Royal Netherlands
Meteorological Institute
Ministry of Infrastructure and the Environment

Statistical analysis of total column ozone data

Jurrien Knibbe

De Bilt, 2013 | Trainee report

Statistical analysis of total column ozone data

Version	1

Date	March 27, 2013
Status	Final

Statistical analysis of total column ozone data

Jurrien Knibbe

March 27, 2013

vrije Universiteit amsterdam

Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

Preface

As a master student in mathematics at the VU in Amsterdam I was looking for a research project related to my large interests in earth sciences. Therefore, I chose to do a research internship at the KNMI, the Royal Netherlands Meteorological Institute.

I am familiar with this institute due to a previous internship for my bachelor graduation in mathematics in 2010, which I also did at the KNMI for three months. Back then, I performed a statistical trend analysis on satellite measurements of the cloud variables 'cloud fraction', which is a measure for the cloud cover, and 'cloud pressure', which is a measure for the height of the cloud. My supervisors were Dr. Piet Stammes and Dr. Ping Wang from the KNMI and Prof. Mathisca de Gunst from the VU in Amsterdam.

For my master graduation I was again welcome to do an internship in 2012-2013, now for six months, to perform a statistical analysis on how ozone is dependent on other variables. Supervisors on this research are Dr. Ronald van der A from the KNMI and again Mathisca de Gunst. I would like to take this opportunity to thank them and Dr. Jos de Laat from the KNMI for the guidance and reflecting comments on this research throughout the project. I also would like to thank Piet Stammes for mentoring conversations.

Contents

1 Introduction

1.1 KNMI

The KNMI is mostly known for its weather forecasts, but it has a much broader field of scientific expertise. The KNMI is the leading research institute for meteorology, seismology and climate research in the Netherlands. As an agency within the ministry of infrastructure and environment, the KNMI is responsible for weather forecasting, providing weather and climate data for the private sector, launching weather alarms and representing the Netherlands in international research organizations such as the IPCC, WMO, ECMWF, EUMETSAT and EUMETNET. The research part of the institute is split between 'Climate and Seismology' and 'Meteorology'. As an intern I am positioned in the 'Earth-observation Climate' group, which is a division of the 'Climate and Seismology' department. The main focus of this group is on satellite retrieval algorithms, data processing and data analysis.

Figure 1: The KNMI, located in de Bilt. Source: www.knmi.nl.

1.2 Research question

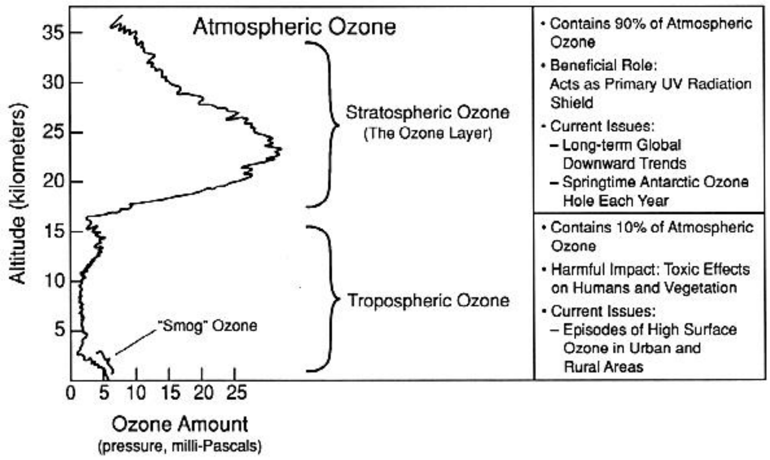


Figure 2: The distribution of a column ozone in the atmosphere. Source: WMO, Scientific assessment of ozone depletion: 1994.

Earth's atmosphere consists of several layers with different characteristics; the troposphere, the stratosphere, the mesosphere and the thermosphere, respectively, in order of increasing height. The layers are generally defined by the temperature profile: in the troposphere temperature decreases in height, in the stratosphere temperature increases in height, in the mesosphere temperature decreases in height and in the thermosphere temperature increases in height.

Figure 2 shows the distribution of ozone in height and why ozone is important in our atmosphere. 90% of all ozone is located in the stratosphere, where it absorbs a large part of the solar UV-radiation. This type of radiation is harmful for all forms of life and vegetation when it reaches earth's surface in high doses. Therefore, this stratospheric ozone layer is of great importance. The other 10% of ozone is located in the troposphere, where it is an important component of air pollution.

In the last decades of the twentieth century a decrease of stratospheric ozone has been detected. This decrease is related to a large amount of emissions of chlorofluorocarbons (CFCs). Above the Antarctic such a decrease is problematic because the Antarctic depletion of ozone is much higher than in other regions. Especially in the September-November months special meteorological conditions, such as the isolation of stratospheric air and extremely low temperatures, cause massive ozone destruction above the South Pole. In presence of CFC gasses in the stratosphere, the ozone can be completely destroyed in this region. Since 1987 the CFC emissions are banned in most countries by the Montreal protocol. In subsequent amendments, other ozone depleting substances (ODS), such as Hydrobromofluorocarbons (HBFCs) and hydrochlorofluorocarbon (HCFCs), were banned. The expectation is that the ozone layer in the stratosphere will restore itself in the coming decades.

The goal of this research is to investigate how monthly variability in ozone depends on explanatory variables spatially across the globe. These variables may affect ozone chemically, by their influence on chemical reactions, or dynamically, by their influence on the spatial distribution of ozone such as wind patterns. We use explanatory variables to describe the dynamical behaviour of ozone instead of spatial mathematical modeling, because it is extremely difficult to realistically model the dynamics of ozone on a monthly time scale by spatial statistical modeling. Instead, there are several climate indices available to account for the spatial dynamics in ozone in a more reliable way. We will build statistical models, which are used to perform statistical inference. Similar statistical analyses have previously been performed on equivalent latitude bands of satellite data by Stolarski et al. [1991], Bodeker et al. [1998, 2001] and Brunner et al. [2006] among others and on ground-based measurement data of several ground stations by e.g. Wohltmann et al. [2007]. In the present study, however, the analysis is performed on 1 by 1.5 degree sized grid cells. Therefore, we are not limited to ozone data of specific locations on earth, such as ground stations, moreover, we are able to analyze ozone locally instead of on equivalent latitude bands. This enables us to investigate spatial patterns that illustrate where each of the explanatory variables affect ozone.

2 Variables and data description

In this section all variables, which are included in this research, are introduced. In the sequel we shall denote the model variables that represent our explanatory variables of interest by 'proxies'.

2.1 Ozone

Ozone is a molecule consisting of three oxygen atoms, chemically denoted as O_3. In the troposphere ozone is an unstable molecule with a short life time. It breaks down to the more stable ordinary oxygen molecule O_2 by the net chemical reaction

$$2O_3 \rightarrow 3O_2. \tag{1}$$

Solar radiation of wavelengths smaller then $315nm$ are known to be dangerous for life forms and vegetation, causing sunburn and direct DNA damage in skin tissue and vegetation. 99% of our atmosphere consists of O_2 and N_2 gasses that absorb all wavelengths smaller than $200nm$. Ozone is responsible for absorbing the solar wavelengths between 200 and 315 nm. Therefore, although ozone is a toxic gas at ground level, at higher altitudes it is a useful gas. UV levels due to low O_3 concentrations have a damaging effect. This, for example, increases the risk of skin cancer. High

At ground level the ozone concentration is between 0.001 and 0.5 parts per million (ppm.). The air at ground level is considered polluted if this concentration exceeds 0.1 ppm. Ozone is more stable at stratospheric heights due to the lower temperature and pressure, which decreases the probability of a collision between ozone molecules for reaction 1. More importantly, ozone is mainly formed under influence of solar UV radiation, with wavelengths smaller than 200 nm, as will be explained in the next section. This type of radiation is completely absorbed by the O_2 and N_2 molecules in the stratosphere and does not reach lower altitudes. This leads to an ozone concentration between 2 to 8 ppm. in the stratosphere. Thus, most of the atmospheric ozone is located in the stratosphere, shielding the lower atmosphere from harmful UV radiation.

The Dobson Unit (DU) is used as a unit for the total amount of ozone in an vertical column of air. One DU of a gas is defined as a $0.01mm$ thick layer of this pure gas at a pressure level of $1000hPa$ (ground level) and 15 degrees in Celsius (ground temperature). The average amount of ozone in our atmosphere is among $300DU$, which thus corresponds to a $3mm$ thick layer of ozone at ground level.

The data that is used for this variable is the ozone Multi Sensor Re-analysis (MSR) data-set from 1979-2008 described in Van der A et al. [2010] combined with two years (2009 and 2010) of data from the SCIAMACHY satellite instrument described in Eskes et al. [2003]. The MSR data are assimilated measurements from the TOMS, SBUV, GOME, SCIAMACHY, OMI and GOME-2 satellite instruments. Independent ground-based ozone data are used to correct for the bias in the satellite measurements. The MSR data-set, together with the two years of SCIAMACHY data, contain 32 years of monthly means of ozone in Dobson Units on a 1 by 1.5 degrees sized grid and the standard errors corresponding to these monthly averaged ozone values in terms of DU.

2.2 Solar radiation

As was mentioned briefly in the introduction, O_2 'absorbs' solar wavelengths smaller than $200nm$. This occurs through the chemical reaction

$$O_2 + photon(< 200nm) \rightarrow 2O. \tag{2}$$

This reaction requires the energy of a photon with short wavelengths. Due to this reaction O radicals are created. These atoms react with O_2 through the reaction

$$O_2 + O \rightarrow O_3 + heat, \tag{3}$$

producing ozone and heat. The latter reaction goes very fast due to the large amount of O_2 molecules to engage a reaction. Solar radiation is a necessary variable for the formation of ozone (see equation 2). Most of the ozone is produced in the tropics, because there the solar radiation is the highest.

On the other hand, as was also mentioned in section 2.1, O_3 molecules 'absorb' solar wavelengths between 200 and 315 nm. This is done by the reaction

$$O_3 + photon(> 200nm, < 315nm) \rightarrow O_2 + O. \tag{4}$$

From this equation it seems that solar radiation is responsible for depletion of ozone. But since the latter reaction generates a free oxygen atom, ozone will regenerate due to the fast reaction of equation 3. For this reason, the negative effect by the chemical reaction of equation 4 on ozone depletion is small. However, solar radiation can cause depletion of ozone indirectly. This happens through a catalytic mechanism involving chloride or bromide if the temperature is sufficiently low.

The intensity of solar radiation has a cycle of roughly eleven years, which particularly dominates the UV-radiation and is, therefore, important for ozone production. For the solar intensity data, the proxy of measured total global solar irradiance from Frohlich et al. [2000] is used. This time series contains monthly means from 1979-2010 of measured total solar irradiance in W/m^2.

2.3 Chlorine and Bromide

In the twentieth century, the industrial countries emitted a lot of ODS containing chloride and bromide atoms. These gasses ascend to stratospheric heights where the strong solar radiation destroys the CFCs, HCFCs and HBFCs, thereby releasing chloride and bromide atoms. These chloride and bromide atoms react with ozone creating chloride or bromide-monoxide through the reactions

$$\begin{aligned} O_3 + Cl &\rightarrow O_2 + ClO, \\ O_3 + Br &\rightarrow O_2 + BrO. \end{aligned} \tag{5}$$

After the above reaction, even more ozone is depleted due to chloride and bromide via the reactions

$$\begin{aligned} O_3 + ClO &\rightarrow 2O_2 + Cl, \\ O_3 + BrO &\rightarrow 2O_2 + Br. \end{aligned} \tag{6}$$

We see that the chemical reactions of equations 5 and 6 form a catalytic cycle until the chloride and bromide are bonded with molecules. Therefore, these reactions contribute a lot to the depletion of ozone.

As a proxy for the chloride and bromide variables we use the effective equivalent stratospheric chlorine (EESC). To calculate this data we follow the procedure described in Newman et al. [2006]. This time series is a measure for the effective amount of chloride and bromide in the stratosphere.

2.4 Stratospheric temperature

The stratospheric temperature has an important role in the chemistry of ozone depletion. In the stratosphere it can get extremely cold. When the temperature drops below -80 degrees Celsius, stratospheric clouds may form. These clouds consist of very small ice particles. Along these ice particles the CFC and HBFC gasses lose chloride and bromide atoms more rapidly under influence of solar radiation. This is speeding up the reactions of equations 5 and 6.

Temperature has a strong seasonal cycle, especially at the Poles where it can attain values below -80 degrees Celsius in the corresponding winter. Due to these low values ozone is mainly depleted at polar regions.

As a proxy for the stratospheric temperature we use the effective ozone temperature, as calculated in Van der A et al. [2010]. This quantity is calculated as the weighted mean temperature in the atmosphere, with the column-wise fraction of ozone as weights of the temperature at corresponding heights. This data is monthly averaged and gridded on 1 by 1.5 degree.

2.5 Aerosols

In order for aerosols to affect ozone significantly, they must ascend to stratospheric heights. This is mainly caused by volcanic eruptions. In the time-span of our ozone data there were 2 major volcanic eruptions that were able to send aerosols to such heights: The El Chicon volcano eruption in Mexico (1982) and the Pinatubo volcano eruption in the Philippines (1991).

Aerosols affect ozone in different ways. They affect ozone because the reactions of equations 5 and 6 are catalyzed along the surface of aerosols, as happens with the ice particles of polar stratospheric clouds. But high aerosols also prohibit solar radiation from reaching the ozone-layer, cooling the stratosphere, while lower aerosols tend to warm the stratosphere.

For the aerosol data we use two proxies corresponding to aerosols between 15-20 km and between 20-25 km. These proxies are measured in terms of optical thickness and consist of 24 zonal longitudinal bands with a width of 7.5 degrees in latitudes. This data is described in Sato et al. [1993].

2.6 The polar vortex

In addition to the chemical factors that play a role in ozone formation and depletion, the air dynamics in the stratosphere are important for the distribution of ozone. The stratospheric polar vortex is of great importance at the Poles and mid-latitudes. This polar vortex is a westerly downward flow at 45 to 70 degrees in latitudes, circulating around the polar regions. The vortex at the Southern Hemisphere is mostly situated above the southern ocean and is, therefore, barely disturbed. This leads to a strong southern polar vortex, whereas the northern vortex is more disturbed by the Himalaya and the Rocky Mountains. These mountains cause so-called Rossby waves, a tropospheric wind circulation, which in turn disturbs the northern polar vortex.

Strong polar vortices hinder the air exchange between the Poles and the mid-latitudes; the polar air masses get isolated. This is one of the mechanisms that is responsible for the ozone hole, which is formed every year at the South Pole. The strength of the polar vortex is seasonally dependent.

For the polar vortex data we use 2 proxies of the vertical EP-flux described in Kanamitsu et al. [2002] corresponding to the northern and southern polar vortex. The vertical EP-flux is a measure of the strength of the polar vortex. The flux is calculated from wind speed measurements performed at ground stations around the world. We average the vertical EP-flux at 100 hPa over 45-70 degrees in latitudes of the corresponding Hemisphere, obtaining 2 time series as a measure for the strength of the polar vortices.

The effect of the polar vortex on ozone lasts more than a month. In the winter period of the corresponding Hemisphere the strength of this effect decreases more slowly than in the summer period. Therefore, we transform the EP-flux proxies as in Brunner et al. [2006]:

$$x_{EP}(t) = x_{EP}(t-i) \cdot exp(\frac{1}{\tau}) + \tilde{x}_{EP}(t), \tag{7}$$

where x_{EP} is the transformed proxy, $\tilde{x}_{EP}$ the original proxy and τ is set to 3 months in the corresponding summer and to 12 months in the corresponding winter months.

2.7 Quasi Biennial oscillation

The Quasi Biennial oscillation (QBO) is a tropical stratospheric wind pattern with a large impact on ozone. The direction of these winds oscillate from eastern to western with a period of about two years. These winds are situated at the equator and thus affect ozone mainly at the equator and mid-latitudes. The direction and intensity of the winds are dependent on the height in the stratosphere. The oscillation starts at the top of the stratosphere, slowly descending to the bottom. Therefore, the QBO has a somewhat different structure at different heights, see figure 3.

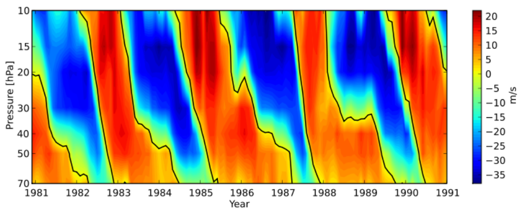


Figure 3: This picture shows the QBO oscillation at the equator. In this figure positive speed corresponds to wind in the western direction. Source: www.wikipedia.org, data from FU Berlin.

As a proxy for the QBO variable we use two time series of measure wind speeds at 30 hPa and 10 hPa in height described in Baldwin et al. [2001]. These time series were calculated from daily wind measurement at several ground stations around the equator. The data is given in m/s.

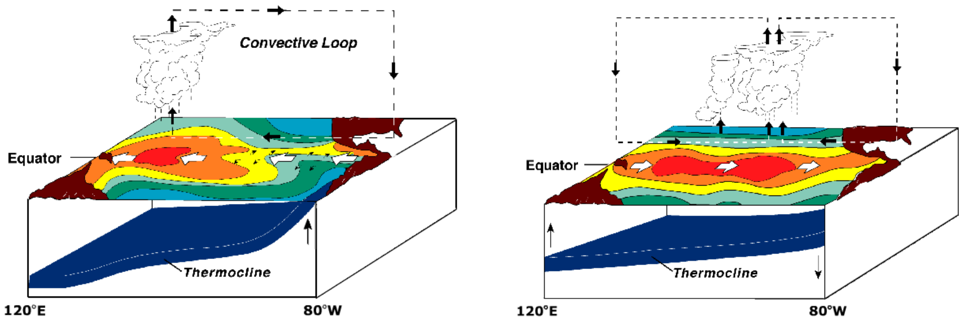


Figure 4: The left picture shows the normal situation in the Pacific, and the right picture shows the situation during an El Nino. Red corresponds to surface water of high temperature and blue corresponds to surface water of low temperature. Source: PMEL/NOAA/TAO.

2.8 El Nino

The El Nino is a weather phenomenon originating in the Pacific Ocean. In normal conditions the surface water in the Pacific Ocean at the equator flows in the western direction. This causes an upwelling of cold water at the eastern Pacific. During an El Nino period, a surface flow in the eastern direction stops the upwelling of cold water. This causes a warmer sea temperature and a warmer tropospheric air temperature. Coupled to this change in the oceanic flow pattern is a change in air flow patterns. In normal conditions there is one convective loop of air above the pacific, while during the El Nino, this loop is broken down in two smaller loops. Figure 4 illustrates the El Nino phenomenon.

Due to these changes in wind and oceanic flows patterns, the El Nino can have a global effect on weather and temperature. Especially in 1998 there was a large El Nino. Even in the stratosphere such an El Nino has dynamically related consequences.

As a proxy for the El Nino variable we use the calculated multivariate El Nino southern oscillation index (ENSO) as described in Wolter and Timlin. [1993, 1998]. This time series contains monthly values from 1979-2010 as a measure for the global intensity of El Nino.

2.9 Geopotential height

The boundary between the troposphere and the stratosphere is called the tropopause. The height of the tropopause is very stable at the tropics. However, at mid-latitudes and Poles the height of the tropopause is strongly dependent on weather systems. When this boundary ascends, it pushes the stratospheric layer upwards. This causes air, and thus ozone, to move to other parts of the stratosphere. Geopotential height is a measure for the tropopause height. It is seasonally dependent, especially at mid-latitudes.

As a proxy for the geopotential height we use data from the ECMWF. The used data-set contains gridded values of geopotential height at 500 hPa.

2.10 Potential vorticity

The potential vorticity is a preserving quantity corresponding to the force of a spinning motion of a fixed mass of air. This quantity is preserved under the vertical stretching of the air mass. For example, the potential vorticity of a rotating figure skater is preserved under vertical stretching of his body, although the rotation speed will increase.

Due to the Coriolis force, the air masses in the Southern Hemisphere rotate in an opposite direction with respect to the air masses in the Northern Hemisphere. This results in a change of sign in the potential vorticity between the air masses of both Hemispheres, which is positive for clockwise rotation. Previous studies have shown high correlations between potential vorticity and ozone (Allaart et al. [1993], Riishøjgaard and Källén. [1997]).

As a proxy for the potential vorticity we again use data from the ECMWF. The used data set contains gridded values of potential vorticity at 150 hPa.

3 Data overview and correlation

3.1 Data overview

In this section an overview of the data is given, and the normalization transformation of the proxies is explained. Table 1 lists all proxies, together with a description and data source. The time series of the proxies in table 1 are all normalized to zero with a standard deviation of one using the formula

$$X_j = \frac{x_j - \bar{x}_j}{sd(x_j)}, \tag{8}$$

with x_j corresponding to the original proxy corresponding to variable j, $\bar{x}_j$ to its mean value and X_j corresponding to the normalized unit less proxy corresponding to variable j. The normalized proxies are plotted in figure 5. This normalization is important for two reasons: first, due to the normalization the regression estimates are in the same order of magnitude, and can be compared to each other. Second and more important, this normalization is necessary when we create 'alternative variables', which we will introduce in chapter 4.

Table 1: list of variables

Proxy	Data description	Source
O3	Globally gridded(1x1.5 degrees) ozone in Dobson Units	www.temis.nl/protocols/O3global.html
SOLAR	Averaged total solar irradiance in $\frac{W}{m^2}$	www.esrl.noaa.gov/psd/enso/mei/
EESC	Effective stratospheric chlorine and bromide	acd-ext.gsfc.nasa.gov/Data_services/automailer/index.html
TEMP	Effective ozone temperature (gridded)	www.atmos-chem-phys-discuss.net/10/11401/2010/acpd-10-11401-2010.pdf
AERO[25-30]	7.5 degree zonal bands of Aerosol Optical Thickness averaged over 25-30 km. level	data.giss.nasa.gov/modelforce/strataer/tau_map.txt
EPFLUX-N	vertical Elijassen-Palm flux at 100 hPa averaged over 45-90 degrees north	www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html
EPFLUX-S	vertical Elijassen-Palm flux at 100 hPa averaged over 45-90 degrees south	www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html
QBO10	index for Quasi Biennial Oscillation at 10 hPa	www.geo.fu-berlin.de/en/met/ag/strat/produkte/qbo/
QBO30	index for Quasi Biennial Oscillation at 30 hPa	www.geo.fu-berlin.de/en/met/ag/strat/produkte/qbo/
ENSO	Multivariate El Nino southern oscillation index	www.esrl.noaa.gov/psd/enso/mei/
GEO	Geopotential height of the 500 hPa level (gridded)	data-portal.ecmwf.int/data/d/interim_moda/levtype=pl/
PV	Potential vorticity at 150 hPa (gridded)	data-portal.ecmwf.int/data/d/interim_moda/levtype=pl/

3.2 Correlation

Correlation between explanatory variables can cause problems in a regression analysis. When two or more explanatory variables are highly correlated, it becomes unclear which variable should be included in the model. Therefore, one has to check for collinearity between the explanatory variables.

From figure 5 we see that the EPFLUX, TEMP, PV and GEO are the proxies with a clear seasonal component. Therefore, we assume that the correlations between this group of variables and the other explanatory variables are low. The correlations between the variables are analyzed separately for these groups. Variables are considered too much correlated when their correlation value exceeds 0.4.

Table 2 shows the correlations between the proxies of the EESC, SOLAR, AERO, EPFLUX, QBO and ENSO proxies, where the AERO proxy is average between -61 to -69 degree latitudes. From this table we conclude that the two aerosol proxies are highly correlated. To avoid problems we choose not to include the AERO[15-20] proxy in this study, because more ozone is between 20-25 km. In addition, we notice the high correlations of the QBO50 proxy with respect to the other QBO proxies. Due to these correlation values we choose not to include the QBO50 proxy in the regressions. Furthermore, the EPFLUX-N and EPFLUX-S proxies are strongly anti correlated. We choose to use the EPFLUX-N proxy for the regressions performed in the Northern Hemisphere and EPFLUX-S for regressions performed in the Southern Hemisphere. Besides these correlation values, none of the values in table 2 are considered problematic.

The correlations between the variables with a strong seasonal component are shown in figure 6. In the plots we see that these variables are highly correlated at some regions. Especially GEO, TEMP and EPFLUX are highly correlated at the Northern Hemisphere. This problem will partially be solved due to the variable selection method, that will be applied in sections 4.2 and 5.3. This method selects variables in the regression

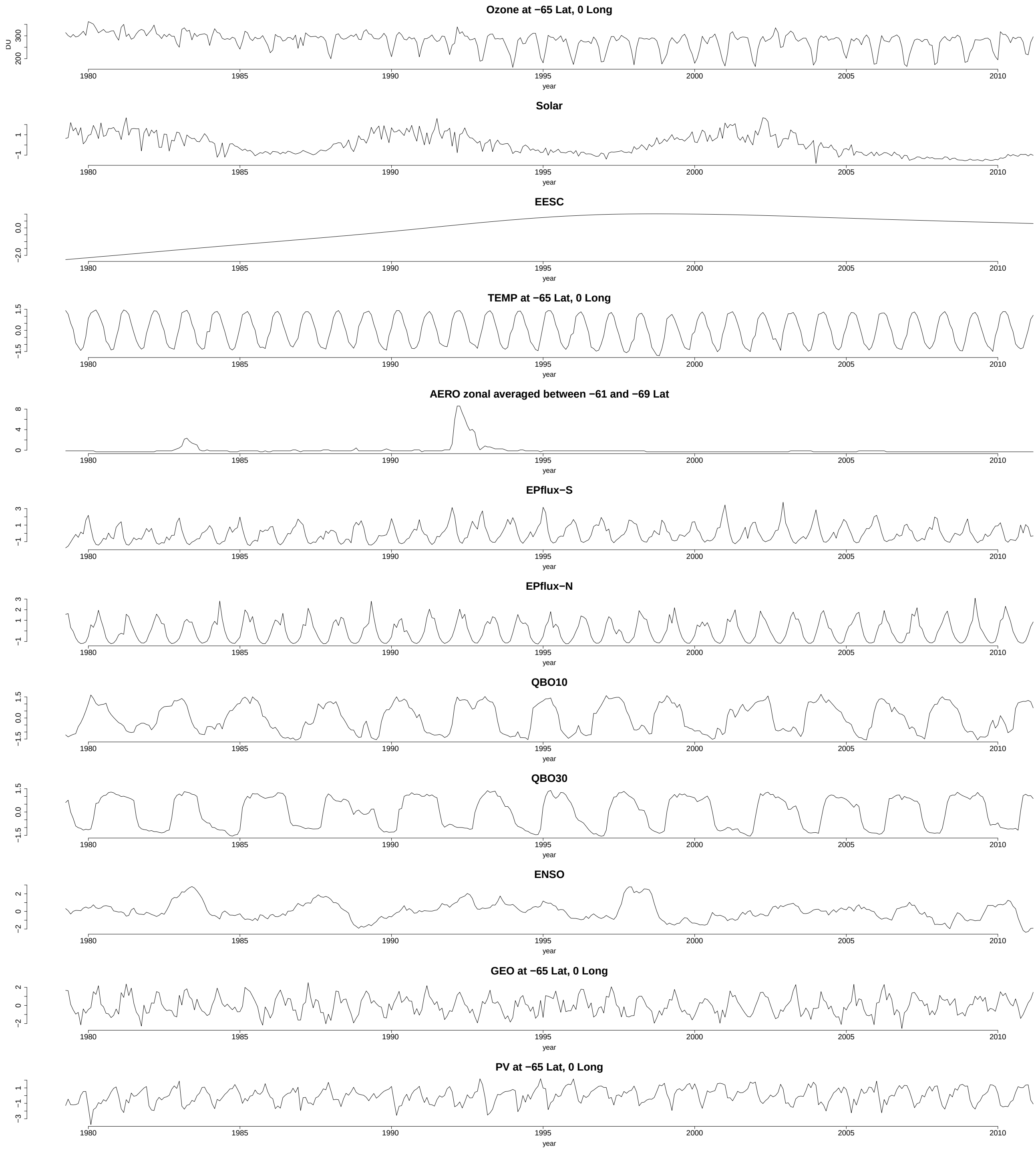


Figure 5: The upper plot shows the ozone in DU at -60 degrees in latitudes and 0 degrees in longitudes. The other plots show the normalized proxies of the explanatory variables, where TEMP, AERO, GEO and PV are taken at the same region as the ozone time series.

model based on the significance of their contribution to the model. This procedure often selects one of the highly correlated explanatory variables for the regression model. This method, however, does not entirely solve the problems corresponding to these high correlation values. This will be a subject of discussion when the models are introduced in section 5.2, and when the results are discussed in chapter 7.

Table 2: Table of correlations of non-gridded proxies

Proxy	Solar	EESC	AERO[15-20]	AERO[20-25]	EPFLUX-N	EPFLUX-S	QBO10	QBO30	QBO50	ENSO
Solar	1	-0.29	0.06	0.18	0.04	-0.09	0.01	0.03	0.02	0.04
EESC	-0.29	1	-0.05	-0.22	0.02	0.18	0.03	0.01	-0.02	-0.12
AERO[15-20]	0.06	-0.05	1	**0.61**	-0.06	0.16	0.08	0.01	-0.06	0.32
AERO[20-25]	0.18	-0.22	**0.61**	1	0.01	0.11	0.13	-0.03	-0.01	0.29
EPFLUX-N	0.04	0.02	-0.06	0.01	1	**-0.52**	0.03	0.14	0.06	-0.05
EPFLUX-S	-0.09	0.18	0.16	0.11	**-0.52**	1	0.05	-0.18	-0.14	0.01
QBO10	0.01	0.03	0.08	0.13	0.03	0.05	1	0.03	**-0.69**	-0.02
QBO30	0.03	0.01	0.01	-0.03	0.14	-0.18	0.03	1	0.34	0.04
QBO50	0.02	-0.02	-0.06	-0.01	0.06	-0.14	**-0.69**	0.34	1	0.1
ENSO	0.04	-0.12	0.32	0.29	-0.05	0.01	-0.02	0.04	0.1	1

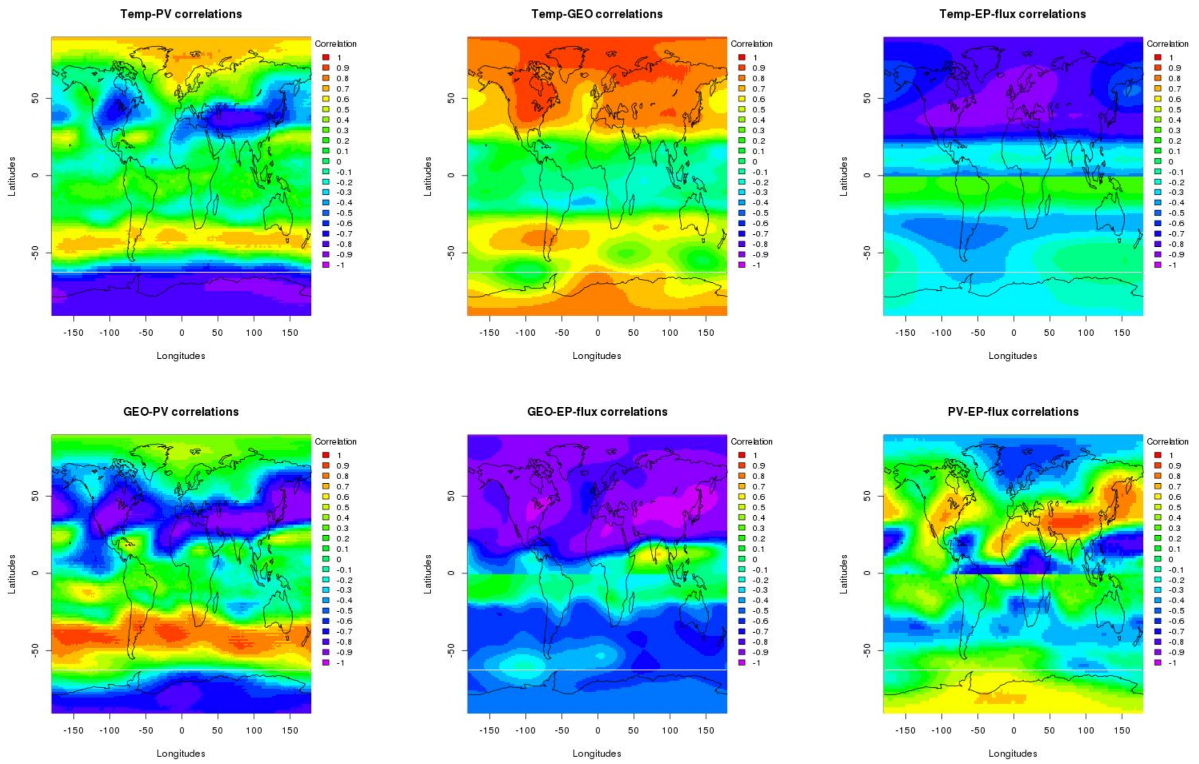


Figure 6: These plots show the correlation values between corresponding variables per grid cell, where EPFLUX-N is used in the Northern Hemisphere, and EPFLUX-S in the Southern Hemisphere.

4 Analysis of seasonality

4.1 Seasonality in ozone dependencies

There are two ways that seasonality plays a role in this analysis. There is the seasonal variation in ozone, which is discussed in the next section, and there is the seasonality in the dependencies between ozone and its explanatory variables, which is analyzed in this section.

The seasonality in dependencies is illustrated by an example considering clouds and ground temperature: Suppose we would like to examine the effect that clouds have on ground temperature. We know that during daytime clouds have a negative effect on ground temperature, because they reflect sunlight before reaching the earth's surface. However, during the night, clouds have a positive effect on ground temperature, because they work as a blanket, keeping the warmth in the lower troposphere.

Now suppose we perform a regression of hourly data of ground temperature against hourly data of cloud fraction measurements. The estimate for the clouds regression coefficient will be based on the mean effect that clouds have on temperature. As a result the effect that clouds have during the day might cancel out the effect that clouds have during the night. Therefore, the estimate might not even be found significantly different from zero.

A way to deal with this problem is to perform the regression of temperature data against an alternative variable to account for the effect caused by clouds. When the alternative variable consists of the clouds measurement data for day time data and minus the clouds measurement data for the night time data, the effect that night time data of the alternative variable has on the regression coefficient does not cancel out the affect of the day time data of the alternative variable. Consequently, this alternative variable should explain more of the variability in the temperature data than the original cloud measurements.

In this study we encounter the same problem on a seasonal timescale. Therefore, before we can model ozone properly, we first need to study these seasonalities in dependencies.

4.2 Monthly regression model and regression method

In this section we perform regressions of annual data for each month to study the effect of the explanatory variables throughout the season, as is done in Brunner et al. [2006]. We assume that the effect of the explanatory variables depends more on latitude than on longitude. Therefore, we average all gridded data, of ozone and explanatory variables, along the longitudes for these regressions. Hereafter, we again normalize these yearly time series using equation 8.

For the regressions we do a backwards selection of variables in linear regression models. This selection is done in the following regression steps:

In the first regression step we use the linear regression model

$$Y = X\beta + \epsilon, \tag{9}$$

where Y denotes the vector of our dependent variable ozone of yearly values on a particular latitude and month, X the matrix with the proxies of the explanatory variables as columns including the intercept as a column of ones, β the vector of the regression coefficients corresponding to the columns in X and ϵ the noise vector, where we assume that the elements of the noise vector are uncorrelated and normally distributed with mean zero and variance σ^2.

The regression coefficients $\hat{\beta}$ are estimated by the least squares estimates

$$\hat{\beta} = (X'X)^{-1}X'Y. \tag{10}$$

Subsequently, the noise vector ϵ and the covariance matrix of $\hat{\beta}$ are estimated by

$$\begin{aligned} \hat{\epsilon} &= Y - X\hat{\beta}, \\ \widehat{Var}(\hat{\beta}) &= (X'X)^{-1}\hat{\sigma}^2, \end{aligned} \tag{11}$$

where $\hat{\epsilon}$ is the vector of residuals, $\hat{\sigma}$ the estimated standard deviation of the noise and $\hat{Var}(\hat{\beta})$ the estimate for the covariance matrix of $\hat{\beta}$. From the regression estimates we calculate a vector of t-statistics T, where the i'th element of T is given by

$$T_i = \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}}. \tag{12}$$

Here $\hat{\sigma}^2_{\hat{\beta}_i}$ is the i'th diagonal element of the co-variance matrix estimate $\hat{Var}(\hat{\beta})$. From these t-statistics we calculate the corresponding P-values. If a P-value exceeds the significance level α we remove the proxy corresponding to the largest obtained P-value from the regression model in equation 9 and we repeat the above procedure. Thus, this proxy's corresponding column in X will be removed in the next step. Note, that in each regression step, the least reliable proxy is removed from the regression model, until all proxies are reliable up to the chosen significance level α. In these regressions we use a loose significance level α of 0.1 instead of the usual values 0.05 and 0.01, because we are not interested in the significance of the results but in whether any patterns arise in the results.

4.3 Results and interpretations

The results of the monthly regressions are shown in figure 7. We use these plots to characterize the seasonalities in dependencies. We choose to characterize these seasonalities as sines and cosines with a period of a year or half a year to stay in line with other literature in this field (Brunner et al. [2006] and Bodeker et al. [2001] among others). These harmonic characterizations will be used to create the alternative variables discussed in section 4.1, by multiplication with the original time series. Note that sines and cosines are oscillating between positive and negative values and are, therefore, applicable in the method that is discussed in section 4.1. A difference between the method in former studies and the method in this study is that here the seasonal analysis is used to find the best phase for the harmonic characterizations to model the seasonality in the dependency for each proxy, whereas the previous studies model these seasonalities by multiple sines and cosines with periods of a year and, for some variables, half a year. Following the method discussed in the previous section, the former studies include up to four alternative variables per explanatory variable. By choosing the phase of the harmonics correctly, we hope to get the same or better results with the use of less alternative variables.

From figure 7 there does not seem to be a clear seasonality in the dependency of ozone on the SOLAR proxy. Therefore, we don't include an alternative variable corresponding to this proxy.

For the EESC variable, we see that there is a large seasonality in the dependency of ozone with EESC at the polar regions. We characterize this seasonality with a sine of a yearly period with a maximum in April and its lowest value in October. This sine aligns the regressions coefficients in figure 7 at the South Pole, and is opposite to those at the North Pole. Therefore, the effect of the alternative variable EESC_2, defined in equation 13 below, on ozone is expected to be positive at the South Pole and negative at the North Pole.

The TEMP proxy particularly has a large effect in the Southern Hemisphere. In this region, the effect is larger in the summer months with respect to the winter months. We model this change in ozone dependency using a cosine starting its yearly period in January. This cosine is opposite to the regression coefficients in the Southern Hemisphere and, based on these results, the TEMP corresponding alternative variable should affect ozone negatively.

The regression coefficients corresponding to the aerosols do not show clear patterns throughout the year. Therefore, we choose not the include any seasonal dependence for this variable.

Considering the results of the EPFLUX proxy, we see that the effect of the polar vortex on ozone is very irregular throughout the season. In the Southern Hemisphere we see different regression estimates for the summer than for the winter. In the Northern Hemisphere the regression estimates differ for the January, April, August and November months. Therefore, we use a cosine with a yearly period and a cosine with a period of half a year, where both period start in January, to model the change in ozone dependency corresponding to EPFLUX.

For QBO we see that there is an obvious seasonality in the regression coefficients at the mid latitudes of both Hemispheres. For the QBO proxy at 10 hPa we characterize the seasonality with a cosine with its maximum in February, and its lowest value in August. This function aligns the seasonality of the southern mid-latitudes, and is opposite to the seasonality in the northern mid-latitudes. For the QBO at 30 hPa this pattern is shifted one month. Therefore we use a cosine with its peak in March and its lowest value in September to characterize the seasonality in the dependency with this variable. These alternative variables are, therefore, expected to affect ozone negatively at Northern mid-latitudes and positively at Southern mid-latitudes.

The regression coefficients of the ENSO proxy are different in the July-September months. We model this by a cosine of a yearly period with its lowest values in these months.

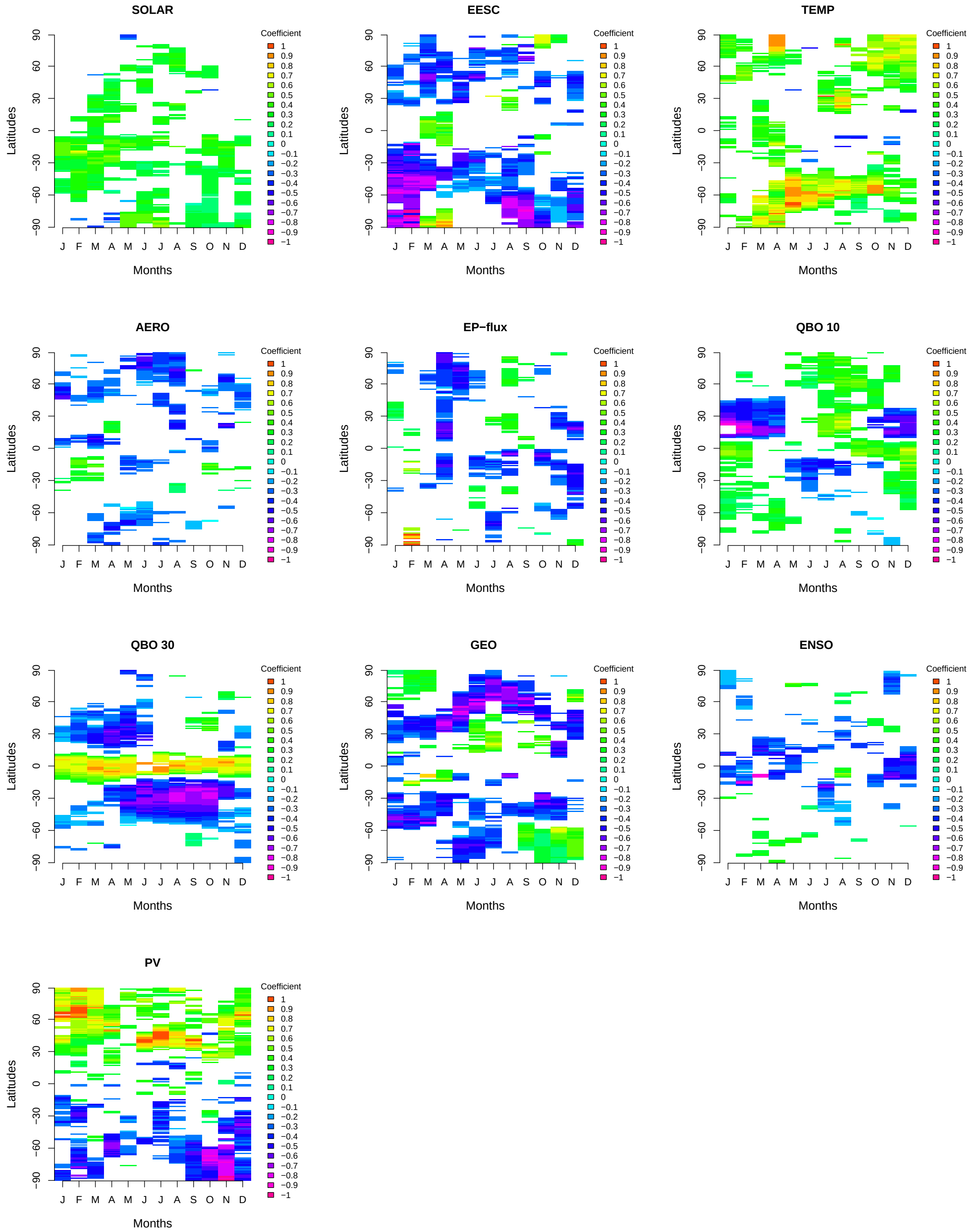


Figure 7: These plots show the coefficient estimate of each proxy of the regression for each month and latitude. White regions indicate non-significant estimation values.

Analyzing the regression coefficients corresponding to geopotential height, we again notice a seasonality in the influence on ozone throughout the year. Especially at the polar and mid-latitude regions figure 7 show a strong seasonality in the regression coefficients of this variable. We choose to model this seasonality with a sine of a yearly period with its peak in December and its lowest value in June.

According to figure 7, ozone has a strong seasonality in the dependence with the potential vorticity variable at the polar regions. We choose to model this seasonality as a sine with a yearly period and its maximum in June.

Based on the choices made above, we create the alternative variables as in equation 13. To emphasize the small short-time variations on the explanatory variables, they are first normalized with their standard deviations. The resulting time series is multiplied with the sines and cosines to account for the periodic dependencies.

$$\begin{aligned}
&X_{EESC_2}(t) = sin(\tfrac{2\cdot\pi\cdot(t-1)}{12}) \cdot X_{EESC}(t), && X_{TEMP_2}(t) = cos(\tfrac{2\cdot\pi\cdot(t)}{12}) \cdot X_{TEMP}(t),\\
&X_{EPFLUX_2}(t) = cos(\tfrac{2\cdot\pi\cdot(t)}{12}) \cdot X_{EPFLUX}(t), && X_{EPFLUX_3}(t) = cos(\tfrac{2\cdot\pi\cdot(t)}{6}) \cdot X_{EPFLUX}(t),\\
&X_{QBO10_2}(t) = cos(\tfrac{2\cdot\pi\cdot(t-1)}{12}) \cdot X_{QBO10}(t), && X_{QBO30_2}(t) = cos(\tfrac{2\cdot\pi\cdot(t-2)}{12}) \cdot X_{QBO30}(t),\\
&X_{ENSO_2}(t) = cos(\tfrac{2\cdot\pi\cdot(t-1)}{12}) \cdot X_{ENSO}(t), && X_{GEO_2}(t) = cos(\tfrac{2\cdot\pi\cdot(t)}{12}) \cdot X_{GEO}(t),\\
&X_{PV_2}(t) = cos(\tfrac{2\cdot\pi\cdot(t)}{12}) \cdot X_{PV}(t),
\end{aligned} \tag{13}$$

where t denotes time in months after January 1979 to 2010.

5 The local regressions

In this chapter we perform regressions independently on monthly time series on 1 by 1.5 degree grids.

5.1 Seasonality in ozone

For some regions, the ozone time series show a strong seasonal variation, as can be seen in figure 5. We use two different methods to model this variation. In the first method we model the seasonal component by including an harmonic series of sines and cosines with periods of a year and half a year in the model, like a Fourier series using only the most abundant frequencies. This model will be referred to as the statistical model throughout this analysis. Most previous studies, such as Stolarski et al. [1991], Brunner et al. [2006], Bodeker et al. [1998, 2001], use these harmonic series to account for the seasonality in ozone in their regression models. As is known from Fourier theory, these harmonic series are useful for extracting a periodic signal. However, when explanatory variables are included in the model which themselves contain a large seasonal component, this can cause problems due to possible high correlations between the harmonic series and the explanatory variable. In that case it is unclear whether the variability in ozone should be explained by the explanatory variable, or by the harmonic series. Therefore we choose not to include explanatory variables with a large seasonal component in the statistical model. Because the seasonal variation is accounted for by the Fourier term, this model is particularly suitable for studying the non-seasonal in ozone.

The second method to model the seasonality in ozone is to include variables in the regression model that affect ozone on seasonal time scales. Such a model will further be referred to as the physical model. This model will be used to analyze how the seasonality in ozone is driven by these explanatory variables. In this model, therefore, Fourier series are not included to account for the seasonality in ozone.

In addition we introduce a third model, where we combine the physical and statistical model. In this model all explanatory variables and the Fourier series are included. We refer to this model as 'the combined model'. Even though the variables in this model are highly correlated, interesting conclusions can be drawn from the overall performance of this model when we compare it to the performance of the statistical and physical model.

5.2 The regression models

First we define the statistical model:

$$Y(t) = a + \sum_{k=1}^{2} \beta_{2k-1} \cdot sin(2\pi t\frac{k}{12}) + \beta_{2k} cos(2\pi t\frac{k}{12}) + \sum_{j=5}^{m+4} \beta_j X_j(t) + \epsilon(t), \tag{14}$$

Where $Y(t)$ is the amount of ozone at time t, a the intercept value, β_j the coefficient of the corresponding harmonic function or explanatory variable X_j, t the time in months after January 1979 and $\epsilon(t)$ the noise at time t. The noise at different points in time are assumed to be uncorrelated and normally distributed with mean zero and variance σ^2. This model can also be written as

$$Y = X\beta + \epsilon, \tag{15}$$

with Y the vector representation of $Y(t)$, X a matrix with the four harmonic series and the proxies of included explanatory variables as columns and the intercept as a column of ones and ϵ the vector representation of $\epsilon(t)$.

As the set of explanatory variables in model 15 we take the QBO, SOLAR, AERO, ENSO and EESC variables and their corresponding alternative variables from section 4.3. Note that these variables do not contain a strong seasonal component.

The next model is the physical model where we include variables with a seasonal cycle to explain the seasonality in ozone physically. This model can also be written as equation 15, where X contains no Fourier series to account for the seasonality in ozone, but more proxies of explanatory variables X_j. In this model X consists of the complete collection of explanatory variables, their corresponding alternative variables and the intercept as a column of ones.

The combined model is again of the form of equation 15, where X consists of the complete collection of explanatory variables, their corresponding alternative variables, the four Fourier terms and the intercept as a column of ones.

An overview of the models and the included variables is given in table 3, where ‘Fourier’ denotes the four harmonic series with periods of a year and half a year.

Table 3: Table of the included proxies in the models

Model \Proxies	Intercept	Solar	EESC	EESC_2	AERO[20-25]	QBO10	QBO10_2	QBO30	QBO30_2	ENSO	ENSO_2
Statistical	X	X	X	X	X	X	X	X	X	X	X
Physical	X	X	X	X	X	X	X	X	X	X	X
Combined	X	X	X	X	X	X	X	X	X	X	X
Model \Proxies	TEMP	TEMP_2	EPFLUX	EPFLUX_2	EPFLUX_3	GEO	GEO_2	PV	PV_2	Fourier	
Statistical										X	
Physical	X	X	X	X	X	X	X	X	X		
Combined	X	X	X	X	X	X	X	X	X	X	

In the next subsection we discuss the method of estimation of the model regression coefficients.

5.3 Coefficient estimation and variable selection method

In the regression we do a backwards selection of variables, as explained in previous chapter. However, the estimation method is done slightly different in these regressions compared to the regressions performed in chapter 4. For each model we proceed with the following steps: in the first step we model ozone with one of the three linear regression models.

We want the ozone data values that have a large standard error to have less weight than the values with small standard deviation, because, especially in the polar regions, the standard error of the monthly means in the ozone proxy varies over time. We choose to give weight reciprocal to their standard errors

$$w(t) = \frac{1}{se_Y(t) \cdot \bar{se}_Y}, \tag{16}$$

where $w(t)$ is the weight given to $Y(t)$, $se_Y(t)$ the standard error of Y at month t and $\bar{se}_Y$ the mean standard error of Y. In our estimation procedure we account for these weights using the weighted least squares estimate for $\hat{\beta}$. We also estimate the noise vector and the covariance matrix of $\hat{\beta}$. The estimates are given by:

$$\begin{aligned} \hat{\beta} &= (X'WX)^{-1}X'WY, \\ \hat{\epsilon} &= \sqrt{w}(Y - X\hat{\beta}), \\ \widehat{Var}(\hat{\beta}) &= (X'WX)^{-1} \cdot \hat{\sigma}^2, \end{aligned} \tag{17}$$

where W the diagonal matrix with weights w on the diagonal and zero in all off-diagonal entries and $\hat{\sigma}$ the estimated standard deviation of the noise. From these regression estimates we calculate the t-statistics and P-values and follow the same variable selection procedure as performed in section 4.2, now with a significance value α of 0.01. We choose this lower value of α because in these regressions we are interested in significant regression coefficient estimates. To investigate how well the models describe the ozone variable, the R^2 value is calculated. This value is the fraction of variation in the ozone proxy that is explained by the explanatory variables.

$$R^2 = 1 - \frac{\sum\limits_t \hat{\epsilon}(t)^2}{\sum\limits_t (Y(t) - \bar{Y})^2}, \tag{18}$$

where $\bar{Y}$ denotes the mean value of Y.

5.4 Results

In this section we show the results of the regressions of the statistical and the physical model. For the regression estimates corresponding to SOLAR, EESC, AERO, QBO and ENSO and their alternative variables only the regression estimates of the statistical model are shown in figure 8, because the patterns in the regression estimates for the results of these variables in the physical model are similar. From the plots in figure 8 we see that solar radiation has a positive influence on ozone, especially in the Southern Hemisphere. It is surprising that the solar cycle does not have a significant effect on ozone in the Northern Hemisphere. Figure 8 shows a negative effect of EESC on ozone at the polar regions, especially at the South Pole. This is an expected result. The corresponding alternative variable also has a strong effect at the South Pole. This variable captures the seasonal effect that ozone is strongly depleted by chemical reactions involving EESC in the spring months of the corresponding Hemisphere. This is a main cause of the ozone hole that appears in this period. Figure 8 shows a negative effect of aerosols on ozone at high latitudes, especially at the North Pole. This is caused by the Pinatubo volcanic eruption. The aerosols from this eruption were mainly located at high northern latitudes. Both QBO proxies have a large effect on ozone at the equator and mid-latitudes. It seems that the QBO westerly winds increase the amount of ozone at the equator, and decrease the amount of ozone at the mid-latitudes. Also the corresponding alternative variables show clear patterns. Obviously, these variables have an opposite effect on ozone at the northern mid-latitudes with respect to the southern mid-latitudes. This is caused by the opposite seasons in the QBO dependency of ozone between the Hemispheres, as was previously seen in figure 7. ENSO has a negative effect on ozone at the pacific and a positive effect at higher latitudes. This indicates a flow of ozone from the stratosphere above the pacific to regions of higher latitudes in the stratosphere due to El Nino. ENSO's alternative variable has an negative influence at the equator and positive influence at mid-latitudes. This pattern is in agreement with the expected pattern, considering figure 7.

For the regression coefficients corresponding to EPFLUX, TEMP, GEO, PV and their alternative variables the regression estimates of the physical model are shown in figure 9. These are the variables that contribute to the seasonality in ozone. In general we see that the variables of figure 9 have a large effect on ozone at the polar regions, and a small effect at the equator. This is in agreement with the large seasonal component of ozone at the Poles, compared to the seasonality of ozone at the tropics. In addition, note that these explanatory variables contribute more to the variability in ozone than the explanatory variables from figure 8, as is seen in the difference in color-bar values. The positive regression coefficient estimates of the TEMP proxy in figure 9 at the Poles confirm the theory that low temperature at polar regions has a negative effect on ozone. The corresponding alternative variable has its main effect at the South Pole. However, the sign of these estimates differs from what we expected based on figure 7. The effect of the EP-flux on ozone at the Poles is opposite to its effect on ozone at mid-latitudes. This is in agreement with the phenomena that the polar vortex hinders air exchange between the Poles and mid-latitudes. Surprisingly, the effect of the northern EP-flux is opposite to the effect of the southern EP-flux. Also in the plots corresponding to the alternative variables show the polar vortex at about 60 degrees in latitude on both Hemispheres. However, from figure 7 we expected EPFLUX_3 to be included at the Northern Hemisphere and EPFLUX_2 at the Southern Hemisphere. Roughly the opposite is the case. Geopotential height has most variability at the mid-latitudes. Therefore, it is not surprising that the geopotential height has its effect on ozone mostly at these latitudes. The negative effect on ozone is in agreement with the theory. The corresponding alternative variable roughly shows the patterns that were expected based on figure 7. The regression coefficient estimates corresponding to the potential vorticity in the Southern Hemisphere are opposite in sign to the estimates at the Northern Hemisphere. This is caused by the opposite sign of the potential vorticity itself between the Hemispheres. For the effect on ozone, the direction of the rotation of air is not important, but the intensity of the rotation is. Since the potential vorticity is positive

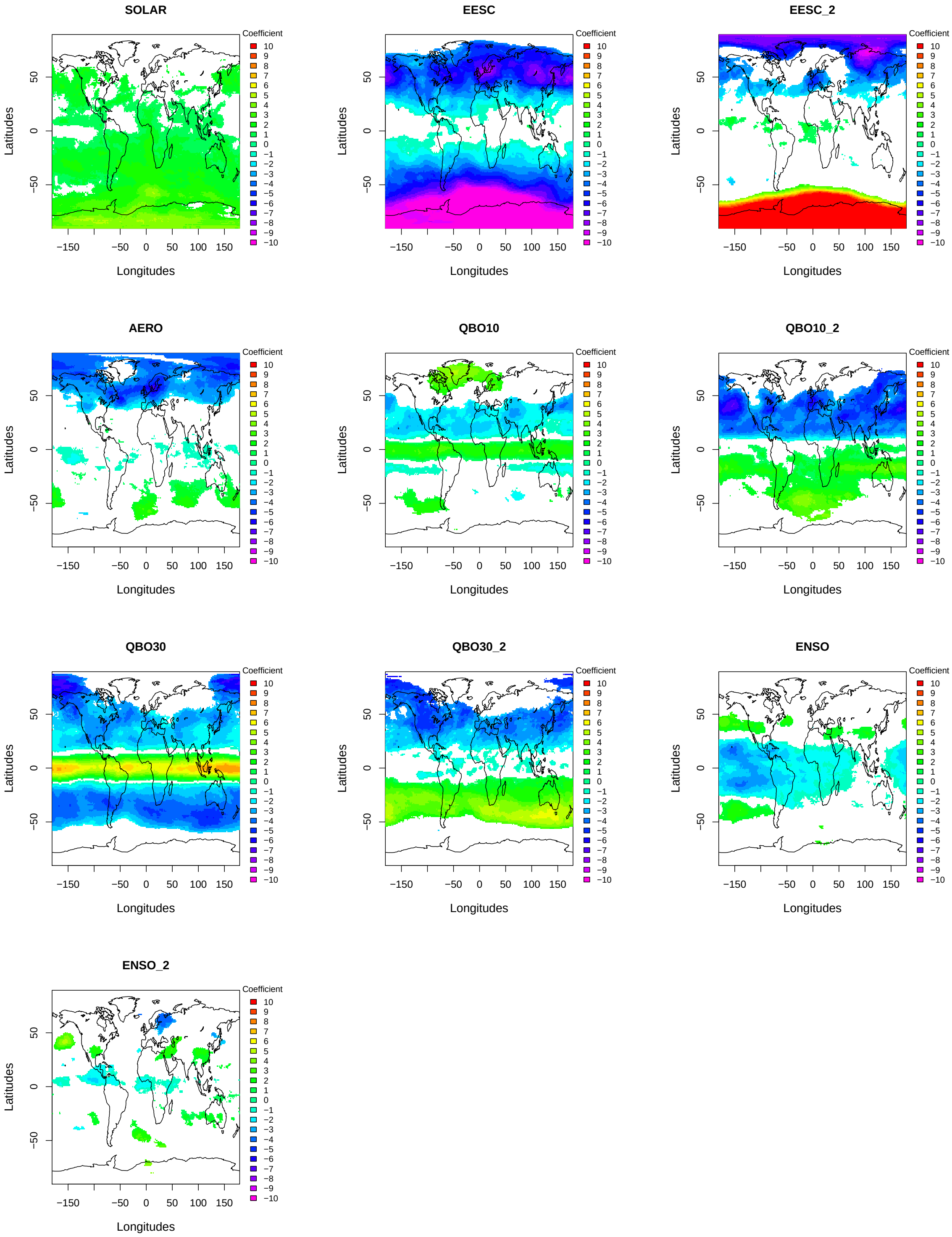


Figure 8: These plots show the regression coefficient estimates of the explanatory variables included in the statistical model. White regions indicate non-significant coefficient estimates.

at the Northern Hemisphere, we conclude that the rotation of air masses has a positive effect on ozone. The alternative variable has an effect at the North Pole and South Pole.

In figure 10 the regression results of three specific locations on the earth for both the physical and the

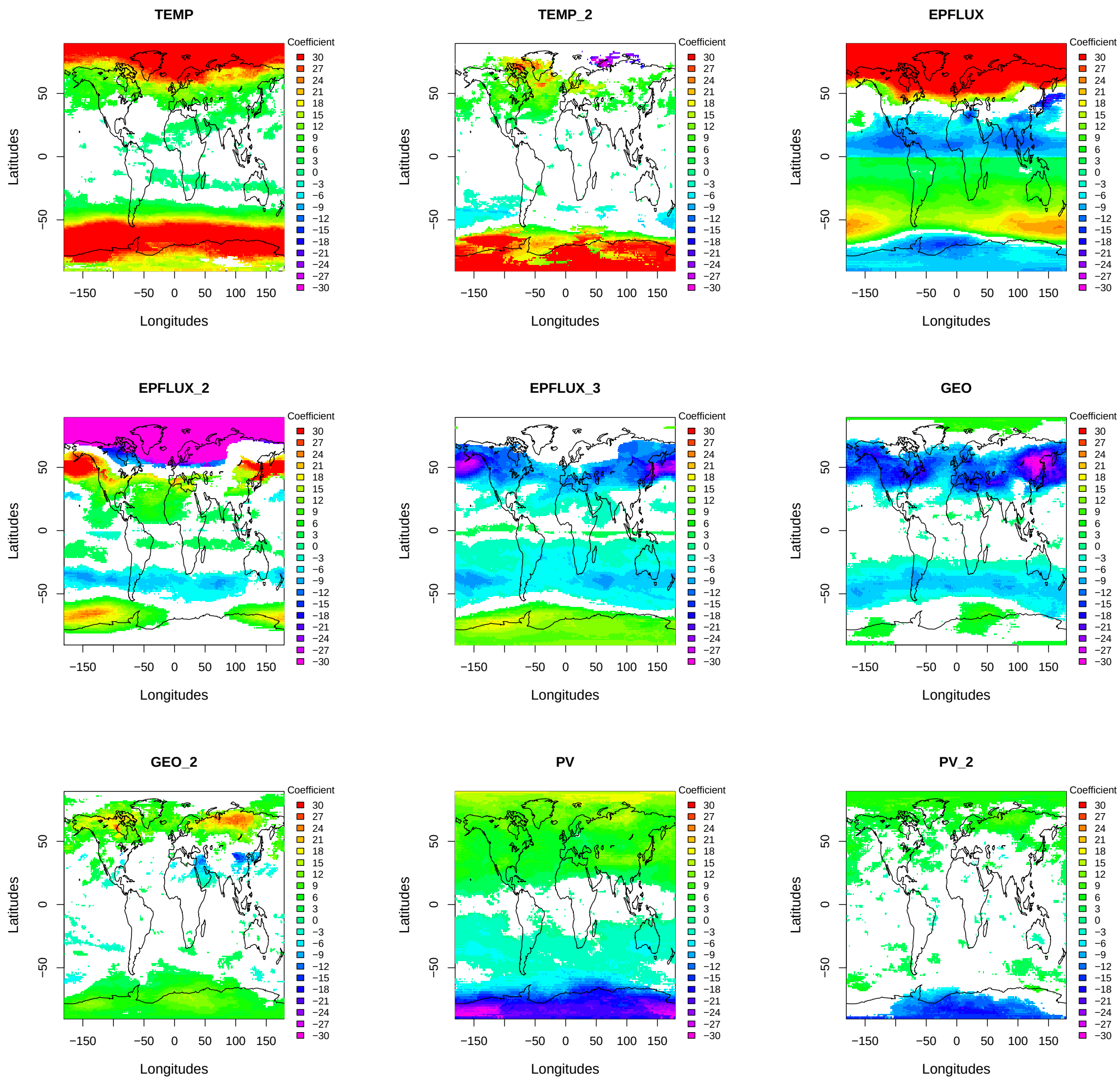


Figure 9: These plots show the regression coefficient estimates of the explanatory variables included in the physical model. White regions indicate non-significant coefficient estimates.

statistical model are shown. For this figure Reykjavik is chosen to show the contribution of aerosols and to show how the seasonality in ozone at high northern latitudes is explained for the physical and the statistical model, respectively. Here, the aerosols have a significant effect on ozone according to both regression models, and the seasonality in ozone is explained by TEMP, EPFLUX, GEO and GEO_2. Furthermore, we notice that the contribution of QBO10 and ENSO are found significant in the statistical model, but are excluded in the physical model. Also, note that the residuals in the physical model still contain a seasonal effect, while the seasonality of ozone is completely explained by the Fourier terms in the statistical model.

Bogota is chosen to show the effect of the QBO and ENSO variables on ozone. We clearly see a cycle of roughly 2 years in the ozone time series, which is explained by the QBO variables. The ENSO proxies have a small but significant effect on ozone, according to the results of both models. The small seasonality in the ozone time series is explained by the EPFLUX and GEO_2 proxies.

Finally, we chose a location on the Antarctic to visualize the effect of the EESC on ozone and the seasonality of ozone at southern latitudes. The decrease in the ozone time series follows the EESC curve nicely, and therefore the EESC is included in both models. However, we notice that EESC has an even larger effect on the seasonal

Figure 10: The left plots show the influence of the explanatory variables in the regressions of the physical model. The right plots show the influence of the explanatory variables in the regressions of the statistical model. This influence is showed in terms of DU. The Fourier term is calculated as the sum of the effect of the harmonic series in the statistical model. For each location, the time series have the same vertical scaling.

minimum of ozone. EESC_2 explains this expanding seasonal effect on ozone. Furthermore, the ENSO and SOLAR proxies were included in the statistical model and not in the physical model. For the ENSO variable, we see that TEMP proxy captures the effect of the El Nino in 1998 by the high values in that year.

6 Conclusions and discussion

6.1 Discussion on ozone depletion

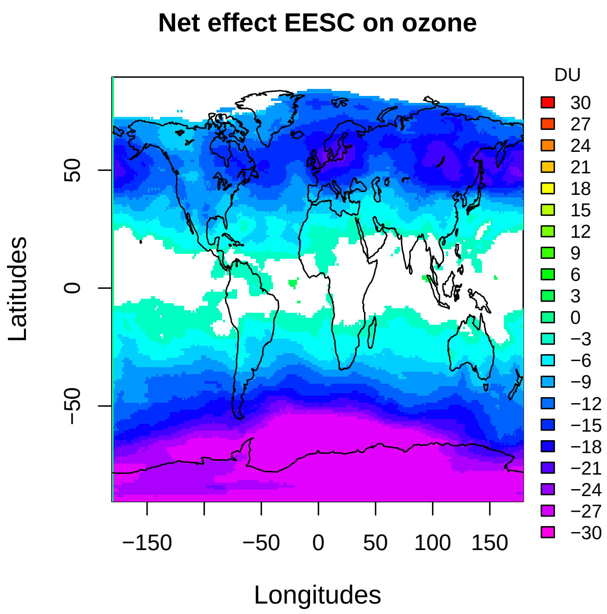


Figure 11: This plot shows the amount of effect of EESC on ozone between 1979 and 2010 in terms of DU.

An interesting explanatory variable is the EESC, because it represents the direct effect that the ODS have on ozone. From the regression coefficients we calculate the depletion of ozone due to the EESC in terms of Dobson units by

$$\Delta_{O_3} = \beta_{EESC} \cdot \Delta_{EESC}, \tag{19}$$

where Δ_j is defined as the difference in the anomaly of variable j between 1979 and 2010. The Δ_{O_3} are shown in figure 11. From this figure we see that the EESC results in a regional decrease of ozone up to 40 DU in 2010. The plots of the Antarctic in figure 10 show the regression results in a region where such a decrease is detected. From this plot we conclude that the ozone layer still suffers from the amount of ozone depleting substances in the stratosphere, although this amount has been shrinking since the Montreal protocol as is seen in the curve in the EESC variable. In addition, we conclude that the effect of the ODS is even larger in the August-November months at the antarctic, and that the amount of depletion in these months again follow the EESC curve. This is an indication that the ODS indeed are an important cause for the yearly appearing ozone hole at the Antarctic.

The curve in the EESC does indicate a recovery in the amount of ODS in the stratosphere. This is hopeful for the future ozone layer.

6.2 Model discussion

In this section we discuss the results and properties of the regression models used in this study.

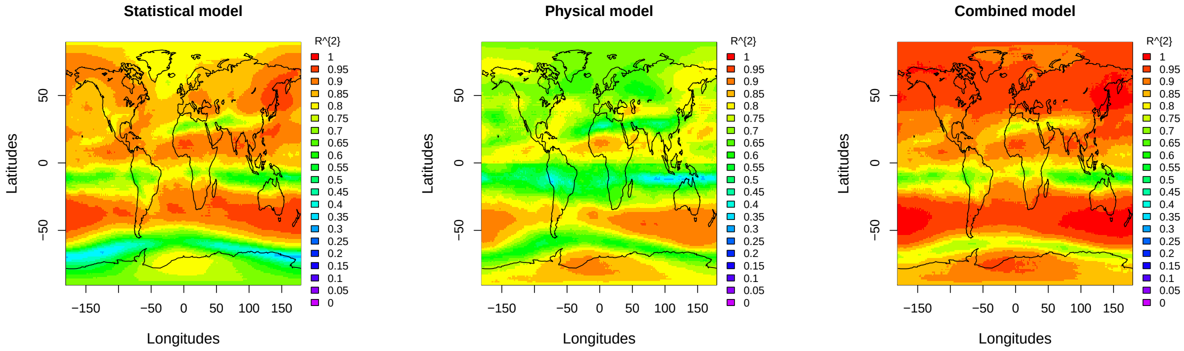


Figure 12: These plots show the explained variance of the models in terms of R^2.

First, we discuss the inclusion of alternative variables in our models to describe the seasonalities in effect on ozone. In the statistical model alternative variables were used for the EESC, QBO and ENSO proxies based on figure 7. The results of these alternative variables in figure 8 show convincing spatial patterns that were expected from our analysis in seasonality, performed in chapter 4. In the physical model we used the alternative variables corresponding to the EPFLUX, TEMP, GEO and PV, again based on the results in figure 7. However, the spatial patterns in the regression coefficient estimates of the alternative variables for TEMP and EPFLUX figure 9 do not show the expected patterns. The reason for these patterns are not understood and require more analysis.

The statistical model has the advantage with respect to the other models that the correlation values between the explanatory variables, including the harmonic series, are low. Therefore, we are certain that the variability explained by the included variables in this model is attributed to the correct variables. The problem with this model is, however, that we do not gain any insight in how the seasonal component of ozone is driven by explanatory variables. This model is useful when one is interested in non-seasonal effects on ozone.

The physical model includes more explanatory variables to describe the seasonal component in ozone. A problem with these variables is that they are highly correlated in some regions. Therefore, it is less certain whether the explained variability is attributed to the correct explanatory variable. The northern EP-flux and the stratospheric temperature, for example, are highly correlated at the North Pole, as is shown is figure 6. Both variables are selected by the backwards selection of variables as significantly important in the model. It is questionable whether the variance explained by these variables in this region should be attributed to both explanatory variables in this ratio. However, the obtained results are as expected. For example, the geopotential height is included mostly at the mid-latitudes, whereas the EP-flux and temperature are included at the Poles. Even though these variables are highly correlated, on each location, the backwards selection method chooses the variables in the model that are expected to affect on ozone. Nevertheless, we have to take the correlations in this model into account when drawing conclusions.

Furthermore, from figure 12 we see that, with the exception of the south polar region, in terms of explained variance the performance of the statistical model is better than the performance of the physical model. From this we conclude that the harmonic series used in the statistical model describes the seasonal component in ozone better than the EP-flux, stratospheric temperature, geopotential height, potential vorticity and their corresponding alternative variables together. Apparently the physical model not only has problems with collinearity, we also conclude that the included variables in the physical model do not fully describe the seasonal component in ozone. This is also seen in the residuals of the Reykjavik regressions shown in figure 10.

The results of the combined model show a high performance in terms of explained variance, as can be seen in figure 12. However, as we remarked before, some explanatory variables in this model are highly correlated, which makes interpreting the results more difficult. Figure 12 shows patterns that can help in the search for ozone related variables to further improve the regression models. The combined model has lower performance in the zonal bands at -10 and -60 degrees latitudes and above the Sahara and south-west Asia. The zonal band at -60 degrees shows the southern polar vortex. Because the exact position of the vortex is not fixed in time, and has a large effect on the amount of ozone, it is difficult to incorporate this effect in this region. For other regions of low explained variance, however, the reasons are not understood and require more detailed analysis.

6.3 Comparison to other studies

Previous ozone regression studies, such as Stolarski et al. [1991], Brunner et al. [2006] and Bodeker et al. [1998, 2001] perform regressions on seasonality in dependencies, as we did in section 4. In Brunner et al. [2006] the regression results are presented in similar plots as in section 4, where they show only results on the QBO, aerosols, solar cycle and EP-flux variables. These plots are shown in figure 13. The results corresponding to the QBO and aerosols are very similar, but the results on EP-flux and solar intensity are slightly different. We think this is caused by the difference in the data-set that is used for the ozone variable. Brunner used ozone data spanning the time period 1979-2004.

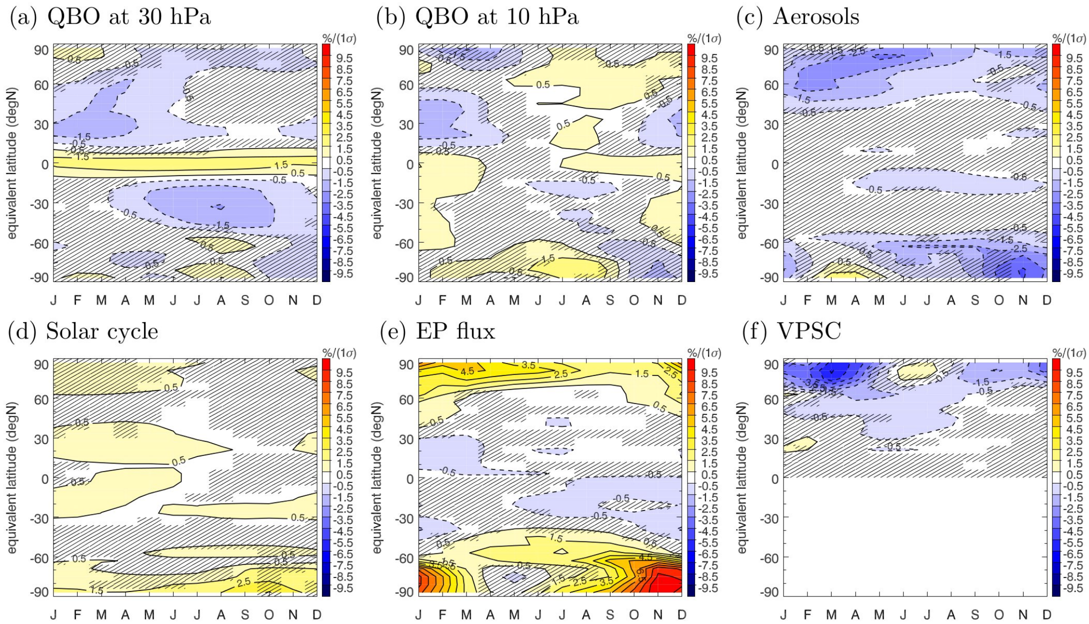


Figure 13: These plots show the regression coefficients of the seasonal regression analysis performed by Brunner. Source: Brunner et al. [2006].

This study is unique in that the regressions are performed per grid cell. The advantage is that local and regional conclusions can be drawn from these results, whereas regressions on ground-based measurements, as performed by Wohltmann et al. [2007], are performed on specific locations, and the regressions in the studies mentioned above are performed on equivalent latitude values of ozone, thus can only be interpreted for the latitude bands as a whole.

In addition, the regression method in this study contains an algorithm for selecting the significant explanatory variables in the regression model. A similar method is used in Mäder et al. [2007] to list explanatory variables in the order of its importance to the model. However, the studies mentioned above did not use this approach, but included all available variables in the regressions.

Furthermore, this study tried to model the seasonal component in ozone using the physical model, where the previous studies used harmonic series to account for the seasonal component. However, we did not achieve satisfying results with this physical model and encountered problems due to high correlation values between the explanatory variables included in this model.

6.4 Suggestions for further research

This study tried to model the seasonal component in ozone with the physical model. This model did not entirely succeed in explaining the seasonality in ozone. This model has problems due to correlations in the explanatory variables. A better understanding of the driving factors behind the seasonal variation in ozone is needed to improve this model.

Furthermore, figure 12 shows regions where the performance of all three models can be improved. Further research on the driving factors behind ozone in these regions is needed to find explanations for these features, or may find variables that explain the unexplained variance in the ozone time series for these regions.

References

[1] van der A, R. J., Allaart, M. A. F. and Eskes, H. J.:*Multi sensor reanalysis of total ozone.* Atmos. Chem. Phys. Discuss., 10,11401-11448, 2010.

[2] Allaart, M. A. F., Kelder, H. and Heijboer, L. C.:*On the relation between ozone and potential vorticity.* Geophys. Res. Lett., 20,811-814, 1993.

[3] Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I. Horinouchi, T., Jones, D. B. A., Kinnersley, J. S. Marquardt, C., Sato, K. and Takahashi, M., G.L.:*The quasi-biennial oscillation.* Rev. of Geophys., 39, 179-229, 2001.

[4] Bodeker, G. E., Boyd, I. S. and Matthews, W. A.:*Trends and variability in vertical ozone and temperature profiles measured by ozonesondes at Lauder, New Zealand: 1986-1996.* J. Geophys. Res., 103,D22,661-681, 1998.

[5] Bodeker, G. E., Scott, J. C., Kreher, K. and Mckenzie, R. L.:*Global ozone trends in potential vorticity coordinates using TOMS and GOME intercompared against the Dobson network: 1978-1998.* J. Geophys. Res., 106,D19,29-42, 2001.

[6] Brunner, D., Staehelin, J., Maeder, J.A., Wohltmann, I. and Bodeker, G.E.:*Variability and trends in total and vertically resolved stratospheric ozone based on the CATO ozone data set.* Atmos. Chem. Phys., 6,4985-5008, 2006.

[7] Eskes, H. J., van der A, R. J., Brinksma, E. J., Veefkind, J. P., de Haan, J. F., and Valks, P. J. M.:*Retrieval an validation of ozone columns derived from measurements of SCIAMACHY on Envisat.* Atmos. Chem. Phys. Discuss., 5,4429-4475, 2005.

[8] Frohlich, C.,:*Observations of Irradiance Variations.* Space Sci. Rev., 94,15-24, 2000.

[9] Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.K., Hnilo, J. J., Fiorino, M., and Potter, G.L.:*NCEP-DOE AMIP-11 REANALYSIS (R-2).* Bull. Amer. Meteor. Soc., 83,1631-1643, 2006.

[10] Mäder, J. A., Staehelin, J., Brunner, D., Stahel, W. A., Wohltmann, I. and Peter, T.:*Statistical modeling of total ozone: Selection of appropiate explanatory variables.* J. Geophys. Res., 112,D11108, doi:10.1029/2006JD007694, 2007.

[11] Newman, P., Nash, E. R. and Kawa, S.R.:*When will the Antarctic ozone hole recover?.* Geophys. Res. Lett., 33,L12814,doi:10.1029/2005GL025232, 2006.

[12] Riishøjgaard, L. P. and Källen, E.:*On the correlation between ozone and potential vorticity for large scale Rossby waves.* J. Geophys. Res., 102,8793-8804, 1997.

[13] Sato, M., Hansen, J.E., McCormick, M.P. and Pollack, J.B.:*Stratospheric aerosol optical depth. 1850-1990.* J. Geophys. Res., 98, 22987-22994, 1993.

[14] Stolarski, R. S., Bloomfield, P. and Mcpeters, R. D.:*Total ozone trends deduced from NIMBUS 7 TOMS data.* Geophys. Res. Lett., 18,1015-1018, 1991.

[15] Wolter, K., Timlin, M.S.:*Monitoring ENSO in COADS with a seasonally adjusted principal component index..* Proc. of the 17th Climate Diagnostics Workshop, 52-57, 1993.

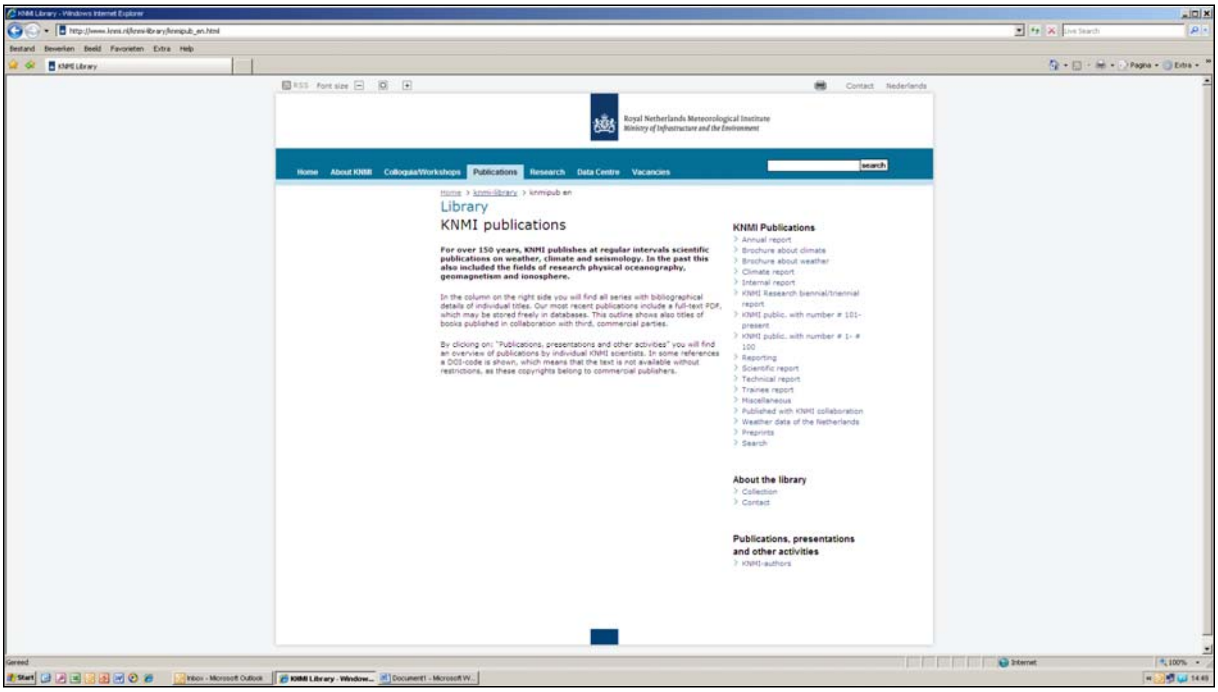
[16] Wolter, K., Timlin, M.S.:*Measuring the strength of ENSO events - how does 1997/98 rank?.* Weather, 53, 315-324, 1998.

[17] WMO,:*Scientific Assessment of Ozone Depletion: 1994.* WMO Glob. Ozone Res. and Mon. Proj., Report No. 37 1995.

[18] Wohltmann, I., Lehmann, R., Rex, M., Brunner, D., Mäder, J.A.,:*A process-oriented regression model for column ozone.* J. Geophys. Res., 112,D12304, doi:10.1029/2006JD007573, 2007.

A complete list of all KNMI-publications (1854 – present) can be found on our website

www.knmi.nl/knmi-library/knmipub_en.html



The most recent reports are available as a PDF on this site.