonal mean anomalies and normalised to unit variance.

The NAO index was computed, as in section 4.4, by subtracting the normalised time series of MSLP at Stykkisholmur, Iceland (65.1^{o}N, 22.7^{o}W) from that at Lisbon, Portugal (38.7^{o}N, 9.1^{o}W). The resulting time series was then normalised to unit variance for plotting purposes. Here we used the NCEP grid points nearest to Stykkisholmur and Lisbon: (65.0^{o}N, 22.5^{o}W) and (37.5^{o}N, 10.0^{0}W), respectively. The Niño 3.4 index is defined as the area-average of SST anomalies over the domain (5^{o}S-5^{o}N, 190^{o}E-240^{o}E). The index was also normalised to unit variance for plotting purposes. Two normalised Niño 3.4 indices are shown in figure 6.1: one computed using the full time range of the NCEP reanalysis (DJF 1951 - DJF 2000) and the other using the AMIP II time range data (DJF 1980 - DJF 1986). Note that both Niño 3.4 indices may be considered the same in the AMIP II years. Their correlation is 0.999 with a significance level lower that 1%.

The correlation between Niño 3.4 index and the observed NAO index is -0.04 and 0.24, when the NCEP time range and the AMIP II time range, respectively, are considered. Although low, these correlations are not significant at the 1% (or even 10%) significance level.

The lack of a significant correlation between Niño 3.4 index and the observed NAO index does not imply necessarily that SST anomalies associated with ENSO do not force MSLP variability associated with the NAO. Note that, as seen in section 4.4, the fraction of the modelled station-based NAO index total variance due to SST-SIC forcing in DJF does not exceed 5%. Feldstein [2000] used NCEP winter daily data from 1958 to 1997 to estimate the signal (interannual variance due to external forcing) to noise (interannual variance from stochastic processes) ratio. The estimated ratio of 0.09 indicates that interannual variability of the NAO arises primarily from climate noise. Thus, it is quite obvious that it is the internal variability of the NAO that is uncorrelated with ENSO. On the other hand, the observed (NCEP) NAO must be regarded as only one realisation of a random variable, and the correlation between this realisation with Niño 3.4 is itself also a single realisation of a random variable.

6.2 Seasonal modulation of the ENSO forcing

In this work, winter seasonal means have been defined as averages of December, January and February monthly means. These are the months traditionally used to define the boreal winter season (e.g., Walker [1925]), and, because of that, it has been called the *standard winter season* in this work. The December to February season has been extensively used to study worldwide ENSO teleconnections [van Loon and Madden, 1981; van Loon and Rogers, 1981; Horel and Wallace, 1981; Pan and Oort, 1983; Kiladis and Diaz, 1989; Hoerling et al., 1997], the impact of ENSO on the EA atmosphere [Dong et al., 2000; Cassou and Terray, 2001; Pozo-Vásquez et al., 2001; Merkel and Latif, 2001; Cassou and Terray, 2004; Lin and Derome, 2004; Mathieu et al., 2004; van Oldenborgh, 2005; Pozo-Vásquez et al., 2005b,a], and also the ENSO impact on the NAO [Rogers, 1984; Rodwell et al., 1999; Peterson et al., 2002; Lin and Derome, 2002; Lin et al., 2005; Coppola et al., 2005; Melo-Gonçalves et al., 2005].

The use of the December to February season to study ENSO teleconnections on the EA region has been questioned since the work of Huang et al. [1998]. These authors

applied a multi-resolution cross-spectral analysis to the time series of monthly mean NAO index (defined by Hurrel [1995] as the difference in MSLP between Lisbon and Stykkisholmur) and monthly mean Niño 3 index, computed for each calendar month from 1965 to 1995, to identify the El Niño years in which the two time series are coherent. For all calendar months, the composite of the Niño 3 index formed with *coherent* years (70% of all El Niño years) have higher values than the composite formed with *incoherent* years (30% of all El Niño years). This means that coherent years are associated to moderate and strong ENSO events. The composite of the NAO index formed with those *coherent* years decays from a maximum (positive) value in November to a minimum (negative) value in February, the change of the sign occurring from about December to January. The negative index value persists for another four months or so (see figure 2 of Huang et al. [1998]).

Latter, Moron and Gouirand [2003] performed a singular value decomposition analysis and a composite analysis to monthly data from 1873 to 1996, and showed that the pattern of north Atlantic MSLP anomalies associated with tropical Pacific SST anomalies projects onto the positive phase of the NAO in November-December, and onto the negative phase of the NAO in February-March.

The results of Huang et al. [1998] and Moron and Gouirand [2003] suggest that the December-February season is not appropriate to study the ENSO signal in the EA region, since the signal in November-December is opposite to the one observed in January-March. In order to test if this is also the case in our model simulations, we performed a SVDA [Bretherton et al., 1992]) between monthly mean SST anomalies in the tropical Pacific and monthly mean MSLP anomalies over the EA region, for each calendar month from December to May.

Let $L_{m_L p}$ represent the monthly mean anomaly of SST at spatial location m_L ($m_L = 1, ..., M_L$) in year p ($p = 1, ..., P$), and let $R'_{m_R pk}$ be the monthly mean anomaly of MSLP at spatial location m_R ($m_R = 1, ..., M_R$) in year p ($p = 1, ..., P$) of the kth ($K = 1, ..., K$) simulation. Note that $R'_{m_R pk}$ represents the departure of the MSLP anomaly from the

ensemble mean (see equation 6.6). Now, let $\mathcal{L}_1$ be a $(M_L \times P)$ random data matrix whose (m, p) entry is $X_{m_L p}$, and let

$$\mathcal{L} = \big[\overbrace{\mathcal{L}_1|...|\mathcal{L}_1}^{K \text{ times}}\big], \tag{6.1}$$

which is a $(M_R \times M_L)$ matrix. Let also $\mathcal{R}_k$ be a $(M_R \times P)$ random data matrix whose (m, p) entry is $R'_{m_L pk}$, and let

$$\mathcal{R} = \big[\mathcal{R}_1|...|\mathcal{R}_k|...|\mathcal{R}_K\big], \tag{6.2}$$

which is also $(M \times PK)$ matrix.

The SVDA of Bretherton et al. [1992] consists of performing a singular value decomposition of the sample cross-covariance matrix between $\mathcal{L}$ and $\mathcal{R}$,

$$\mathbf{S}_{LR} = \frac{1}{P(K-1)} \mathcal{L}\mathcal{R}^{\mathrm{T}}, \tag{6.3}$$

which is a $(M_R \times M_L)$ matrix, that is,

$$\mathbf{S}_{LR} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}, \tag{6.4}$$

where $\mathbf{U}$ is a M_L-dimensioned square matrix whose columns form an orthonormal set of vectors called the *left singular vectors*, $\mathbf{U}$ is a M_R-dimensioned square matrix whose columns form an orthonormal set of vectors called the *right singular vectors*, and $\mathbf{\Lambda}$ is a $(M_R \times M_L)$ diagonal matrix whose non-zero $M = \min(M_R, M_L)$ diagonal elements are called the *singular values* of $\mathbf{S}_{LR}$.

The strength of coupling between $\mathbf{L}$ and $\mathbf{R}$ represented by the mth pair of patterns is measured by the fraction of total squared covariance - the *squared covariance fraction* (SCF, Bretherton et al. [1992]) - given by

$$SCF_m = \frac{\lambda_m^2}{\sum_{m'=1}^{M} \lambda_m^2}, \quad m = 1, \ldots, M = \min(M_L, M_R) \tag{6.5}$$

where λ_m is the mth singular value, that is, the value of the (m, m) entry in Λ.

Bretherton et al. [1992] defined the mth left (right) *homogeneous correlation map* as the spatial pattern of correlations between the grid point values of the left (right) field and the time series associated to the left (right) mth spatial pattern, that is, the time series given by the mth line of the $M_L \times PK$ ($M_R \times PK$) matrix $\mathbf{U}^{\mathrm{T}}\mathcal{L}$ ($\mathbf{V}^{\mathrm{T}}\mathcal{R}$).

SVDA was performed using the monthly SST anomalies in the tropical Pacific (20^{o}S-20^{o}N, 120^{o}E-290^{o}E) to build the left matrix, $\mathcal{L}$ (equation 6.1), and the monthly MSLP anomaly departures from the MUGCM ensemble mean in the EA region (20^{o}N-80^{o}N, 90^{o}W-60^{o}E) to build the right matrix, $\mathcal{R}$ (equation 6.2). Both time series at each grid point were normalised to unit variance prior to the computation of the cross-covariance matrix $\mathbf{S_{LR}}$ (equation 6.4). The analysis was done for each month from December to May. The homogeneous correlation maps of the first SVDA mode for each month is shown in figure 6.2 where the SCF (equation 6.5 with $m = 1$) is printed on the top left-hand corner of each panel.

The pattern of the right correlation map projects onto the negative phase of the NAO in all months, which confirms the results found in section 5.6.1. Note that, unlike the observational results of Huang et al. [1998] and Moron and Gouirand [2003], our modelled ENSO teleconnection on the EA region does not changes sign between December and January. Therefore, we will continue to use the December-February season, instead of January-March, as the boreal winter season to study the ENSO forcing on the EA sector.

6.3 NAO patterns and indices

In this section we use the MUGCM AMIP II ensemble of winter mean (DJF) MSLP anomalies over the EA sector to obtain two NAO patterns and associated indices: EOF-based and correlation-based NAO patterns and indices. First, these patterns are computed using internal variability data because the NAO arises primarily from internal dynamics (see section 4.4 and Feldstein [2000]). Analogous patterns are also computed with the 33-year ACYC experiment with the MUGCM (see chapter 2) and also with

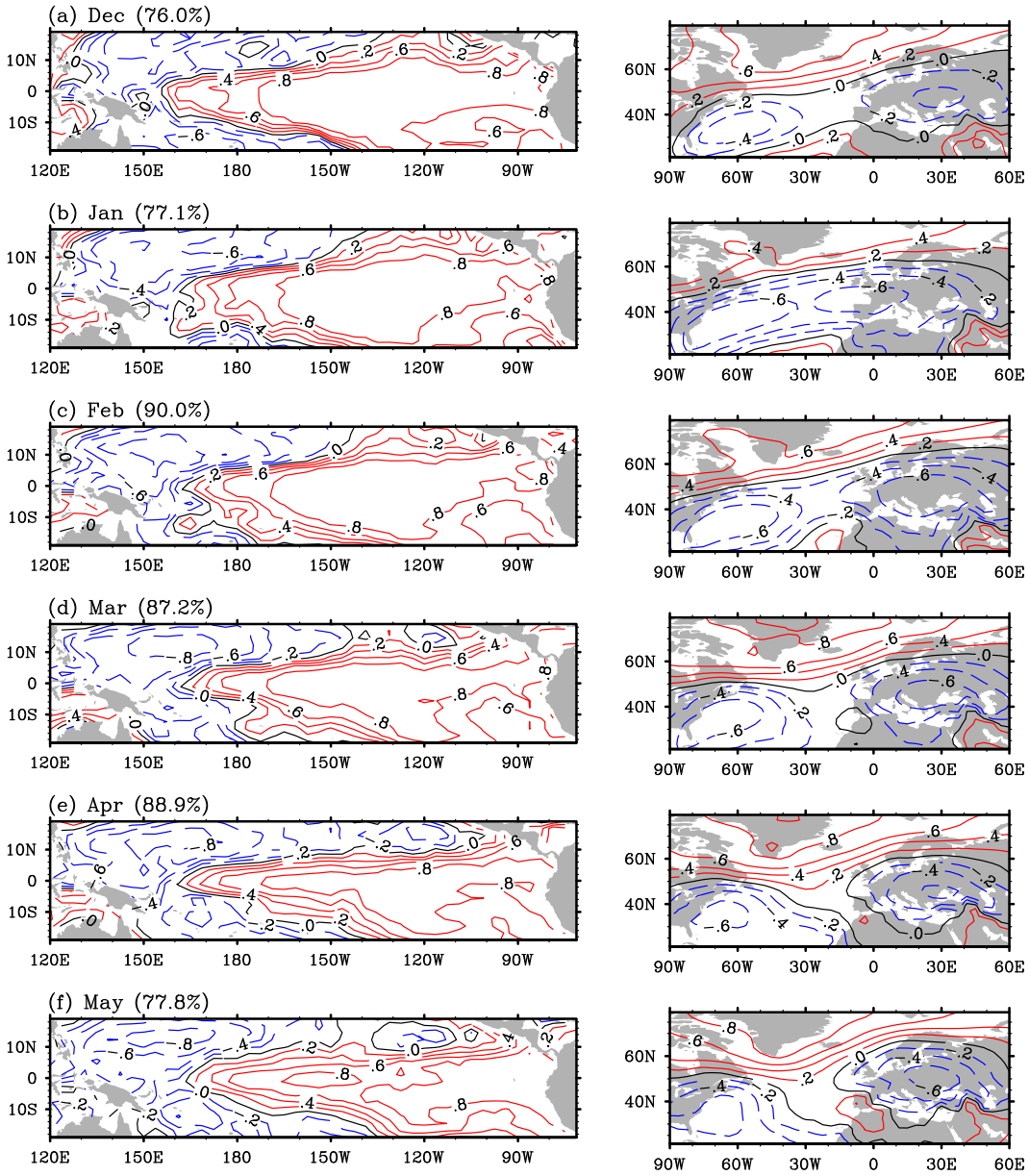

Figure 6.2: Homogenous correlation maps of the MUGCM first SVDA mode of tropical Pacific SST anomalies (left panels) and EA MSLP anomalies (right panels) in (a) October, (b) November, (c) December, (d) January, (e) February, (f) March, and (g) April. Negative, zero and positive correlations plotted with dashed blue, solid black and solid red isolines, respectively. Contour interval is 0.2. The percentages represent squared covariance fractions.

.

NCEP data. The ACYC NAO pattern is computed only to prove that the size of the AMIP ensemble is large enough to estimate the MUGCM internal variability. The

NAO pattern from the NCEP data is shown to test the capability of the MUGCM in simulating the internal dynamics of the atmosphere in the EA region.

Once the EOF-based and correlation-based NAO patterns are obtained, two NAO indices, associated with each pattern, are computed using the AMIP II data: a index representing the internal variability of the NAO and another representing its total variability.

Consider, as in section 5.1, that the random variable X_{mpk} represents the AMIP seasonal mean anomaly of MSLP at spatial location m ($m = 1, ..., M$) in year p ($p = 1, ..., P = 17$) of the kth ($k = 1, ..., K = 30$) simulation.

Each random variable X_{mpk} can be expressed as the sum of the ensemble mean in year p and the departure of the kth ensemble member from this mean:

$$X_{mpk} = \overline{X}_{mp\bullet} + X'_{mpk}, \tag{6.6}$$

where the dot notation represents, as before, averaging over the missing subscript.

Let $\mathcal{X}_k$ and $\mathcal{X}_{I,k}$ be $(M \times P)$ random data matrices whose (m, p) entries are X_{mpk} and X'_{mpk}, respectively. Also let $\mathcal{X} = \left[\mathcal{X}_1|...|\mathcal{X}_k|...|\mathcal{X}_K\right]$ and $\mathcal{X}_I = \left[\mathcal{X}_{I,1}|...|\mathcal{X}_{I,k}|...|\mathcal{X}_{I,K}\right]$ be $(M \times PK)$ random data matrices. These represent the *total variability* and *internal variability* data, respectively.

Both EOF-based and correlation based patterns are obtained from the the sample covariance matrix of internal variability $\mathcal{X}_I$,

$$\mathbf{S}_I = \frac{1}{P(K-1)} \mathcal{X}_I \mathcal{X}_I^{\mathrm{T}}, \tag{6.7}$$

which is a multivariate version of equation 4.10 and is an unbiased estimator of the *true* covariance matrix of internal variability, $\mathbf{\Sigma}_I$.

6.3.1 PCA-based NAO pattern

To obtain the model internal NAO pattern we performed an Orthogonal Empirical Function (EOF) analysis onto the internal variability data $\mathcal{X}_I$. We use $\mathcal{X}_I$ instead of the total variability data, $\mathcal{X}$, because, as already mentioned above, the NAO arises primarily from internal dynamics. Note that in section 3.4.2 the EOF analysis was performed onto a (biased) estimate of total variability data $\mathcal{X}$ because the purpose was to validate the total variability of the modelled data (MUGCM) against the observed data (NCEP).

The internal variability EOFs are computed as the unit-length eigenvectors of the sample covariance matrix of internal variability $\mathbf{S}_I$ (equation 6.7). The two leading patterns of internal variability of winter (DJF) seasonal mean MSLP anomalies is presented in figure 6.3 (a,b). They represent 30.7% and 25.0% of total internal variance. The NAO pattern is identified as the EOF-1 pattern.

In order to test if the AMIP II EOFs of internal variability represent well the true internal variability of the MUGCM, the two leading EOFs of the ACYC (see chapter 2) seasonal mean DJF MSLP anomalies (annual cycle removed) were also computed for the EA region. Note that, since the SST-SIC forcing is the same for all 33 ACYC simulations, internal variability can be estimated by the inter-member variability [Chervin, 1986; Chervin and Schneider, 1976] and, thus, the spatial patterns of internal variability in the EA region can be estimated by these EOF patterns [Saravanan, 1998; Cassou and Terray, 2001].

The first and second ACYC EOFs, presented in figure 6.3 (c,d), represent 37.6% and 25.0% of the variability. The NAO pattern, identified as the first EOF, agrees very well with the AMIP II NAO pattern. Both patterns show the characteristic meridional NAO dipole, and the amplitudes of the centres of action are similar.

In order to validate our simulated patterns of internal variability, we also computed the EOFs using NCEP data. The first and second EOFs are presented in figure 6.3

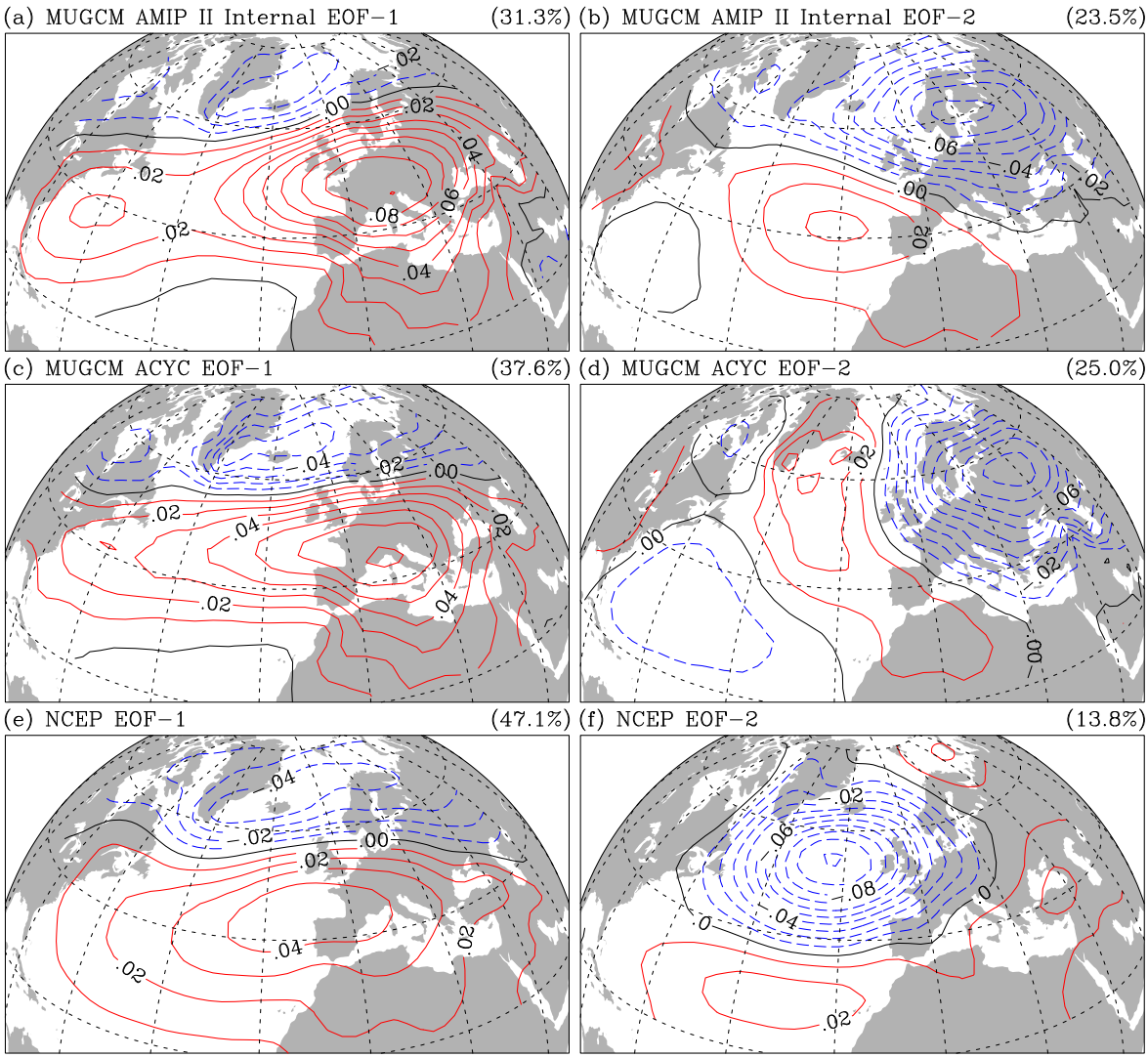

Figure 6.3: First (left) and second (right) EOFs of DJF seasonal mean MSLP anomalies of (a,b) MUGCM AMIP II internal variability, (c,d) MUGCM ACYC and (e,f) NCEP. Negative, zero and positive loadings plotted with blue, black and red isopleths, respectively. EOFs are unit length and dimensionless. Grid lines represented every 20^o.

(e,f), representing 47.1% and 13.8% of total variability. The NAO pattern, identified as the first EOF, is well simulated by the MUGCM.

The NCEP MSLP data represent, of course, total variability data containing the component due to internal dynamics of the atmosphere and the component due to external forcing agents (SST, volcanic eruptions, anthropogenic gases, etc). A traditional

method to extract internal variability from the observed data (or from a single AGCM simulation) is based on the principle that internal variability is originated from dynamical weather processes which have a time scale smaller than the timescale of the processes due to external forcing. Usually, the daily signal of a meteorological variable is decomposed into high frequency variability signal and a low frequency variability signal. The high frequency signal is then averaged over a season to obtain the seasonal internal variability signal (see, for example, von Storch and Zwiers [1999]). Melo-Gonçalves et al. [2005] filtered out low frequency variability from the NCEP HGT-850 winter seasonal mean anomalies by subtracting a 5-year running mean from the data. The aim of authors was to remove the observed positive trend of the NAO observed in the last four decades [Hurrel, 1995] which is considered as a result of external forcing. This procedure is not able, however, to remove all external forced signals and even worst, it may remove some components of interannual variability that result from seasonal integration of synoptic instabilities, that is, it may dump out some components of internal variability. Furthermore, since the variability at midlatitudes arises primarily from internal dynamics, total variability EOFs in the EA region are good approximations of internal variability EOFs. For these reasons, we simply show in figure 6.3 (e,f) the EOFs of total variability. These are the same patterns of patterns shown in figure 3.22 (b,d) and are reproduced here for better comparison with the MUGCM AMIP II EOFs.

In summary, the NAO pattern was identified as the leading EOF of both simulated (AMIP II and ACYC) and observed data. The similarity between AMIP II and ACYC leading EOFs prove that the AMIP II ensemble is large enough to estimate the leading mode of MUGCM internal variability. Besides the discrepancies with the NCEP NAO pattern, the AMIP II NAO pattern seem to be a fair simulation for the present purposes.

6.3.2 Correlation-based NAO pattern

The simulated NAO teleconnection is also demonstrated computing the teleconnectivity maps [Wallace and Gutzler, 1981] of the model AMIP internal variability. The

teleconnectivity at mth grid point is computed by

$$T_m = -100 \times \min(r_{mm'}), \quad m' = 1, ..., M \tag{6.8}$$

where $r_{mm'}$ is the sample correlation between the mth and m'th stations and is the (m, m') entry in the sample correlation matrix

$$\mathbf{R}_I = \mathbf{D}^{-1}\mathbf{S}_I\mathbf{D}^{-1}, \tag{6.9}$$

where $\mathbf{D}$ is a diagonal matrix whose elements are the square roots of the corresponding elements in $\mathbf{S}_I$ (equation 6.7), that is, $d_{mm} = \sqrt{s_{mm}}, m = 1, ..., M$.

Note that the sample correlation matrix $\mathbf{R}$ is a $(M \times M)$ matrix where the elements in the mth column are the correlations of the data at the mth station with all the stations. Thus, equation 6.8 is equivalent to finding the minimum value (strongest negative correlation) in the mth column of $\mathbf{R}$ and multiplying this value -100.

Figure 6.4 maps the teleconnectivities T_m computed from $\mathbf{R}_I$, that is, the teleconnectivity map of winter MSLP anomalies of internal variability, $\mathbf{X}_I$. The grid points of the dipole with highest teleconnectivity are marked with black points along with its teleconnectivity value.

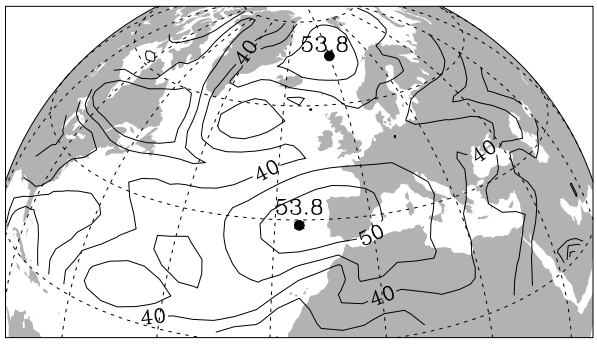

Figure 6.4: Teleconnectivity map of internal variability of MUGCM AMIP II DJF MSLP anomalies. The dots identify the pair of grid points with highest teleconnectivity value. Isoline interval is 5. Grid lines represented every 20^o.

The highest teleconnectivity (T = 53.2) dipole data has its north and south poles located at (70.4^{o}N, 0^{o}E) and (39.1^{o}N, 15^{o}W), respectively, that is, close to the locations usually used to compute the NAO index: Iceland and Azores or Lisbon. The northern grid point of the AMIP II dipole was used as base point to compute one-point correlation maps [Wallace and Gutzler, 1981] for AMIP II, ACYC and NCEP MSLP anomalies in winter. These are presented in figure 6.5. Both maps show the characteristic meridional NAO dipole pattern.

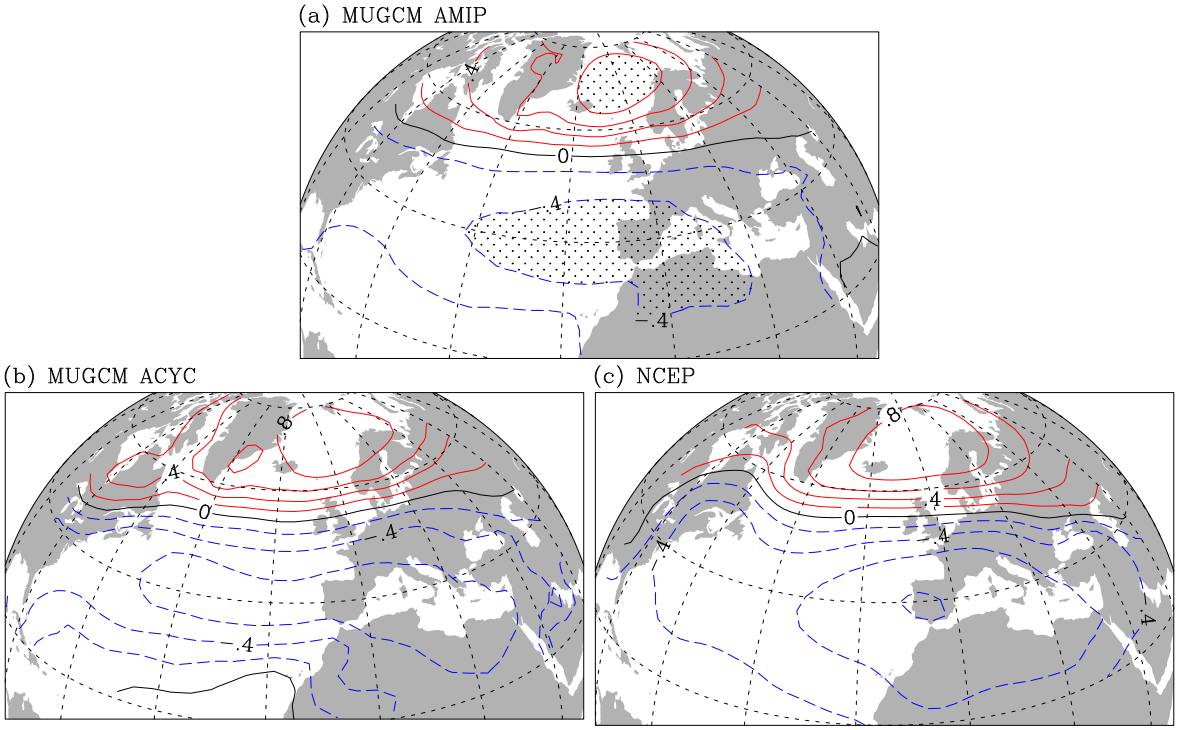

Figure 6.5: One-point correlation maps of DJF MSLP anomalies of (a) MUGCM AMIP II internal variability, (b) MUGCM ACYC and (c) NCEP. Isoline spacing is 0.2. Negative, zero and positive correlations plotted with blue, black and red isopleths, respectively. Grid lines represented every 20^o.

6.3.3 NAO indices

From the AMIP EOF-based NAO pattern (figure 6.3 (a)) two indices were computed: an internal variability index,

$$\mathrm{NAOpc_T} = \mathbf{e}_{I,1}^{\mathrm{T}} \mathcal{X}_I, \tag{6.10}$$

which is the principal component associated with the leading EOF of internal variability, $\mathbf{e}_{I,1}$, and a total variability index

$$\mathrm{NAOpc_T} = \mathbf{e}_{I,1}^{\mathrm{T}} \mathcal{X} \tag{6.11}$$

which is the projection of total variability data, $\mathcal{X}$, onto the same EOF. Each index was then normalised by its standard deviation.

A second pair of indices is based on the AMIP one-point correlation map of figure 6.5(a). We computed the area-weighted averages of the internal, X'_{mpk}, and total, X_{mpk}, winter anomalies inside each stippled region (correlations greater or equal to 0.8 in the northern pole and less or equal to -0.4 in the southern pole). For each centre, both internal and total area-weighted averaged anomalies were normalised by their respective standard deviation. Internal and total NAO indices were then defined as the difference between the northern and southern normalised series of the internal and total variabilities, respectively. Each index was then normalised by its standard deviation. We designate these indices by $\mathrm{NAOi_I}$ and $\mathrm{NAOi_T}$ for the internal and total variability, respectively.

6.4 ENSO episodes

In order to detect the winters of occurrence of cold and warm ENSO episodes, the Niño 3.4 index computed from the AMIP II DJF seasonal mean anomalies is shown in figure 6.6 as blue (negative values) and red bars (positive values). Note that this is exactly the same index shown in figure 6.1 as a solid heavy line.

Since ENSO is a coupled oceanic-atmospheric mode of variability, the Multivariate ENSO Index (MEI) from PSD/ESRL/NOAA [1] is also shown in figure 6.6 as black bars. The MEI is derived from tropical Pacific Comprehensive Ocean-Atmosphere Data Set (COADS) records and is a multivariate measure of the ENSO signal, since it is the first

[1] Physical Sciences Division/Earth System Research Laboratory/National Oceanic and Atmospheric Administration - www.cdc.noaa.gov/ENSO/enso.mei_index.html

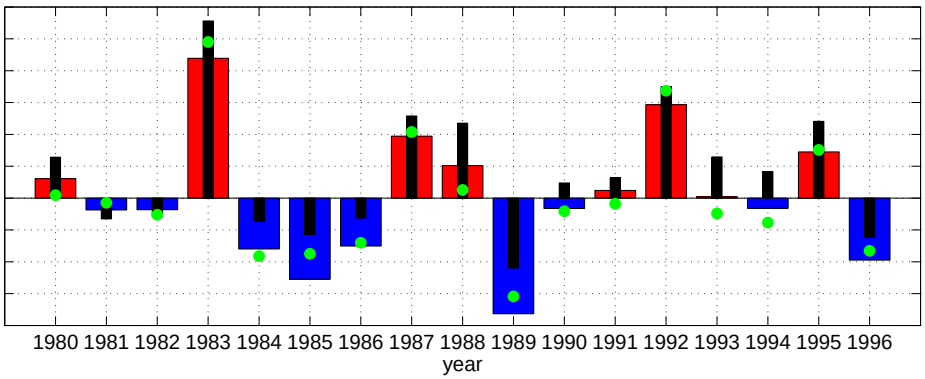

Figure 6.6: Niño 3.4 index computed from SST AMIP II data (blue/red bars for negative/positive values), Multivariate ENSO index (MEI) from PSD/ESRL/NOAA (black bars) and Convective Index (CI) computed from PRECIP AMIP II data (green dots). Niño 3.4 and CI computed from DJF seasonal mean anomalies and normalised to unit variance. MEI values are DJ seasonal mean anomalies and renormalised to unit variance with respect to the 1980-96 period.

principal component of six observed variables over the tropical Pacific: SST, TMP-SFC, MSLP, surface zonal and meridional wind components and cloudiness [Wolter and Timlin, 1998]. Positive (negative) values of the MEI represent the warm (cold) ENSO phase. The MEI was normalised to unit variance with respect to the 1980-1996 period.

Figure 6.6 also shows, as green dots, an index representing the convection in the tropical Pacific, the Convective Index (CI, Lin et al. [2005]). This index was constructed by averaging the DJF seasonal mean MUGCM AMIP II PRECIP anomalies over the area of (5^oS-5^oN, 120^oE-90^oW), and normalising to unit variance.

The years with occurrence of cold and warm ENSO episodes were selected by the non-normalised Niño 3.4 index plotted in figure 6.6, using the criterion presented in table 6.1, with the exception of 1988 that was not considered a weak warm event because its CI value is very low. Under this criterion, 53% of the years are characterised by either an El Niño (warm ENSO SST) or La Niña (cold ENSO SST) event, which is consistent with the definition of El Niño and La Niña given by [Trenberth, 1997].

Table 6.1: Years used for composites of ENSO episodes in the 1980-1996 period, based on the normalised Niño 3.4 index. The winter of 1988 that was not considered a weak warm event because its CI value is very low (see the plots of Niño 3.4 index and CI in figure 6.6).

ENSO composite name	Niño 3.4 (s)	Years
Cold ENSO	$s \leq -1$	1985 1989 1996
Cold ENSO & Weak cold ENSO	$s \leq -0.5$	1984 1985 1986 1989 1996
Warm ENSO & Weak Warm ENSO	$s \geq 0.5$	1983 1987 1992 1995
Warm ENSO	$s \geq 1$	1983 1987 1992

6.5 Probability Density Functions

In section 6.3.3 we defined four indices to measure the strength of the simulated NAO: $\mathrm{NAOpc_I}$ and $\mathrm{NAOi_I}$ for the internal variability and $\mathrm{NAOpc_T}$ and $\mathrm{NAOi_T}$ for the total variability. To assess the NAO sensitivity to ENSO polarity, we extracted two subsets from these unit variance indices: one consisting of cold ENSO winters and another composed by warm ENSO winters.

We estimated the Probability Density Functions (PDFs) for all years, cold and warm ENSO phases composites, and for 'neutral' years, i.e. all years from 1980 to 1996 that are not included in table 6.1. The PDFs were estimated by the Kernel method [Silverman, 1986] using a normalised Gaussian Kernel function.The smoothing parameter was objectively determined by the least-squares cross-validation procedure for each index ($\mathrm{NAOpc_I}$, $\mathrm{NAOpc_T}$, $\mathrm{NAOi_I}$ and $\mathrm{NAOi_T}$) and used for all composites.

Figure 6.7 shows the estimated PDFs for the simulated internal and total NAO variabilities represented by $\mathrm{NAOpc_I}$ and $\mathrm{NAOpc_T}$ indices, respectively, and the corresponding PDFs of the $\mathrm{NAOi_I}$ and $\mathrm{NAOi_T}$ indices.

The PDFs were also estimated for the ENSO phases when the weak cold and weak warm ENSO episodes were included. These are presented in figure 6.8 along with the

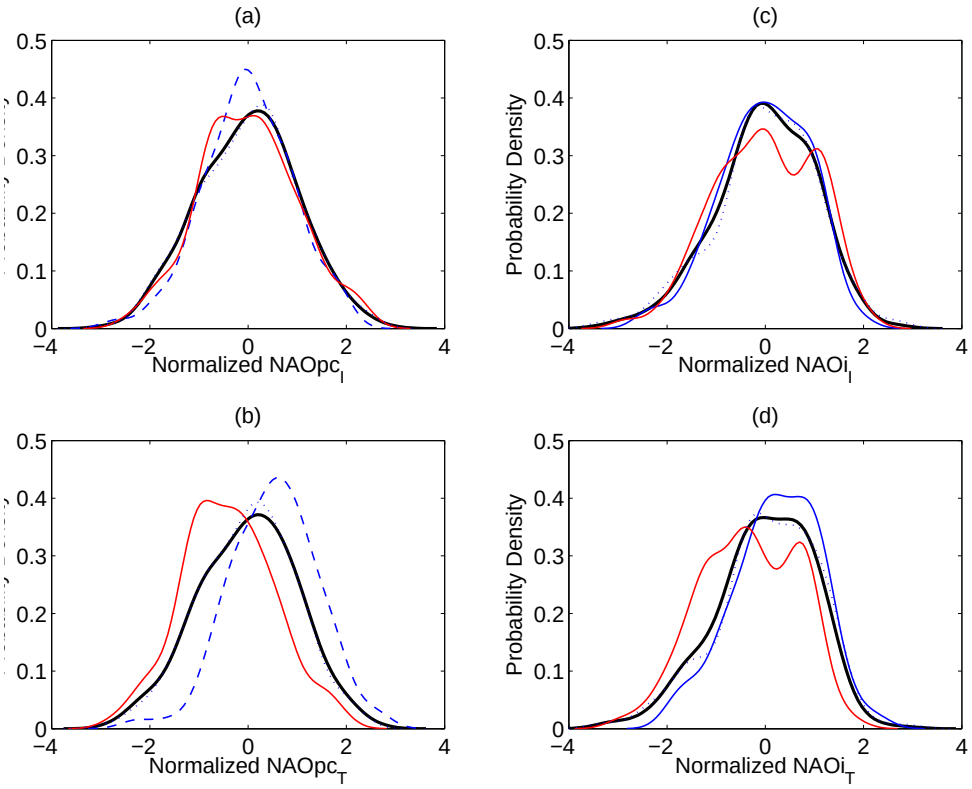

Figure 6.7: Estimated Probability Density functions (PDFs) for the NAO indices: (a) $NAOpc_I$, (b) $NAOpc_T$, (c) $NAOi_I$, and (d) $NAOi_T$. In each panel solid thick black, solid thin red, dashed blue and dotted black lines represent the PDFs for all, warm ENSO, cold ENSO, and neutral years composites, respectively. PDFs estimated by the Kernel method using a normalised Gaussian Kernel function. The smoothing parameter, computed by the least-squares cross-validation procedure using the all years composite and used for all composites, is 0.32 (a,b,d), and 0.31 (c).

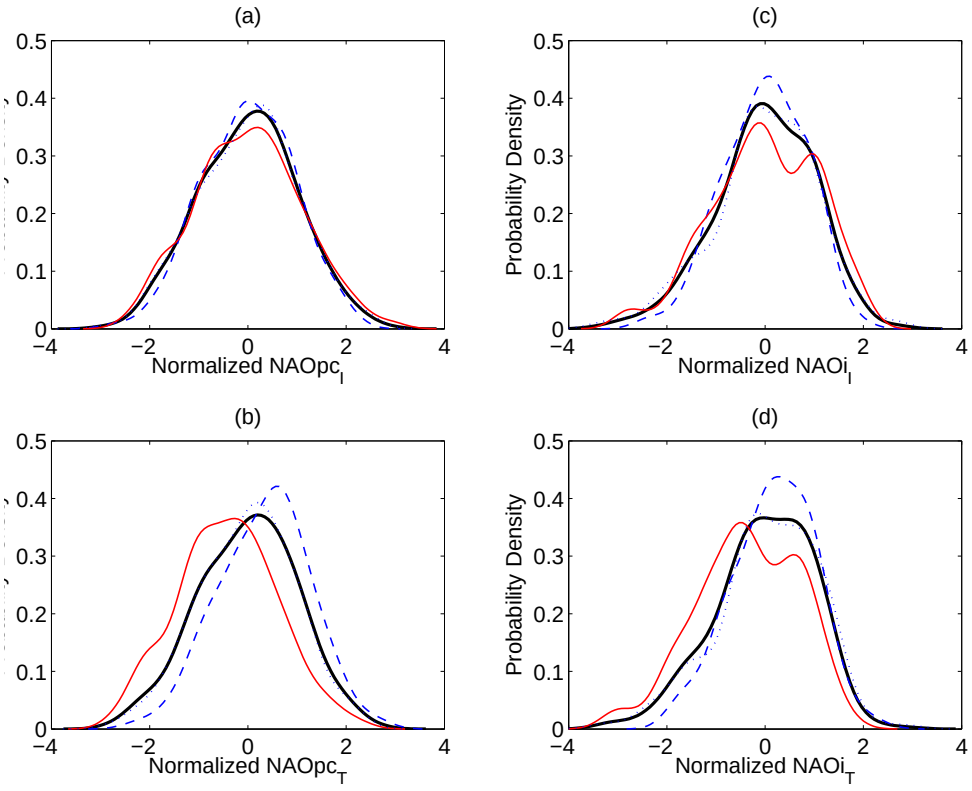

Figure 6.8: As in Figure 6.7 but including the weak ENSO episodes in the Cold ENSO composite and the weak warm ENSO episodes in the warm ENSO composite.

.

PDFs for all years and neutral years.

Figures 6.7 and 6.8 suggest a different NAO behaviour in cold and warm ENSO phases. In order to test if the PDFs of the NAO signal are statistically different for cold and warm ENSO phases, we applied the Smirnov test [Conover, 1971], also known as Kolmogorov-Smirnov two-sample test, to the cold ENSO and warm ENSO series. This is a non-parametric test whose test statistic is the largest difference between the Empirical Distribution Functions of the two samples. The Probability Distribution Functions are given by

$$F(x) = \int_{-\infty}^{x} f(x')dx', \tag{6.12}$$

where $f(x')$ is the PDF.

Table 6.2: p-values of the hypothesis tests applied to the cold ENSO and warm ENSO series.

Hypothesis Test	NAOpc_I	NAOpc_T	NAOi_I	NAOi_T
Smirnov test (equation 6.13)	0.379	0.000	0.556	0.001
t-test (equation 6.14)	–	0.000	–	0.000
F-test (equation 6.15)	0.002	0.039	0.000	0.000

Since the lower panels of Figures 6.7 and 6.8 suggest that smaller indices values are more probable in warm ENSO, the Empirical Distribution Functions for warm ENSO must be greater than the respective Functions for the cold ENSO phase. To assess the significance of this difference, we performed the one-sided test

$$H_0 : \forall x, F_w(x) \leq F_c(x) \ vs \ H_1 : \exists x, F_w(x) > F_c(x), \tag{6.13}$$

where F_c and F_w are the true Probability Distribution Functions of the cold ENSO and warm ENSO populations, respectively. The p-values (Table 6.2) obtained for internal variability are too high to reject the null hypothesis with a reasonably level of significance. However, for total variability, the Smirnov test rejects the null hypothesis at a significance level smaller than 0.1%, meaning that warm and cold ENSO samples come from different populations. The NAO indices tend to have lower values for the warm than for the cold ENSO phase.

The PDF of a population fully describes it. However, for the sake of interpretability, it is useful to extract some statistics from it. In Table 6.3 we present first and second order statistics (μ, σ^2) computed from the estimated PDFs of cold ENSO and warm ENSO series for internal and total variability. First and second order sample statistics $(\overline{X}, s^2)$ are also included for latter discussion.

For internal variability both cold and warm ENSO estimated populations have zero mean. This must be the case because the internal variability data is centred for each year. However, the variance of NAOpc_I (NAOi_I) is 18.9% (28.4%) greater for warm

Table 6.3: Sample and estimated population statistics of cold ENSO and warm ENSO internal and total variabilities.

		$NAOpc_I$	$NAOpc_T$	$NAOi_I$	$NAOi_T$
$\overline{X}$	WARM	0	-0.404	0	-0.338
	COLD	0	0.522	0	0.232
s^2	WARM	1.012	0.895	1.103	0.981
	COLD	0.764	0.753	0.700	0.664
μ	WARM	0	-0.404	0	-0.338
	COLD	0	0.522	0	0.232
σ^2	WARM	1.075	1.145	1.139	1.185
	COLD	0.904	1.128	0.887	0.912
$P(X \leq 0)$	WARM	0.485	0.613	0.490	0.618
	COLD	0.459	0.358	0.475	0.362
$P(X < 1)$	WARM	0.193	0.274	0.193	0.280
	COLD	0.136	0.100	0.127	0.091
$P(X > 1)$	WARM	0.197	0.083	0.215	0.082
	COLD	0.123	0.184	0.127	0.179

ENSO years. Despite the difference in the internal variances, the applied Smirnov test did not reject the null hypothesis that the $\mathrm{NAOpc_I}$ and $\mathrm{NAOi_I}$ indices for the warm and cold ENSO composites are samples of a same population.

Besides the differences of variances given in Table 6.3, the cold and warm ENSO distributions have different shapes. The PDFs for warm ENSO show a bimodality, more evident for correlation-based indices. Such a bimodality is not apparent in the cold ENSO phase. To test the robustness of this bimodality we recomputed the warm ENSO PDFs removing one year from the data subset at each estimation. This bimodality appeared in the three cases and seems to be a robust characteristic of the data. Furthermore, the bimodality is still present even when the weak episodes of each phase are included in the analysis.

For total variability, the statistically different cold and warm ENSO estimated populations have positive and negative means, respectively. The variance of $\mathrm{NAOpc_T}$ ($\mathrm{NAOi_T}$) is 1.5% (29.9%) greater for warm ENSO years. The rejection of the null hypothesis by the Smirnov test seems to be due more to the shifts (different means) of the distributions than to the differences in the variances. Note that the variances of the warm and cold ENSO phases in $\mathrm{NAOpc_T}$ are not very different, and still the null hypothesis was rejected. Note also that in the case of internal variability, with larger variance differences, the null hypothesis was not rejected.

Table 6.3 also shows first and second order sample statistics. To take in consideration the subtraction of the ensemble means, the internal variances for the warm and cold ENSO phases were computed using $3 \times (29 - 1)$ degrees of freedom. We performed the following hypothesis tests:

$$H_0 : \mu_c = \mu_w \ vs \ H_1 : \mu_c > \mu_w \tag{6.14}$$

$$H_0 : \sigma_w^2 = \sigma_c^2 \ vs \ H_1 : \sigma_w^2 > \sigma_c^2 \tag{6.15}$$

For the test of equation 6.14 we do not assume that the true variances are equal. We used a Z random variable, with the true variances replaced by the sample variances, as

test statistic. Invoking the Central Limit Theorem, this test statistic has, if H_0 is true, a t-Student distribution with a number of degrees of freedom that can be estimated by the Smith-Satterthwaite procedure [Milton and Arnold, 1995]. We will call this test, the t-test. For the test of equation 6.15, we used the usual test statistic which have, if H_0 is true, a Fisher distribution (we also invoke, here, the Central Limit Theorem because we do not assume that our samples come from Gaussian distributions). We will refer to this test as the F-test.

The p-values of the t-test and F-test are listed in Table 6.2. By the t-test we can conclude with a high level of significance that, for total variability, cold and warm ENSO samples come from populations with different means, the cold ENSO mean being higher that the warm ENSO mean. The F-test indicates that the warm ENSO population variance is higher than the cold ENSO population variance, for both internal and total variability. All variances are distinct at a significance level of lower than 0.1% except the $\mathrm{NAOpc_T}$ whose significance level is lower than 4%.

6.6 Summary and conclusions

We analysed here a large ensemble of NAO indices simulated by the MUGCM. The indices were constructed to represent the internal and total variabilities of the NAO. To assess the NAO sensitivity to the ENSO polarity, the indices were partitioned into two subsets: one for cold ENSO years and another for warm ENSO years. Then we computed the PDF for each index. The obtained PDFs for the total variability are statistically different at a significance level lower than 0.1%. It is also worth to note the difference in the shape of the PDFs: in warm ENSO years the PDF shows a bimodality while it is unimodal for the cold ENSO phase.

A t-test applied to the sample means showed that the NAO index has a higher mean during the cold ENSO phase. A F-test of the sampled variances showed that both the internal and total variabilities are statistically different for the two ENSO phases.

We cannot disregard the possibility that these results may be model dependent. However, the results are useful to illustrate a methodology to uncover a forced signal in a climate variable with high climate noise. As may be seen in Table 6.3, although the mean of total variability NAO indices is negative for warm ENSO years, the probability to get positive indices is very high (almost 40%). A similar comment applies to the cold ENSO years. Furthermore, for total variability, the probability of an index to get extreme values of the opposite sign of its mean is about 8 to 10%. These non-negligible probabilities help us to understand the difficulty to uncover the climate signal in a single realisation of the atmospheric circulation.

The results also show as the SST forcing may be found in the second and higher moments of internal variability. Indeed, in the present modelling experiment, we show that the internal NAO variances for cold and warm ENSO phases are different at a significance level lower than 1%. Figure 6.8 and 6.8 also suggests a bimodality on the internal variability in the warm ENSO phase but not in the cold ENSO phase.

We note that our results are consistent with some recent published works. First, the bimodality on the warm ENSO phase may contribute for a nonstationary relationship between the ENSO SST anomalies and the variability of North Atlantic climate [Walter and Graf, 2002; Sutton and Hodson, 2003]. We could hypothesise SSTs anomalies over other oceanic basins that may 'choose' between one of the variability modes (the positive or the negative maxima in the PDF). Note that it was shown in section 5.6.2 that interannual variability of SST anomalies in the tropical Atlantic forces the NAO in spring: a positive (negative) northward Atlantic Inter-Hemispheric SST anomaly Gradient (AISG) induces a negative (positive) phase of the NAO. Figure 6.9 shows the homogeneous correlation maps of the MUGCM first SVDA mode of tropical Atlantic SST anomalies and the EA MSLP anomalies in winter, accounting for 32.8% of total squared covariance fraction. The patterns suggest that the AISG forcing of the NAO in spring also occurs in winter. Furthermore, north tropical Atlantic SSTs are known to be forced by ENSO through the NAO itself playing the role of the so called *atmospheric bridge* [Klein et al., 1999; Saravanan and Chang, 2000; Giannini et al., 2001; Alexander

et al., 2002; Huang, 2004]. All together, these processes may interact contributing to the bimodality of the warm ENSO signal in the NAO. This issue is under current investigation.

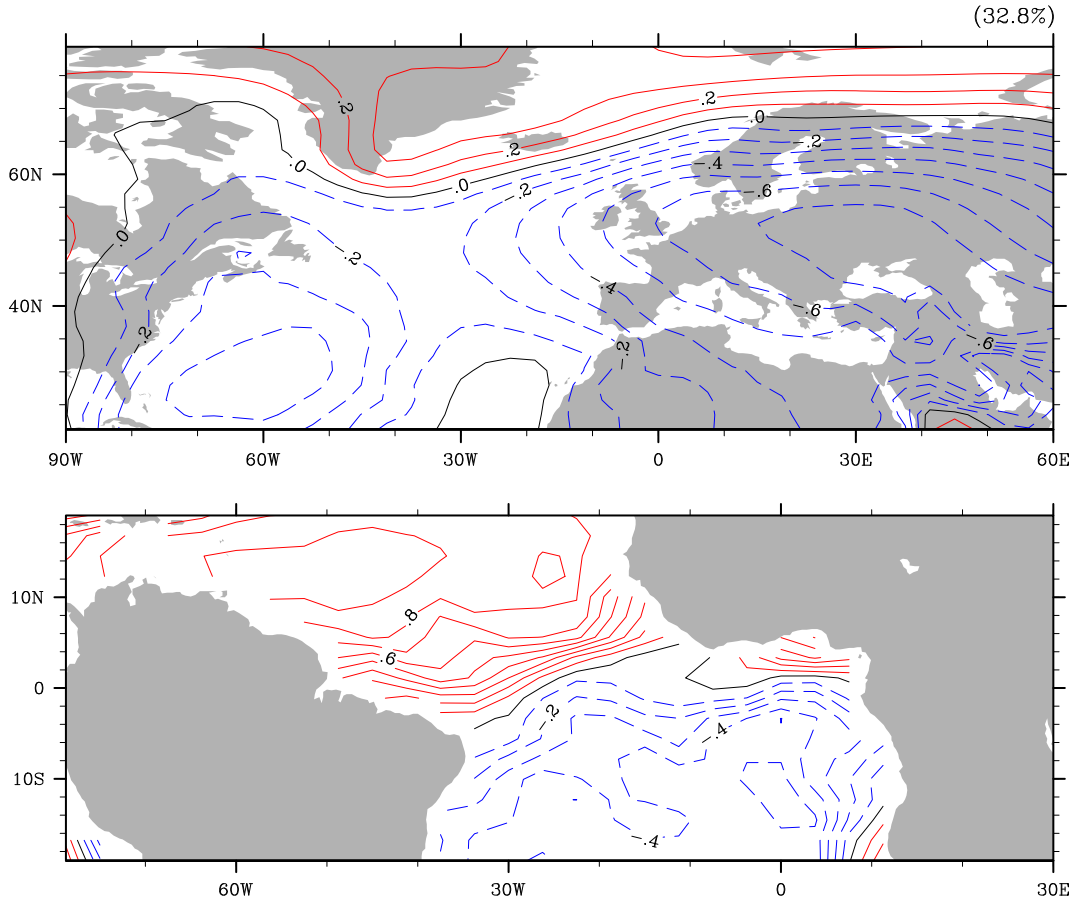

Figure 6.9: Homogenous correlation maps of the MUGCM first SVDA mode of tropical Atlantic SST anomalies (lower panel) and EA MSLP anomalies (upper panel) in winter (DJF). Negative, zero and positive correlations plotted with dashed blue, solid black and solid red isolines, respectively. Contour interval is 0.1. The percentage printed on the right-hand top of the figure is the squared covariance fraction.

.

The bimodality may also be important to understand the results of Pozo-Vásquez et al. [2001] and Cassou and Terray [2001]. These authors found a statistically significant SLP anomaly pattern resembling the NAO associated with the cold ENSO episodes, but no statistically significant pattern was found during the warm phase.

We also note that the bimodality is consistent with the work of Wu and Hsieh [2004]. They showed that the main component of the the Euro-Atlantic climate to ENSO is nonlinear.

A nonlinear component of atmospheric response to ENSO anomalies is also found over the Pacific-North American sector [Wu et al., 2003]. A major characteristic of the nonlinearity of the Pacific-North America climate response to ENSO is that there is an eastward phase shift of the circulation anomalies (by about 35°) between the composites of warm ENSO episodes and cold ENSO episodes [Hoerling et al., 1997]. Figure 6.10 shows this zonal shift reproduced by our simulations. The composites in this figure also show a positive (negative) NAO teleconnection associated with the cold (warm) ENSO phase. This is in agreement with the shift of the means derived from the PDF-statistics. If and how the zonal shift of the Pacific-North America wave train may impact on the cyclogenesis over the western North Atlantic [Fraedrich, 1994] contributing to a nonlinear response over the Atlantic-European sector is still an unsolved problem.

Finally, we note that the higher simulated NAO variability during warm ENSO episodes is consistent with the results of Gouirand and Moron [2003]. These authors found higher variance of observed MSLP anomalies over the North Atlantic during the warm ENSO phase.

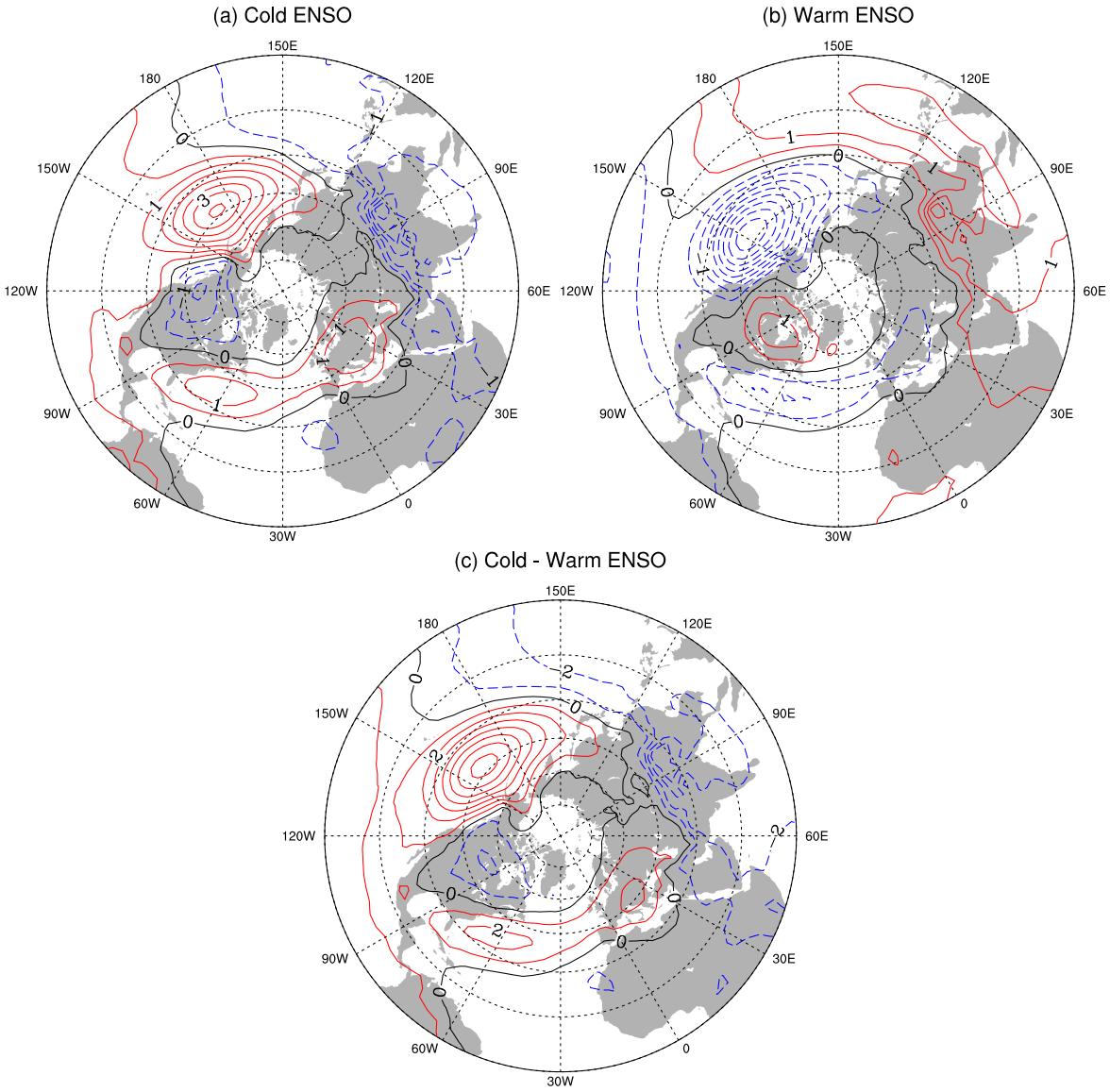

Figure 6.10: Composites of DJF seasonal mean MUGCM MSLP anomalies for the (a) cold and (b) warm ENSO phases, and (c) their difference (cold - warm ENSO). The contour interval is 0.5 mb in the composites for each phase, and 1 mb in the difference map.

Chapter 7

Summary and Conclusions

A large AMIP II ensemble of the MUGCM was used to assess SST anomaly forcing of MSLP seasonal mean anomalies over the EA region. Tropical SST forcing on interannual time scales was focused.

The validation of MUGCM's climatology and variability was performed using the reanalysis from the NCEP. The validation was performed using seasonal means of winter (DJF), spring (MAM), summer (JJA) and autumn (SON). The model climatology was assessed using: MSLP, TMP-SFC, PRECIP, HGT-850, HGT-500 and HGT-200 seasonal means in winter, spring, summer and autumn; vertical structures of zonal-mean zonal circulation and temperature in winter and summer; upper tropospheric zonal wind, in particular the jet streams, in winter and summer; planetary standing waves and eddy components (departures from zonal symmetry) in winter; and meridional and zonal mass overturning circulations in the Pacific and Atlantic (Hadley, Ferrel and Walker circulation) in winter.

Overall, the model is able to reproduce the main features of the observed seasonal climatology, but some important biases were detected. The pressure cells over the Atlantic ocean are overestimated in winter, resulting in the overestimation of the meridional MSLP gradient between the Azores High and the Icelandic Low. Accordingly, the low-tropospheric circulation in the EA sector was also found to be overestimated by

the model, as well as surface temperature and precipitation.

The modelled Pacific jet stream is biased 10° (30°) downstream in winter (summer), is displaced 10° to the north in both seasons, and is 15 m s^{-1} stronger in winter. The Atlantic jet stream is biased 20° downstream (10° upstream) in winter (summer), is shifted 10° (5°) to the north in winter (summer), and is 10 m s^{-1} stronger in summer.

The modelled meridional mass circulation is overestimated. In particular, the thermally direct Hadley cells are too intense, which may have important consequences on the extratropical atmospheric response to tropical SST forcing. The Pacific jet stream overestimation may be a consequence of the intensified Pacific Hadley cells. It was also shown that the Walker cell in the equatorial Pacific and Atlantic are also overestimated by the model. The overestimation of the tropical atmospheric response to SST climatology indicates that its response to SST anomaly variability may also be overestimated. In conjuction with the overestimation of the mean state of the Pacific jet stream, this may lead to an overestimation of the extratropical atmospheric response to tropical Pacific SST anomalies, in particular, a stronger ENSO forcing in the NPA and EA regions.

Modelled and observed variability was tested using two different methodologies. First, the correlation at each grid point of the MUGCM's grid in the EA sector, between the MUGCM-AMIP II ensemble mean and the NCEP MSLP, TMP-SFC, PRECIP, HGT-850, HGT-500 and HGT-200, was determined for each of the four standard seasons. Second, the spatial patterns of variability of MUGCM and NCEP seasonal mean anomalies of MSLP, in the EA region, obtained by PCA, were compared.

The correlations between modelled and observed grid point time series are significantly different from zero, at a significance level of 1%, in almost the entire area of the EA region.

PCA of seasonal mean MSLP in the EA region captured the major observed modes of MSLP anomaly variability in the EA region, namely, the North Atlantic Oscillation, the East Atlantic mode, the Scandinavia or EURASIA I mode, and the East Atlantic/Western Russia or EURASIA II mode. These modes are better captured by the model in winter and spring.

The NAO modelled pattern has the characteristic meridional see-saw in the atmospheric mass between the Icelandic Low and a broad east-west belt, centred near 40^oN, extending from the east coast of the United States to the Mediterranean. However, the southern the southern belt of NAO pattern has its highest loadings shifted to the east by about 15^o in winter, and is broken in two centres of equal sign located in western Atlantic and Europe in spring.

An ANOVA showed that seasonal mean MSLP is significantly SST-forced in almost the entire area of the EA, except in SON, the higher forcing occurring in winter and spring.

The ANOVA of the time series associated to the four modes of variability, identified by the PCA, in winter and spring, showed that these are significantly forced in both seasons. The NAO was found to be the mode with strongest forcing in spring (30.0%).

The ANOVA was also performed to the NAO index in overlapping 3-month mean seasons and it was found that the NAO is significantly forced in all seasons except in autumn and OND, the higher forcing found again in spring.

In summary, the MUGCM MSLP variability in the EA region, in general, and its major mode of variability (the NAO), in particular, have a weak but significant SST-contribution, the forcing being strong in winter and specially in spring.

The two leading SST-forced variability modes of MSLP anomalies in the EA region, in winter and spring, were estimated using three different methods: PCA of the ensemble-mean (standard PCA), SVDA and optimal detection PCA (optimal PCA).

Regression and correlation analysis using SST anomalies and the time series associated to the forced patterns, gave statistical significant evidence that

- SST anomalies in the tropical Pacific associated to the ENSO forces the NAO in winter and spring. A warm (cold) ENSO phase induces a negative (positive) phase of the NAO, specially in winter; and
- SST anomalies in the tropical Atlantic associated to the AISG forces the NAO in spring. A positive (negative) phase of the AISG induces a negative (positive) phase of the NAO.

SVDA analysis of tropical Pacific monthly mean SST anomalies and EA monthly mean MSLP anomalies, from December to May, showed that the pattern of the MSLP correlation map projects onto the negative phase of the NAO in all months, which confirms the results given above. Unlike the observational results of other authors, our modelled ENSO teleconnection on the EA region does not changes sign between December and January.

A large ensemble of two NAO indices simulated by the MUGCM were constructed to represent the internal and total variabilities of the NAO. To assess the NAO sensitivity to the ENSO polarity, the indices were partitioned into two subsets: one for cold ENSO years and another for warm ENSO years. Then we computed the PDF for each index. The obtained PDFs for the total variability are statistically different at a significance level lower than 0.1%. It is also worth to note the difference in the shape of the PDFs: in warm ENSO years the PDF shows a bimodality while it is unimodal for the cold ENSO phase.

A t-test applied to the sample means showed that the NAO index has a higher mean during the cold ENSO phase. A F-test of the sampled variances showed that both the internal and total variabilities are statistically different for the two ENSO phases.

Using composites of the NAO indices associated to the cold and warm ENSO phases, it was shown that the ENSO forces the NAO through the PNA wave train.

We cannot disregard the possibility that these results may be model dependent. However, the results are useful to illustrate a methodology to uncover a forced signal in a climate variable with high climate noise. Although the mean of total variability NAO indices is negative for warm ENSO years, the probability to get positive indices is very high (almost 40%). A similar comment applies to the cold ENSO years. Furthermore, for total variability, the probability of an index to get extreme values of the opposite sign of its mean is about 8 to 10%. These non-negligible probabilities help us to understand the difficulty to uncover the climate signal in a single realisation of the atmospheric circulation.

The results also showed that ENSO SST forcing is found in the second and higher moments of internal variability. Indeed, in the present modelling experiment, we showed that the internal NAO variances for cold and warm ENSO phases are different at a significance level lower than 1%. The PDFs suggests a bimodality on both internal and forced variabilities in the warm ENSO phase but not in the cold ENSO phase. This intriguing result shows that internal variability is not independent from SST forced variability. This paradox shows that the ANOVA model used to estimate potential predictability is inaccurate.

The bimodality of the NAO found during the warm ENSO phase could be explained hypothesising SSTs anomalies over other oceanic basins that may 'choose' between one of the variability modes (the positive or the negative maxima in the PDF). It was suggested that the AISG forcing of the NAO in spring also occurs in winter. Furthermore, north tropical Atlantic SSTs are known to be forced by ENSO through the NAO itself playing the role of the so called *atmospheric bridge*. The zonal shift of the PNA wave train may also impact the cyclogenesis over the western North Atlantic. All together, these processes may interact, contributing to a nonlinear response over the Atlantic-European sector and inducing a bimodality of the warm ENSO signal in the NAO. This issue is still an unsolved problem and is under current investigation.

Appendix A

Principal Component Analysis

Here brief summary of the mathematical formulation of Principal Component (EOF) Analysis is presented. The EOF analysis is a multivariate statistical technique generally known as Principal Component Analysis (PCA). The name *Empirical Orthogonal Function* is due to Lorenz [1956] and is the the more popular term in the geophysical sciences. Book-length descriptions of PCA can be found in Jolliffe [1986], in a general context, and in Preisendorfer [1988], in the geophysical context. It merits a chapter in all multivariate statistical analysis books such as Johnson and Wichern [1998], just to name one, and also in statistics books oriented to the atmospheric sciences such as von Storch and Zwiers [1999], with a formal approach, and Wilks [1995] with a more descriptive approach, but both rich in real world examples. Geophysicists may also find a very short and simple introduction to PCA in the papers of Jolliffe [1990, 1993].

Principal Vectors Consider that the random variable X_m represents the seasonal mean anomaly of some variable at spatial location m, with $m = 1, ..., M$. Remember that seasonal mean anomaly is the deviation of the seasonal mean variable from the season climatology, which implies that $\mathcal{E}\big[X_m\big] = 0$, where $\mathcal{E}$ is the expectation operator. These M random variables can be assembled in a $(M \times 1)$-dimensional random vector $\mathbf{X} = \big[X_1|X_2|...|X_M\big]^{\mathrm{T}}$, where the upperscipt T denotes the transpose operator. The variance-covariance matrix of $\mathbf{X}$,

$$\mathbf{\Sigma}_X = \mathcal{E}\big[\mathbf{X}\mathbf{X}^{\mathrm{T}}\big], \tag{A.1}$$

where $\mathcal{E}$ denotes the expectation operator, is a $(M \times M)$ simmetric matrix that assumes M non-negative eigenvalues and M orthogonal eigenvectors. The first principal vector (PV) of $\mathbf{X}$, $\mathbf{e}_1$, is defined as the normalized eigenvector of $\mathbf{\Sigma}_X$ with the largest eigenvalue, λ_1, the second PV, $\mathbf{e}_2$, is the normalized eigenvector with the second largest eigenvalue, λ_2, and so on.

Each PV is $(M \times 1)$ dimensioned, that is, it has M elements which are called *loadings*. Note that the mth loading of the kth PV, e_{mk}, is associated with the mth spatial location. Consequently, each PV is a field of loadings that can be plotted with smooth contours. This graphical representation of an PV is called the *PV pattern*.

An important property of the PVs is that they diagonalize the covariance matrix $\mathbf{\Sigma}_X$. This can be seen by first grouping the PVs into the $(M \times M)$ matrix $\mathbf{E} = \left[\mathbf{e}_1|\mathbf{e}_2|...|\mathbf{e}_M\right]$:

$$\mathbf{E}^{-1}\mathbf{\Sigma}_X\mathbf{E} = \mathbf{E}^{\mathrm{T}}\mathbf{\Sigma}_X\mathbf{E} = \Lambda, \tag{A.2}$$

where Λ is a diagonal matrix whose entries (1,1), (2,2), ..., (M,M) are the eigenvalues $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_M \geq 0$. Furthermore, noting that $\mathbf{e}_k^{\mathrm{T}}\mathbf{\Sigma}_X\mathbf{e}_k$ represents the variance of the random vector $\mathbf{X}$ in the direction defined by kth PV, $\mathbf{e}_k$, equation A.2 reveals that the first PV defines the direction in which the variance of $\mathbf{X}$ is maximized, the second PV defines the direction in which the variance of $\mathbf{X}$ is maximized subjected to the constraint that it must be orthogonal with direction defined by the first PV, and so on.

An useful equation equivelent to equation A.2 is

$$\mathbf{\Sigma}_X = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^{\mathrm{T}}, \tag{A.3}$$

which is a *singular value decomposition* of the covariance matrix.

Principal Components The PVs determine a new M-dimensional space in which we can view the data. Projecting the $(M \times 1)$-dimensional random vector $\mathbf{X}$ onto the $(M \times M)$ matrix of the PVs we obtain another $(M \times 1)$-dimensional random vector

$\mathbf{Y} = \left[Y_1|Y_2|...|Y_M\right]^{\mathrm{T}}$,

$$\mathbf{Y} = \mathbf{E}^{\mathrm{T}}\mathbf{X}, \tag{A.4}$$

whose components are called the *principal components* (PCs) because they are the components (coordinates) of $\mathbf{X}$ in the basis of the principal vectors. The variance-covariance matrix of $\mathbf{Y}$, $\mathbf{\Sigma}_{\mathrm{Y}} = \mathcal{E}\left[\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\right]$ may be written, using equation A.4, as $\mathbf{\Sigma}_{\mathrm{Y}} = \mathcal{E}\left[\mathbf{E}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{E}\right]$, and, using equations A.1 and A.2, we have

$$\mathbf{\Sigma}_{\mathrm{Y}} = \mathbf{E}^{\mathrm{T}}\mathbf{\Sigma}_{\mathrm{X}}\mathbf{E} \tag{A.5}$$

$$= \mathbf{\Lambda}. \tag{A.6}$$

Equation A.5 shows that the variance of the kth PC of Y, $\mathcal{E}\left[Y_k Y_k^{\mathrm{T}}\right]$, is equal to the variance of $\mathbf{X}$ in the direction defined by the k PV, $\mathbf{e}_k^{\mathrm{T}}\mathbf{\Sigma}_{\mathrm{X}}\mathbf{e}_k$. This must be so, because Y_k is the projection of $\mathbf{X}$ onto the kth PV, $Y_k = \mathbf{e}_k^{\mathrm{T}}\mathbf{X}$. Equation A.6 reveals two very important properties of the $\mathbf{Y}$ random vector. The first is that its components (PCs) are uncorrelated and the second one is that the variance of the kth PC is the kth eigenvalue of $\mathbf{\Sigma}_{\mathrm{X}}$. Furthemore, since $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_M$, $Var\left[Y_1\right] \geq Var\left[Y_2\right] \geq ... \geq Var\left[Y_M\right]$.

By Equation A.4, the kth PC, Y_k, may be seen as a linear combination of the components of $\mathbf{X}$ at the M spatial locations, $X_1, X_2, ..., X_M$. The weight at the mth spatial location on the kth linear combination is the kth PV loading at mth location, e_{mk}. Thus, the stations with higher PV loadings contribute more to the PC. Furthermore, the weights of the first linear combination - the first PV loadings - are such that maximize the variance of Y_1, the weights of the second linear combination- the second PV loadings - are such that maximize the variace of Y_2 subjected to the constraint that the second PV must be orthogonal to the first PV, and so on.

A final property remains to be exposed: the total variance of $\mathbf{X}$ is equal to the sum of the eigenvalues of $\mathbf{\Sigma}_{\mathrm{X}}$ which is equal to the total variance of $\mathbf{Y}$. In fact, from equation A.3 we have that the total variance of $\mathbf{X}$ is $\sum_{m=1}^{M}\mathcal{E}\left[X_m X_m\right] = tr\left(\mathbf{\Sigma}_{\mathrm{X}}\right) = tr\left(\mathbf{E}\mathbf{\Lambda}\mathbf{E}^{\mathrm{T}}\right)$, where tr is the trace matrix operator, that is the sum of the diagonal elements of the operand matrix. Then, noting that $tr\left(\mathbf{E}\mathbf{\Lambda}\mathbf{E}^{\mathrm{T}}\right) = tr\left(\mathbf{\Lambda}\right)$ and using

equation A.6, we obtain

$$\sum_{m=1}^{M} \mathcal{E}\big[X_m X_m\big] = tr\big(\mathbf{\Lambda}\big) = \sum_{k=1}^{M} \mathcal{E}\big[Y_k Y_k\big]. \tag{A.7}$$

Finally, from equations A.6 and A.7 we obtain that proportion of the total population variance due to the kth PC is $\lambda_k / tr(\mathbf{\Lambda})$.

Briefly, PCA can be described as follows: Given a $M \times 1$-dimensional random vector $\mathbf{X} = \big[X_1 | X_2 | ... | X_M\big]^{\mathrm{T}}$, whose components are correlated with each other, the PCA analysis finds an orthonormal basis of M $(M \times 1)$-dimensional vectors - the PVs - that defines the directions of maximum variability of $\mathbf{X}$ subjected to the constraint that they must be orthogonal, that is, the components of $\mathbf{X}$ on this new basis - the PCs - have the maximum possible variance and are uncorrelated with each other.

Estimation of Principal Vectors: the EOFs The above presentation of PCA was based on the assumption that the covariance matrix of the random vector $\mathbf{X}$ is known. In practice, the probability distribution of $\mathbf{X}$ is unkown and, consequently, the covariance matrix $\mathbf{\Sigma}_{\mathrm{X}}$ is also unkown. Fortunately, it can be estimated recurring to the sample theory.

Let $\big(X_{m1}, X_{m2}, ..., X_{mN}\big)$ represent a random sample of the random variable $\mathbf{X}_m$. One realization of this random sample would be $\big(x_{m1}, x_{m2}, ..., x_{mN}\big)$, that is, N realizations (or observations) of the seasonal mean anomaly at station m. Since there are M spatial locations, we have M random samples, of length N, of $\mathbf{X}$, which we can cluster in a $(M \times N)$ random matrix $\mathcal{X}$ whose mth row $\big[X_{m1}, X_{m2}, ..., X_{mN}\big]$ is the mth random sample, of length N, of the random variable X_m.

The covariance matrix Σ_{X} of the random vector $\mathbf{X}$ can then be estimated by the sample covariance matrix of $\mathcal{X}$

$$\mathbf{S}_{\mathrm{X}} = \frac{1}{N-1} \mathcal{X} \mathcal{X}^{\mathrm{T}}. \tag{A.8}$$

Since $\mathbf{S}_{\mathrm{X}}$ is an estimator of $\mathbf{\Sigma}_{\mathrm{X}}$, the eigenvalue-eigenvector pairs computed from it will also be estimators of the eigenvalue-eigenvector pairs of $\mathbf{\Sigma}_{\mathrm{X}}$. Furthermore, the projection of the N realizations of $\mathbf{X}$ onto these eigenvectors will yield N realizations of $\mathbf{Y}$:

$$\mathcal{Y} = \hat{\mathbf{E}}^{\mathrm{T}}\mathcal{X}. \tag{A.9}$$

where $\mathcal{Y}$ is a $(M \times N)$ matrix whose N columns are N observations of the M PCs. The element y_{mn} of $\mathcal{Y}$ is the nth observation of the mth PC. The elements y_{mn} $(n = 1, ..., N)$ are called the *scores* of the mth PC.

The principal vectors of the *sample* covariance matrix $\mathbf{S}_{\mathrm{X}}$ are generally called *sample* principal vectors, but the term Empirical Orthogonal Functions (EOFs) is preferred to the geophysical community, which also prefers to call PCA by EOF analysis, as already said. Some authors, such as von Storch and Zwiers [1999] use the term EOF to name the eigenvectors of both $\mathbf{\Sigma}_{\mathrm{X}}$ and $\mathbf{S}_{\mathrm{X}}$, noticing that the latter are estimators of the former. However, as pointed out by Wilks [1995], the term *empirical* means that the *orthogonal functions* are "defined *empirically* according to the particular data set at hand", a realization of $\mathcal{X}$. Thus, it makes no sense to refer to the eigenvectors of $\mathbf{\Sigma}_{\mathrm{X}}$ as EOFs, because they are not empirically defined. They are parameters that characterize the covariance matrix which is a populational parameter (not a sample parameter) of the random vector $\mathbf{X}$.

An important property of the EOFs of a $(M \times N)$ data matrix $\mathcal{X}$ is that only the first $\min(M, N)$ EOFs will have non-zero eigenvalues. Since the variance of the kth $(k = 1, ..., M)$ PC is equal to the kth eigenvalue (equation A.6), only the EOFs with non-zero eigenvalues, the first $\min(M, N)$ EOFs, will be meaningful.

Finally, we present a relation that will be useful in chapter 5 (equation 5.44). Multiplying both sides of equation A.9 by $\mathcal{X}^{\mathrm{T}}$, using equations A.8 and the sample version of equation A.3, and the orthonormality of $\hat{\mathbf{E}}$ we obtain

$$\mathcal{Y}\mathcal{X}^{\mathrm{T}} \;=\; (N-1)\hat{\mathbf{\Lambda}}\hat{\mathbf{E}}^{\mathrm{T}}.$$

Now, multiplying this equation by the pseudo-inverse of $\hat{\mathbf{\Lambda}}$, noting that $\hat{\mathbf{\Lambda}}^{-1}\hat{\mathbf{\Lambda}} = \mathbf{I}_{(K)}$, where $\mathbf{I}_{(K)}$, with $k = \min(M, N)$, is a $(M \times M)$ matrix with entries (i, j) equal to unit if $i = j < K$ and zero elsewhere, we have

$$\hat{\mathbf{E}}_{(K)} = \frac{1}{N-1}\mathcal{X}\mathcal{Y}^{\mathrm{T}}\hat{\mathbf{\Lambda}}^{-1} \tag{A.10}$$

where $\hat{\mathbf{E}}_{(K)} = \hat{\mathbf{E}}\mathbf{I}_{(K)}$ is a $(M \times M)$ matrix whose first K columns are the first K columns of $\mathbf{E}$, with the rest $M - K$ columns filled with zeros, and $\hat{\mathbf{\Lambda}}^{-1}\mathcal{Y}$ contains the standardized PCs. This equation is equivelent to regressing $\mathcal{X}$ onto the PCs.

Bibliography

Alexander, M. A., Blandé, I., Newman, M., Lanzante, J. R., Lau, N.-C., 2002. The atmospheric bridge: the influence of ENSO teleconnections on air-sea interaction over the global oceans. J. Clim. 15, 2205–2231.

Allen, M. R., Smith, L. A., 1997. Optimal filtering in singular spectrum analysis. Phys. Lett. 234, 419–428.

Alves, I., Rocha, A., 2003. Modelling the influence of climate change on precipitation and temperature in Western Mediterranean. Livro de Resumos do 3^o Simpósio de Meteorologia e Geofísica da Associação Portuguesa de Meteorologia e Geofísica, 20 pp.

Apostol, T. M., 1969. Calculus, Second Edition. Vol. II. John Wiley & Sons, 673 pp.

Barnett, T. P., 1981. Statistical relations between Ocean/Atmosphere fluctuations in the tropical Pacific. J. Phys. Oceanogr. 1, 1043–1058.

Barnett, T. P., 1985. Variations in near global sea level pressure. J. Atmos. Sci. 42, 478–501.

Barnett, T. P., 1993. ENSO and ENSO-related predictability. Part I: Prediction of equatorial Pacific sea surface temperature with a hybrid coupled ocean-atmosphere model. J. Clim. 6, 1545–1566.

Barnett, T. P., Arpe, K., Bengtsson, L., Ji, M., Kumar, A., 1997. Potential predictability and AMIP implications of midlatitude climate variability in two general circulation models. J. Clim. 10, 2321–2329.

Barnston, A., 1994. Linear statistical short-term climate predictive skill in the northern hemisphere. J. Clim. 7, 1513–1564.

Barnston, A., Livezey, R. E., 1987. Classification, seasonality and persistence of low-circulation patterns. Mon. Wea. Rev. 115, 1083–1126.

Bjerknes, J., 1969. Atmospheric teleconnections from the equatorial Pacific. Mon. Wea. Rev. 97, 162–172.

Bjerknes, J., 1972. Large-scale atmospheric response to 1964-65 Pacific equatorial warming. J. Phys. Oceanogr. 2, 212–217.

Blackmon, M. L., Lee, Y.-H., Wallace, J. M., 1984. Horizonal structure of 500 mb height fluctuations with long, intermediate and short time scales. J. Atmos. Sci. 41, 961–979.

Boer, G. J., Arpe, K., Blackburn, M., Déqué, M., Gates, T. L., Hart, T. L., le Treut, H., Roeckner, E., Sheinin, D. A., Simmonds, I., Smith, R. N. B., Tokioka, T., Wetherald, R. T., Williamson, D., 1991. An intercomparison of the climates simulated by 14 atmospheric general circulation models. Tech. rep., World Meteorological Organization/International Council of Scientific Unions World Climate Research Programme, Report No. 15 WMO/TD - No. 425, 37 pp plus maps.

Boer, G. J., Arpe, K., Blackburn, M., Déqué, M., Gates, T. L., Hart, T. L., le Treut, H., Roeckner, E., Sheinin, D. A., Simmonds, I., Smith, R. N. B., Tokioka, T., Wetherald, R. T., Williamson, D., 1992. Some results from an intercomparison of the climates simulated by 14 atmospheric general circulation models. J. Geophys. Res 97(D12), 12771–12786.

Bourke, W., McAvaney, B., Thurling, R., 1977. Global modelling of atmospheric flow by spectral models. Methods in Comp. Phys. 17, General Circulation Models of the Atmosphere. Academic Press, New York, 267-324.

Branković, C., Palmer, T. N., Ferranti, L., 1994. Predictability of seasonal atmospheric variations. J. Clim. 7, 217–237.

Bretherton, C. S., Smith, C., Wallace, J. M., 1992. An intercomparison of methods for finding coupled patterns in climate data. J. Clim. 5, 541–560.

Cassou, C., Terray, L., 2001. Oceanic forcing of wintertime low-frequency atmospheric variability in the North Atlantic European sector: a study with the ARPEGE model. J. Clim. 14, 4266–4291.

Cassou, C., Terray, L., 2004. North Atlantic winter climate regimes: spatial asymmetry, stationarity with time, and oceanic forcing. J. Clim. 17, 1055–1068.

Castanheira, J. M., 2000. Climatic variability of the atmospheric circulation at the global scale. Ph.D. Thesis. University of Aveiro, Portugal, 186 pp.

Chang, P., Saravanan, R., Ji, L., Hegerl, G. C., 2000. The effect of local sea surface temperatures on atmospheric circulation over the tropical Atlantic sector. J. Clim. 13, 2195–2216.

Charney, J. G., Eliassen, A., 1949. A numerical method for predicting the perturbations of the middle latitude westerlies. Tellus 1, 38–54.

Chen, W. Y., 1982. Assessment of Southern Oscillation sea-level pressure indices. Mon. Wea. Rev. 110, 800–807.

Chen, W. Y., 1989. Estimates of dynamical predictability from NMC DERF experiments. Mon. Wea. Rev. 117, 1227–1236.

Chen, W. Y., Van den Dool, H. M., 1997. Atmospheric predictability of seasonal, annual, and decadal climate means and the role of the ENSO cycle: a model study. J. Clim. 10, 1236–1254.

Cherry, S., 1996. Singular Value Decomposition and Canonical Correlation analysis. J. Clim. 9, 2003–2009.

Chervin, R. M., 1986. Interannual variability and seasonal climate predictability. J. Atmos. Sci. 43, 233–251.

Chervin, R. M., Schneider, S. H., 1976. On determining the statistical significance of climate experiments with General Circulation Models. J. Atmos. Sci. 33, 405–412.

Collins, M., Frame, D., Sinha, B., Wilson, C., 2002. How far can we predict El Niño? Geophys. Res. Lett. 29, 131–134.

Conover, W. J., 1971. Practical nonparametric statistics. John Wiley & Sons, 462 pp.

Coppola, E., Kucharski, F., Giorgi, F., Molteni, F., 2005. Bimodality of the North Atlantic Oscillation in simulations with greenhouse gas forcing. Geophys. Res. Lett. 32, L23709.

Czaja, A., Frankignoul, C., 2002. Observed impact of Atlantic SST anomalies on the North Atlantic Oscillation. J. Clim. 15, 606–623.

Davis, J. R., Rowell, D. P., Folland, C. K., 1997. North Atlantic and European seasonal predictability using an ensemble of multidecadal atmospheric GCM simulations. Int. J. Climatol. 17, 1263–1284.

Derome, J., Wiin-Nielsen, A., 1971. The response of a middle-latitude model atmosphere to forcing by topography and stationary heat sources. Mon. Wea. Rev. 99, 564–576.

Deser, C., Blackmon, M. L., 1993. Surface climate variations over the North Atlantic Ocean during winter: 1900-1989. J. Clim. 6, 1743–1753.

Dickson, R. R., Namias, J., 1976. North American influences on the circulation and climate of the North Atlantic sector. Mon. Wea. Rev. 104, 1255–1265.

Dix, M. R., Hunt, B. G., 1995. Chaotic influences and the problem of deterministic seasonal predictions. Int. J. Climatol. 15, 729–752.

Dong, B.-W., Sutton, R. T., Jewson, S. P., O'Neil, A., Slingo, J. M., 2000. Predictable winter climate in the north Atlantic sector during the 1997-99 ENSO cycle. Geophys. Res. Lett. 27, 985–988.

Feldstein, S. B., 2000. The timescale, power spectra, and climate noise properties of teleconnection patterns. J. Clim. 13, 4430–4440.

Fraedrich, K., 1994. An ENSO impact in Europe? Tellus 46A, 541–552.

Frankignoul, C., Friederichs, P., Kestenare, E., 2003. Influence of Atlantic SST anomalies on the atmospheric circulation in the Atlantic-European sector. Annal. Geophys. 46, 71–85.

Gates, W. L., 1992. AMIP: The Atmospheric Model Intercomparison Project. Bull. Amer. Meteor. Soc. 73, 1962–1970.

Gates, W. L., Boyleand, J., Coveyand, C., Deaseand, C., Doutriauxand, C., Drachand, R., Fiorinoand, M., Glecklerand, P., Hniloand, J., Marlaisand, S., Phillipsand, T., Potterand, G., Santerand, B., Sperberand, K., Taylor, K., Williams, D., 1998. An overview of the results of the Atmospheric Model Intercomparison Project (AMIP). Bull. Amer. Meteor. Soc. 80, 29–55.

Giannini, A., Chiang, J. C. H., Cane, M. A., Kushnir, Y., Seager, R., 2001. The ENSO teleconnection to the tropical Atlantic Ocean: contributions of the remote and local SSTs to rainfall variability in the tropical Americas. J. Clim. 14, 4530–4544.

Glowienka-Hense, R., 1990. The North Atlantic Oscillation in the Atlantic-European SLP. Tellus 42A, 497–507.

Gouirand, I., Moron, V., 2003. Variability of the impact of El Niño-Southern Oscillation on sea-level pressure anomalies over the North Atlantic in January to March (1874-1996). Int. J. Climatol. 23, 1549–1566.

Harzallah, A., Sadourny, R., 1995. Internal versus SST-forced atmospheric variability as simulated by an Atmospheric General Circulation Model. J. Clim. 8, 474–495.

Hasselmann, K., 1979. On the signal-to-noise problem in atmospheric response studies. Meteorology over the Tropical Oceans, D. B. Shaw, Ed., Royal Meteorological Society, 251–259.

Hastenrath, S., 2001. In search of zonal circulations in the equatorial Atlantic sector from the NCEP-NCAR reanalysis. Int. J. Climatol. 21, 37–47.

Hastenrath, S., Lamb, P. J., 2004. Climate dynamics of the atmosphere and ocean in the equatorial zone: a synthesis. Int. J. Climatol. 24, 1601–1612.

Hodson, D., Sutton, R., Pohlmann, H., Rodwell, M., Stendel, M., Terray, L., 2003. Influence of the oceans on north Atlantic climate variability: a comparison of results from four atmospheric GCMs. Geophys. Res. Abs. 5, 5592.

Hoerling, H. P., Kumar, A., Zhong, M., 1997. El Niño, La Niña and the nonlinearity of their teleconnections. J. Clim. 10, 1769–1786.

Holton, J. R., 1992. An introduction to Dynamic Meteorology, Third Edition. International Geophysics Series, Vol. 48. Academic Press, 507 pp.

Horel, J. D., 1981. A Rotated Principal Component Analysis of interannual variability of northern hemisphere 500 mb height field. Mon. Wea. Rev. 109, 2080–2092.

Horel, J. D., Wallace, J. M., 1981. Planetary-scale atmospheric phenomena associated with the Southern Oscillation. Mon. Wea. Rev. 109, 813–821.

Hoskins, B. J., Karoly, D. J., 1981. The steady linear response of a spherical atmosphere to thermal and orographic forcing. J. Atmos. Sci. 38, 1179–1196.

Huang, J., 2004. Remotely forced variability in the tropical Atlantic ocean. Clim. Dyn. 23, 133–152.

Huang, J., Higuchi, K., Shabbar, A., 1998. The relationship between the North Atlantic Oscillation and El Niño-Southern Oscillation. Geophys. Res. Lett. 25, 2707–2710.

Hurrel, J. W., 1995. Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. Science 269, 676–679.

Johns, T. C., Carnell, R. E., Crossley, J. F., Gregory, J. M., Mitchell, J. F. B., Senior, C. A., Tett, S. F. B., R. A, W., 1997. The second Hadley Centre coupled ocean-atmosphere GCM: model description, spinup and validation. Clim. Dyn. 13, 103–134.

Johnson, R. A., Wichern, D. W., 1998. Applied Multivariate Statistical Analysis, Fourth Edition. Prentice-Hall, New Jersey, 816 pp.

Jolliffe, I. T., 1986. Principal Component Analysis. Springer, New York, 271 pp.

Jolliffe, I. T., 1990. Principal Component Analysis: A beginner's guide - I. Introduction and application. Weather 45, 375–382.

Jolliffe, I. T., 1993. Principal Component Analysis: A beginner's guide - II. Pitfalls, myths and extensions. Weather 48, 246–253.

Julian, P. R., Chervin, R. M., 1978. A study of the Southern Oscillation and Walker circulation phenomenon. Mon. Wea. Rev. 106, 1433–1451.

Kalnay, E., co authors, 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Amer. Meteor. Soc. 77, 473–471.

Kiladis, G. N., Diaz, H. F., 1989. Global climatic anomalies associated with extremes in the Southern Oscillation. J. Clim. 2, 1069–1090.

Klein, S. A., Soden, B. J., Lau, N.-C., 1999. Remote sea surface temperature variations during ENSO: evidence for a tropical atmospheric bridge. J. Clim. 12, 917–932.

Kumar, A., Hoerling, M., 1995. Prospects and limitations of seasonal atmospheric GCM predictions. Bull. Amer. Meteor. Soc. 76, 335–345.

Kumar, A., Hoerling, M., Ji, M., Leetmaa, A., Sardeshmukh, P., 1996. Assessing a GCM's suitability for making seasonal predictions. J. Clim. 9, 115–129.

Latif, M., Arpe, K., Roeckner, E., 2000. Oceanic control of decadal north Atlantic sea level pressure variability in winter. Geophys. Res. Lett. 27, 727–730.

Lau, K.-M., Lim, H., 1984. On the dynamics of equatorial forcing of climate teleconnections. J. Atmos. Sci. 41, 161–176.

Lau, N.-C., 1997. Interactions between global SST anomalies and the midlatitude atmospheric circulation. Bull. Amer. Meteor. Soc. 78, 21–33.

Lau, N.-C., Nath, M. J., 1994. A modeling study of the relative roles of tropical and extratropical SST anomalies in the variability of the Global Atmosphere-Ocean system. J. Clim. 7, 1184–1207.

Leith, C. E., 1973. The standard error of time-averaged estimates of climatic means. J. Appl. Meteor. 12, 1066–1069.

Lin, H., Derome, J., 2002. Tropical links of the Arctic Oscillation. Geophys. Res. Lett. 29, 1943.

Lin, H., Derome, J., 2004. Nonlinearity of the extratropical response to tropical forcing. J. Clim. 17, 2597–2608.

Lin, H., Derome, J., Brunet, G., 2005. Tropical Pacific link to the two dominant patterns of atmospheric. Geophys. Res. Lett. 32, L03801.

Lorenz, E. N., 1956. Empirical Orthogonal Functions and Statistical Weather Predictions. Sci. Rep. 1. Statistical Forecasting Project, Department of Meteorology, MIT, (NTIS AD 110268), 49 pp.

Lorenz, E. N., 1963. Deterministic nonperiodic flow. J. Atmos. Sci. 20, 130–141.

Lorenz, E. N., 1965. A study of the predictability of a 28-variable atmospheric model. Tellus 17, 321–333.

Lorenz, E. N., 1982. Atmospheric predictability experiments with a large numerical model. Tellus 34, 505–513.

Madden, R. A., 1976. Estimates of Natural Variability of time-averaged sea-level pressure. Mon. Wea. Rev 104, 942–952.

Manabe, S., Terpstra, B., 1974. The effects of mountains on the general circulation of the atmosphere as identified by numerical experiments. J. Atmos. Sci. 31, 3–42.

Mathieu, P.-P., Sutton, R. T., Dong, B., Collins, M., 2004. Predictability of winter climate over the North Atlantic European region during ENSO events. J. Clim. 17, 1953–1974.

McAvaney, B. J., Bourke, W., Puri, K., 1978. A global spectral model for simulation of the general circulation. J. Atmos. Sci. 35, 1557–1583.

McPhaden, M. J., Busalacchi, A. J., Cheney, R., Donguy, J.-R., Gage, K. S., Halpern, D., Ji, M., Julian, P., Meyers, G., Mitchum, G. T., Niiler, P. P., Picaut, J., Reynolds, R. W., Smith, N., Takeuchi, K., 1998. The Tropical Ocean-Global Atmosphere observing system: A decade of progress. J. Geophys. Res. 103, 14169–14240.

Melo-Gonçalves, P., Rocha, A. C., Castanheira, J. M., Ferreira, J. A., 2005. North Atlantic Oscillation sensitivity to El Niño/Southern Oscillation polarity in a large-ensemble simulation. Clim. Dyn. 24, 599–606.

Merkel, U., Latif, M., 2001. A high resolution AGCM study of the El Niño impact on the North Atlantic/European sector. Geophys. Res. Lett. 29 (9), 1291.

Milton, J. S., Arnold, J. C., 1995. Introduction to probability and statistics. McGraw-HIll, 811 pp.

Moron, V., Gouirand, I., 2003. Seasonal modulation of the El Niño-Southern Oscillation relationship with sea level pressure anomalies over the north Atlantic in October-Marsh 1873-1996. Int. J. Climatol. 23, 143–155.

Murteira, B. J. F., 1990. Probabilidades e Estatística, Segunda Edição. Vol. II. McGraw-HIll, 463 pp.

National Research Council, 1996. Learning to predict climate variations associated with El Niño and the Southern Oscillation: Accomplishments and legacies of the TOGA program. National Academy Press, Washington DC.

Neelin, J. D., Battisti, D. S., Hirst, A. C., Jin, F.-F., Wakata, Y., Yamagata, T., Zebiak, S. E., 1998. ENSO theory. J. Geophys. Res. 103, 14261–14290.

Opsteegh, J. D., den Dool, H. M. V., 1980. Seasonal differences in the stationary response of a linearized primitive equation model: prospects for long-range weather forecasting? J. Atmos. Sci. 37, 2169–2185.

Paeth, H., Hense, A., Glowienka-Hense, R., Voss, R., Cubasch, U., 1999. The North Atlantic Oscillation as an indicator for greenhouse-gas induced regional climate change. Clim. Dyn. 15, 953–960.

Pan, Y. H., Oort, A. H., 1983. Global climate variations connected with sea surface temperature anomalies in the eastern equatorial Pacific ocean for the 1958-73 period. Mon. Wea. Rev. 111, 1244–1258.

Peixoto, J. P., Oort, A. H., 1992. Physics of Climate. AIP Press, Springer-Verlag, New York, 520 pp.

Peterson, K. A., Greatbatch, R. J., Lu, J., Lin, H., Derome, J., 2002. Hindcasting the NAO using diabatic forcing of a simple AGCM. Geophys. Res. 29, 50,1–4.

Philander, S. G. H., 1981. The response of the equatorial oceans to a relaxation of the trade winds. J. Phys. Oceanogr. 11, 176–189.

Philander, S. G. H., 1985. El Niño and La Niña. J. Atmos. Sci. 42, 2652–2662.

Pozo-Vásquez, D., Esteban-Parra, M. J., Rodrigo, F. S., Castro-Déz, Y., 2001. The association between ENSO and winter atmospheric circulation and temperature in the north Atlantic region. J. Clim. 14, 3408–3420.

Pozo-Vásquez, D., Gámiz-Fortis, S. R., Tovar-Pescador, J., Esteban-Parra, M. J., Castro-Díez, Y., 2005a. El Niño-Southern Oscillation events and associated european winter precipitation anomalies. Int. J. Climatol. 25, 17–31.

Pozo-Vásquez, D., Gámiz-Fortis, S. R., Tovar-Pescador, J., Esteban-Parra, M. J., Castro-Díez, Y., 2005b. North Atlantic winter SLP anomalies based on the autumn ENSO state. J. Clim. 18, 97–103.

Preisendorfer, R. W., 1988. Principal Component Analysis in Meteorology and Oceanography. Elsevier, Amsterdam, 425 pp.

Rasmunsson, E. M., Carpenter, T. H., 1982. Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. Mon. Wea. Rev 110, 354–384.

Renshaw, A. C., Rowel, D. P., Folland, C. K., 1998. Wintertime low-frequency weather variability in the North Pacific-American sector 1949-93. J. Clim 11, 1073–1093.

Rocha, A., 1992. The influence of global sea surface temperatures on southern african summer climate. Ph.D. Thesis. University of Melbourne, Australia, 255 pp.

Rocha, A., Simmonds, I., 1997. Interannual variability of south-eastern african summer rainfall. Part I: Relationships with air-sea interaction processes. Int. J. Climatol. 17, 235–265.

Rodwell, M. J., Rowell, D. P., Folland, C. K., 1999. Oceanic forcing of the wintertime North Atlantic Oscillation and european climate. Nature 398, 320–323.

Rogers, J. C., 1984. The association between the North Atlantic Oscillation and the Southern Oscillation in the northern hemisphere. Mon. Wea. Rev. 112, 1999–2015.

Ropelewski, C. F., Halpert, M. S., 1987. Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. Mon. Wea. Rev. 115, 1606–1626.

Rowell, D. P., 1998. Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. J. Clim. 11, 109–120.

Rowell, D. P., Folland, C. K., Maskell, K., Ward, M. N., 1995. Variability of summer rainfall over tropical north Africa (1906-92): Observations and modelling. Q. J. L. Meteorol. Soc. 121, 669–704.

Saravanan, R., 1998. Atmospheric low-frequency variability and its relationship to Mid-latitude SST variability: studies using the NCAR climate system model. J. Clim. 11, 1386–1404.

Saravanan, R., Chang, P., 2000. Interaction between tropical Atlantic variability and El Niño-Southern Oscillation. J. Clim. 13, 2177–2194.

Sardeshmukh, P. D., Hoskins, B. J., 1988. The generation of global rotational flow by idealized tropical divergence. J. Atmos. Sci 45, 1228–1251.

Scheffé, H., 1959. The analysis of variance. John Wiley & Sons, 477 pp.

Shukla, J., 1983. Comments on Natural Variability and Predictability. Mon. Wea. Rev. 111, 581–585.

Silverman, B. W., 1986. Density estimation for statistics and data analysis. Chapman and Hall, 175 pp.

Simmonds, I., 1985. Analysis of the 'spinup' of a General Circulation Model. J. Geophys. Res. 90, 5637–5660.

Simmonds, I., 1990. Improvements in General Circulation Model performance in simulating Antarctic climate. Antarctic Sci. 2, 287–300.

Simmonds, I., Dix, M., 1989. The use of mean atmospheric parameters in the calculation of modeled mean surface heat fluxes over the word's oceans. J. Phys. Oceanogr. 19, 205–215.

Simmonds, I., Lynch, A., 1992. The influence of pre-existing soil moisture content on australian winter climate. Int. J. Climatol. 12, 33–54.

Simmonds, I., Smith, N., 1986. The effect of the prescription of zonally-uniform sea surface temperatures in a general circulation model. Int. J. Climatol. 6, 641–659.

Simmonds, I., Trigg, G., Law, R., 1988. The climatology of the Melbourne University General Circulation Model. Public. No 31, Dept. of Meteorol., University of Melbourne, 67 pp.

Smagorinsky, J., 1953. The dynamical influence of large-scale heat sources and sinks on the quasi-stationary mean motions of the atmosphere. Quart. J. Roy. Meteor. Soc. 79, 342–366.

Smith, I. N., 1995. A GCM simulation of global climate interannual variability: 1950-1988. J. Clim 8, 709–718.

Stern, W. F., Miyakoda, K., 1995. Feasibility of seasonal forecasts inferred from multiple GCM simulations. Journal of Climate 8, 1071–1085.

Sutton, R. T., Hodson, D. L. R., 2003. Influence of the ocean on north Atlantic climate variability 1871-1999. J. Clim. 16, 3296–3313.

Taylor, K. E., Williamson, D., Zwiers, F., 2000. The sea surface temperature and sea-ice concentration boundary conditions from AMIP II simulations. Program for Climate Model Diagnosis and Intercomparison, University of California, Lawrence Livermore National Laboratory, PCMDI Report No 60, 24 pp.

Trenberth, K. E., 1997. The definition of El Niño. Bull. Amer. Meteor. Soc. 78, 2771–2777.

Trenberth, K. E., Branstator, G. W., Karoly, D., Kumar, A., Lau, N.-C., Ropelewski, C., 1998. Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. J. Geophys. Res. 103, 14291–14324.

Trenberth, K. E., Caron, J. M., 2000. The Southern Oscillation revisited: sea level pressures, surface temperatures, and precipitation. J. Clim. 13, 4358–4365.

Trenberth, K. E., Stepaniak, D. P., Caron, J. M., 2000. The global monsoon as seen through the divergent atmospheric circulation. J. Clim. 13, 3969–3993.

van Loon, H., Madden, R. A., 1981. The Southern Oscillation. Part I: Global associations with pressure and temperature in northern winter. Mon. Wea. Rev 109, 1150–1162.

van Loon, H., Rogers, J. C., 1978. The seesaw in winter temperatures between Greenland and Northern Europe. Part I: General description. Mon. Wea. Rev 106, 296–310.

van Loon, H., Rogers, J. C., 1981. The Southern Oscillation. Part II: Associations with the middle troposphere in the northern winter. Mon. Wea. Rev 109, 1163–1168.

van Oldenborgh, G. J., 2005. Comments on "Predictability of winter climate over the North Atlantic European region during ENSO events. J. Climate 18, 2770–2772.

Venzke, S., Allen, M. R., Sutton, R. T., Rowell, D. P., 1999. The atmospheric response over the North Atlantic to decadal changes in sea surface temperature. J. Clim. 12, 2562–2584.

von Storch, F., Zwiers, F. W., 1999. Statistical Analysis in Climate Research. Cambridge University Press, 484 pp.

Walker, G. T., 1925. Correlation in seasonal variations of weather - a further study of World Weather. Mon. Wea. Rev. 53, 252–254.

Walker, G. T., 1928. World Weather. Mon. Wea. Rev. 56, 167–170.

Walker, G. T., Bliss, E. W., 1932. World Weather IV. Mem. Roy. Meteorol. Soc. 4, 53–84.

Wallace, J. M., Gutzler, D. S., 1981. Teleconnections in the geopotential height field during the northern hemisphere winter. Mon. Wea. Rev. 109, 784–812.

Wallace, J. M., Rasmusson, E. M., Mitchell, T. P., Kousky, V. E., Sarachik, E. S., 1998. On the structure and evolution of ENSO-related climate variability in the tropical Pacific: lessons from TOGA. J. Geophys. Res. 103, 14,241–14,259.

Wallace, J. M., Smith, C., Bretherton, C. S., 1992. Singular Value Decomposition of wintertime sea surface temperature and 500-mb height anomalies. J. Clim. 5, 561–576.

Walland, D. J., Simmonds, I., 1997. Modelled atmospheric response to changes in northern hemisphere snow cover. Clim. Dyn. 13, 25–34.

Walter, K., Graf, H.-F., 2002. On the changing nature of the regional connection between the north atlantic oscillation and sea surface temperature. J. Geophys. Res. 107(D17), 4338.

Wang, B., 1992. The vertical structure and development of ENSO anomaly mode during 1979-1989. J. Atmos. Sci. 49, 698–712.

Wang, C., 2002a. Atlantic climate variability and its atmospheric circulation cells. J. Clim. 15, 1516–1536.

Wang, C., 2002b. Atmospheric circulation cells associated with El Niño-Southern Oscillation. J. Clim. 15, 399–419.

Wang, X. L., Zwiers, F. W., 1999. Interannual variability of precipitation in an ensemble of AMIP climate simulations conducted with the CCC GCM2. J. Clim. 12, 1322–1335.

Ward, M. N., Navarra, A., 1997. Pattern analysis of SST-forced variability in ensemble GCM simulations: Examples over Europe and the tropical Pacific. J. Clim. 10, 2210–2219.

Weare, B. C., 1982. El Niño and tropical Pacific ocean surface temperatures. J. Phys. Oceanogr. 12, 17–27.

Weare, B. C., Navato, A. R., Newell, R. E., 1976. Empirical Orthogonal Analysis of Pacific sea surface temperatures. J. Phys. Oceanogr. 6, 671–678.

Webster, P. J., 1981. Mechanisms determining the atmospheric response to sea surface temperature anomalies. J. Atmos. Sci. 38, 554–571.

Webster, P. J., 1982. Seasonality in the local and remote atmospheric response to sea surface temperature anomalies. J. Atmos. Sci. 39, 41–52.

Wilks, D. S., 1995. Statistical Methods in the Atmospheric Sciences. Vol. 59 of International Geophysics Series. Academic Press, 464 pp.

Wolter, K., Timlin, M. S., 1998. Measuring the strength of ENSO events: How does 1997/98 rank? Weather 53, 315–324.

Wu, A., Hsieh, W. W., 2004. The nonlinear association between ENSO and the Euro-Atlantic winter sea level pressure. Clim. Dyn. 23, 859–868.

Wu, A., Hsieh, W. W., Zwiers, F. W., 2003. Nonlinear modes of north american winter climate variability derived from a General Circulation Model simulation. J. Clim. 16, 2325–2339.

Wyrtki, K., 1975. El Niño - The dynamic response of the equatorial Pacific ocean to atmospheric forcing. J. Phys. Oceanogr. 5, 572–584.

Xie, P., Arkin, P. A., 1996. Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. Bull. Amer. Meteor. Soc. 78, 2539–2558.

Zwiers, F. W., 1996. Interannual variability and predictability in an ensemble of AMIP climate simulations conducted with the CCC GCM2. Clim. Dyn. 12, 825–848.

Zwiers, F. W., Wang, X. L., Sheng, J., 2000. Effects of specifying bottom boundary conditions in an ensemble of atmospheric GCM simulations. J. Geophys. Res. 105, 7295–7315.