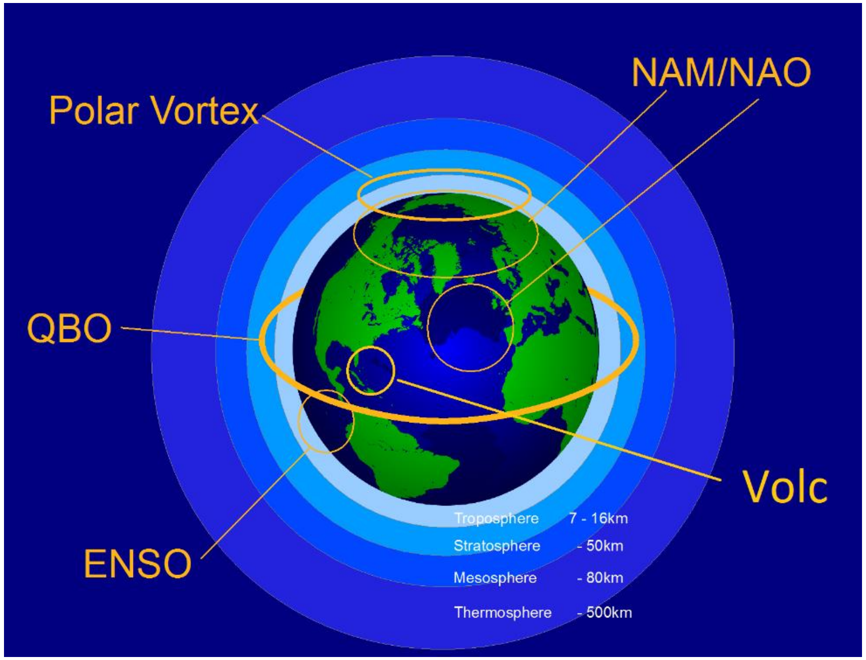


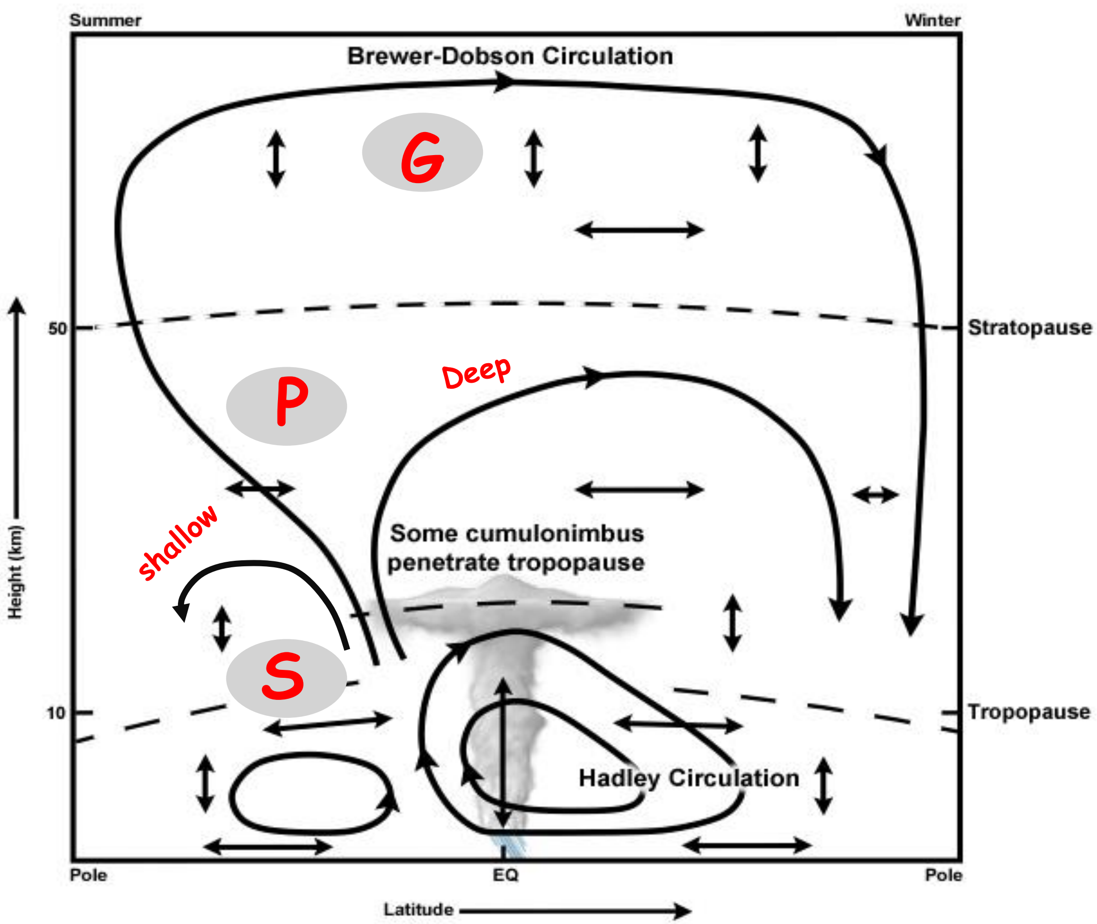
NATURAL CLIMATE VARIABILITY AND STATISTICAL ANALYSIS APPROACH

MOHAMADOU DIALLO

May 29th 2020 | Atmospheric modeling lecture | University of Wuppertal, Germany

JÜLICH Forschungszentrum

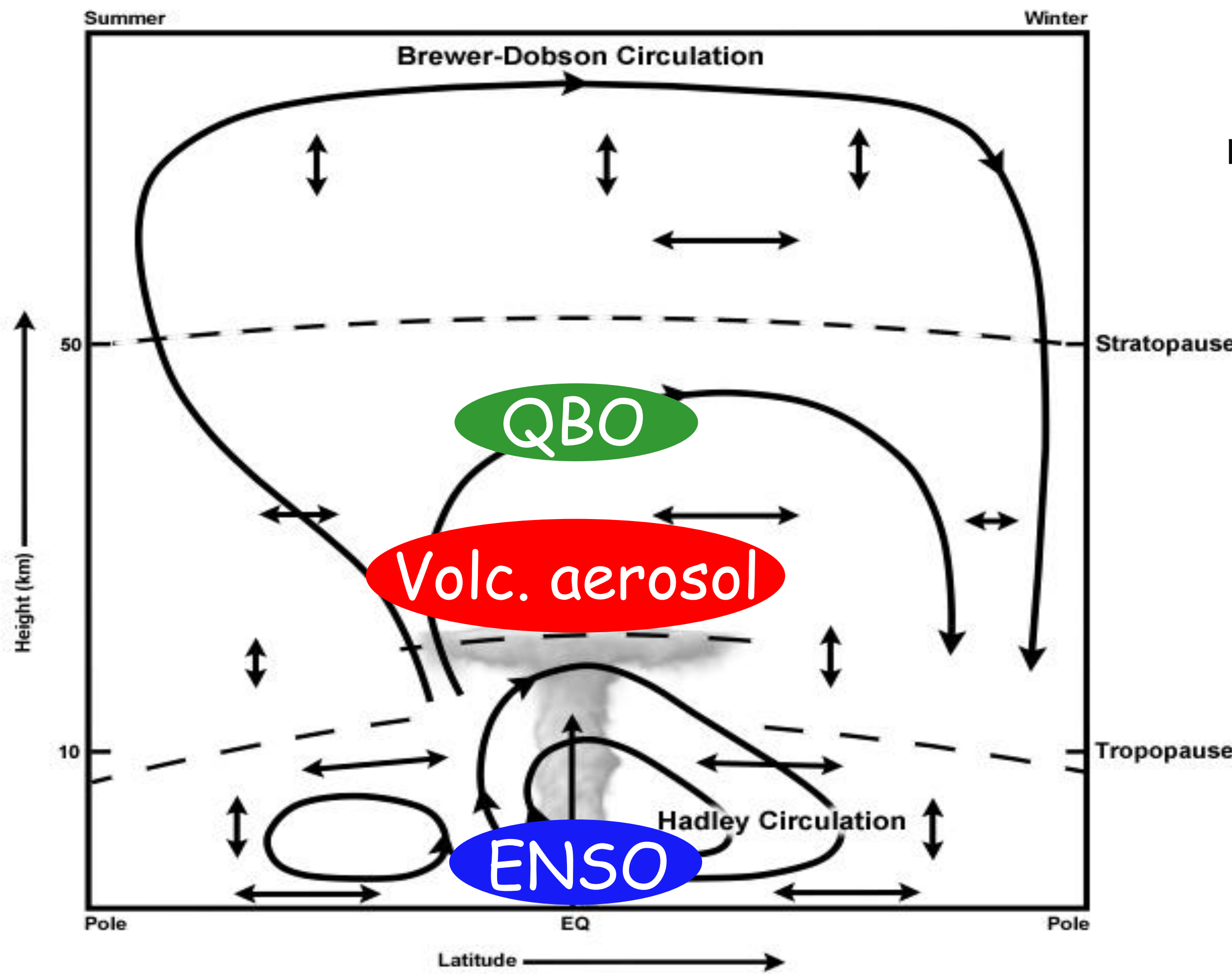
What is the Brewer-Dobson Circulation (BDC)?



[WMO, 1985]

- **BDC** is the transport of air mass from the inner tropics to the higher latitudes via **transition**, **shallow** and **deep** branches [Brewer, 1949].
- **BDC** is driven by the **waves breaking** [Plumb, 2002]:
  - ✓ Gravity waves (G)
  - ✓ Planetary-scale waves (P)
  - ✓ Synoptic-scale waves (S)

Modulations of the BDC by natural variability



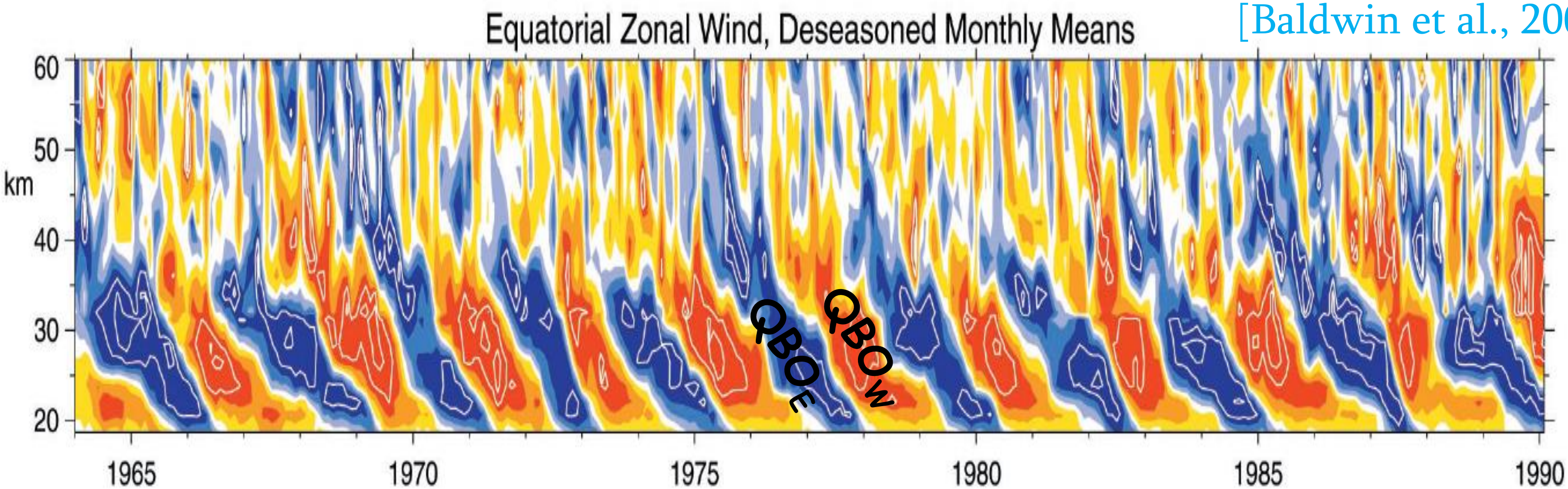
[WMO, 1985]

BDC is modulated by naturally varying signals:

- **Quasi-Biennial Oscillation phases (QBO)** [Plumb & Bell, QJRMS 1982].
- **El Nino-Southern Oscillation (ENSO)** [Randel et al., GRL 2009].
- **Volcanic aerosols** [Robock, GRL 2010].

The Quasi-Biennial Oscillation (QBO)

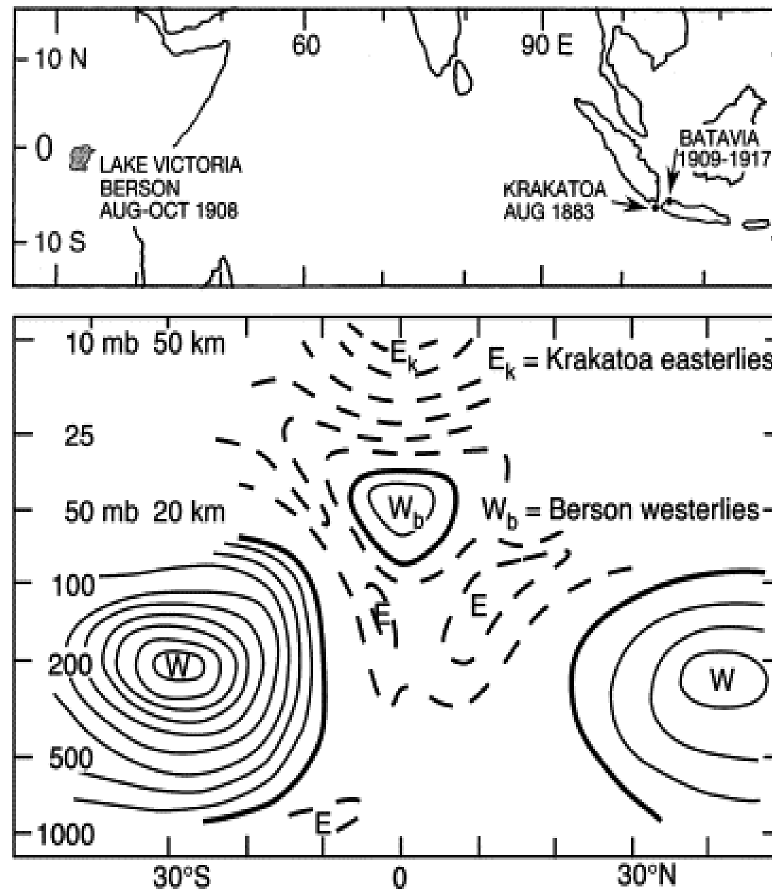
[Baldwin et al., 2001]



- ✓ The QBO is the mean 28 months oscillation of zonal mean winds.
- ✓ The QBO propagates downward into the lower stratosphere at 1 km/month.
- ✓ Transition between QBO_W and QBO_E regimes is often delayed between 30 and 50 hPa.
- ✓ Both the QBO period and amplitude exhibit considerable variability.

Discovery of the QBO

- ✓ Global spread of Krakatao volcanic plime in Aug 1883 led to a discovery of Krakatau-easterlies at 10hPa.
- ✓ In 1910, the discover of Berson-westerly, leading to a conundrum of the Krakatau-easterlies.
- ✓ In 1924, Van Bemmelen confirmed the existing Krakatau-easterlies.
- ✓ Alternating $QBO_{W/E}$ was fully described in 1960s by Edmon and Reed.
- ✓ Westerly momentum induced by equatorially trapped Kelvin waves and easterly momentum by Rossby-gravity [Holton & Lindzen, 1972; Plumb, 1977].



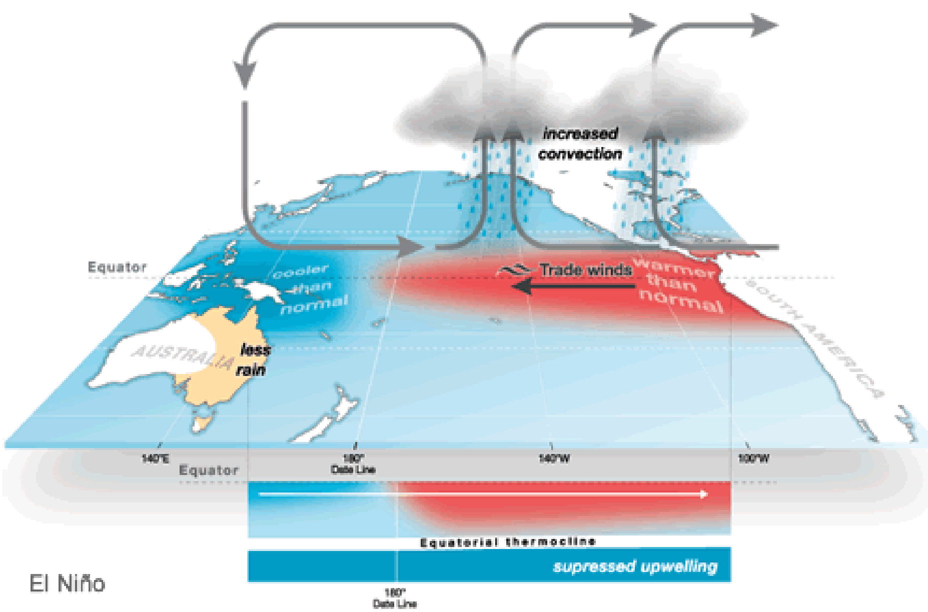
[Lettau, 1956]

The El Niño Southern Oscillation (ENSO)

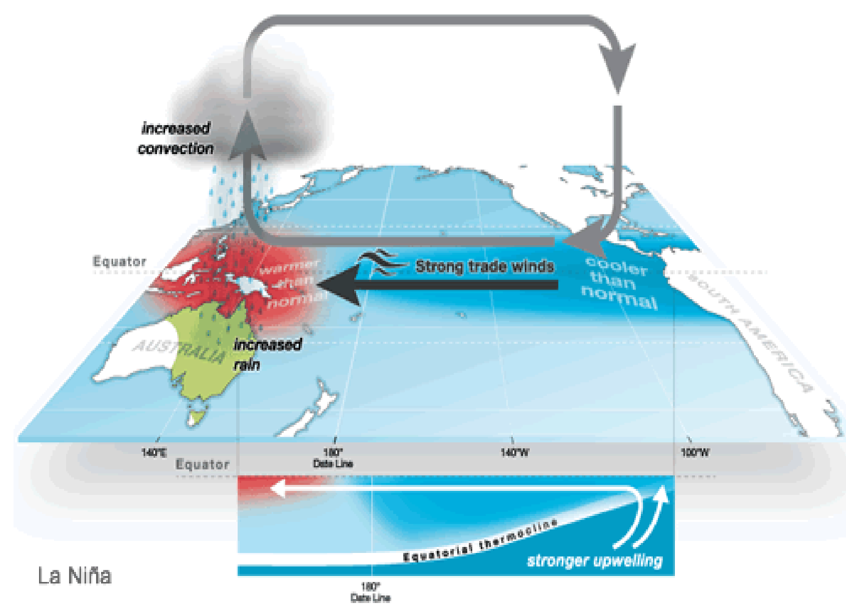
- ✓ ENSO is a coupled ocean-atmosphere phenomenon caused by recurring redistributions of heat and atmospheric momentum in the tropical Pacific [McPhaden, 2002].

The El Niño Southern Oscillation (ENSO)

- ✓ ENSO is a coupled ocean-atmosphere phenomenon caused by recurring redistributions of heat and atmospheric momentum in the tropical Pacific [McPhaden, 2002].
- ✓ The extremes of ENSO, termed El Niño and La Niña, encompass a wide range of climate conditions [Philander, 1990].

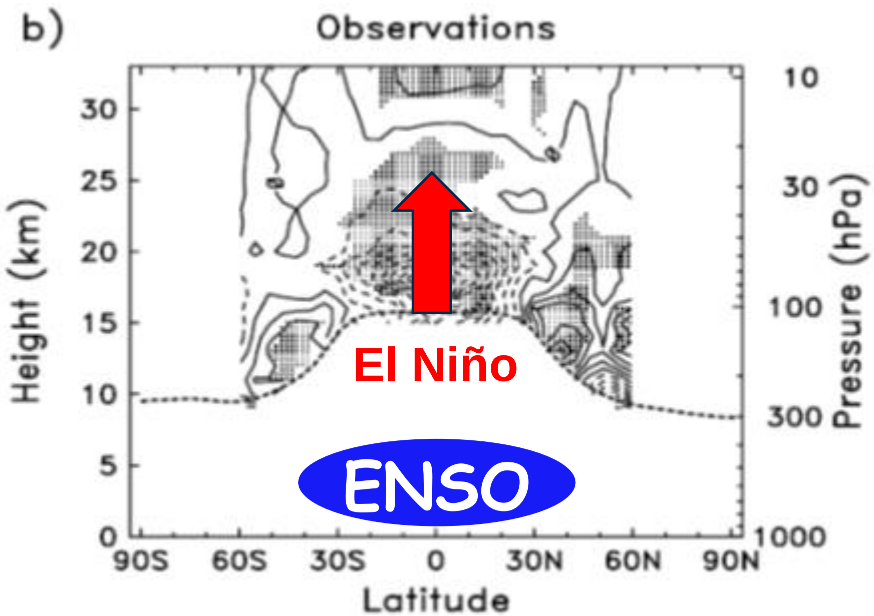


El Niño (anomalously warm SSTs)



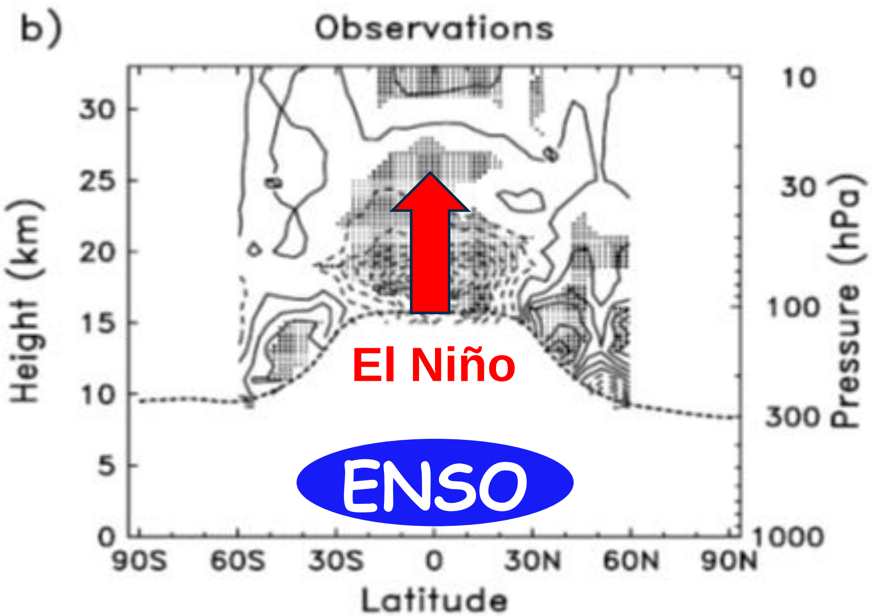
La Niña (anomalously cool SSTs)

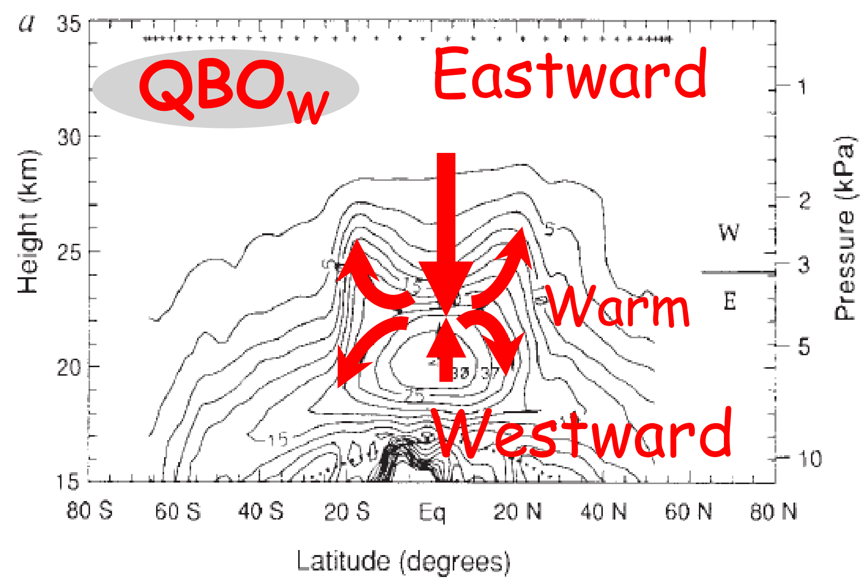
Well-known QBO & ENSO modulations of BDC



- **El Niño enhances the tropical upwelling** [Randel et al., 2009; Calvo et al. 2010; Konopka et al., 2016; Diallo et al., 2018].

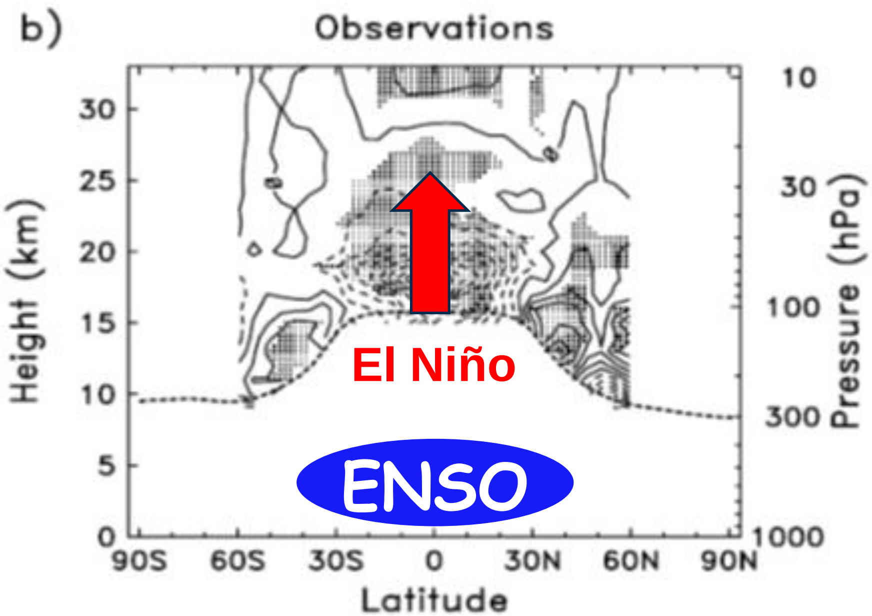
Well-known QBO & ENSO modulations of BDC

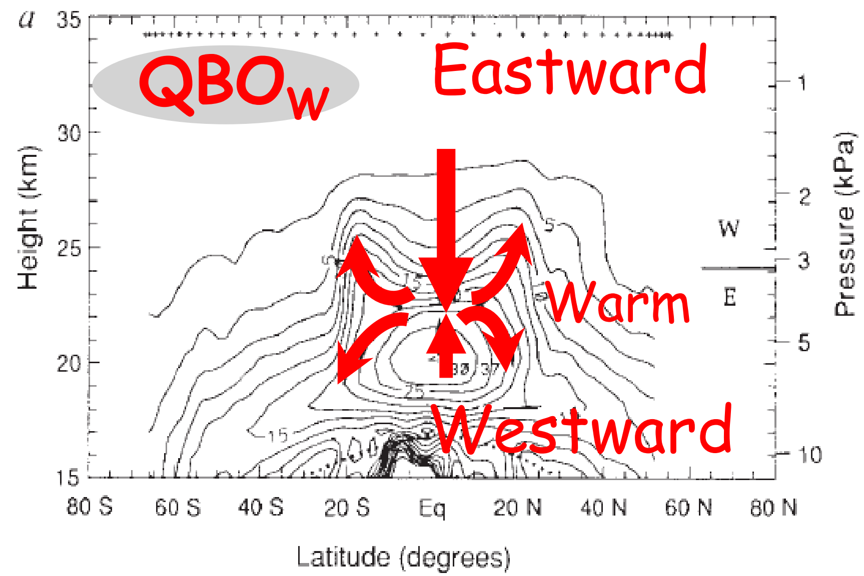


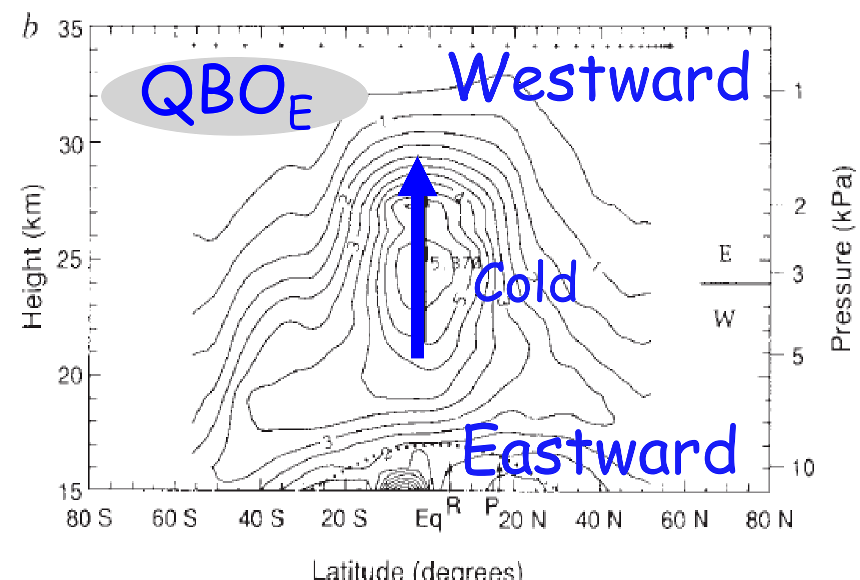


- El Niño enhances the tropical upwelling [Randel et al., 2009; Calvo et al. 2010; Konopka et al., 2016; Diallo et al., 2018].
- QBO$_W$ reduces tropical upwelling but enhances horizontal transport.
- QBO$_E$ enhances tropical upwelling and anomalous cold tropopause [Plumb & Bell, 1982; Trepte et al., 1992]

Well-known QBO & ENSO modulations of BDC

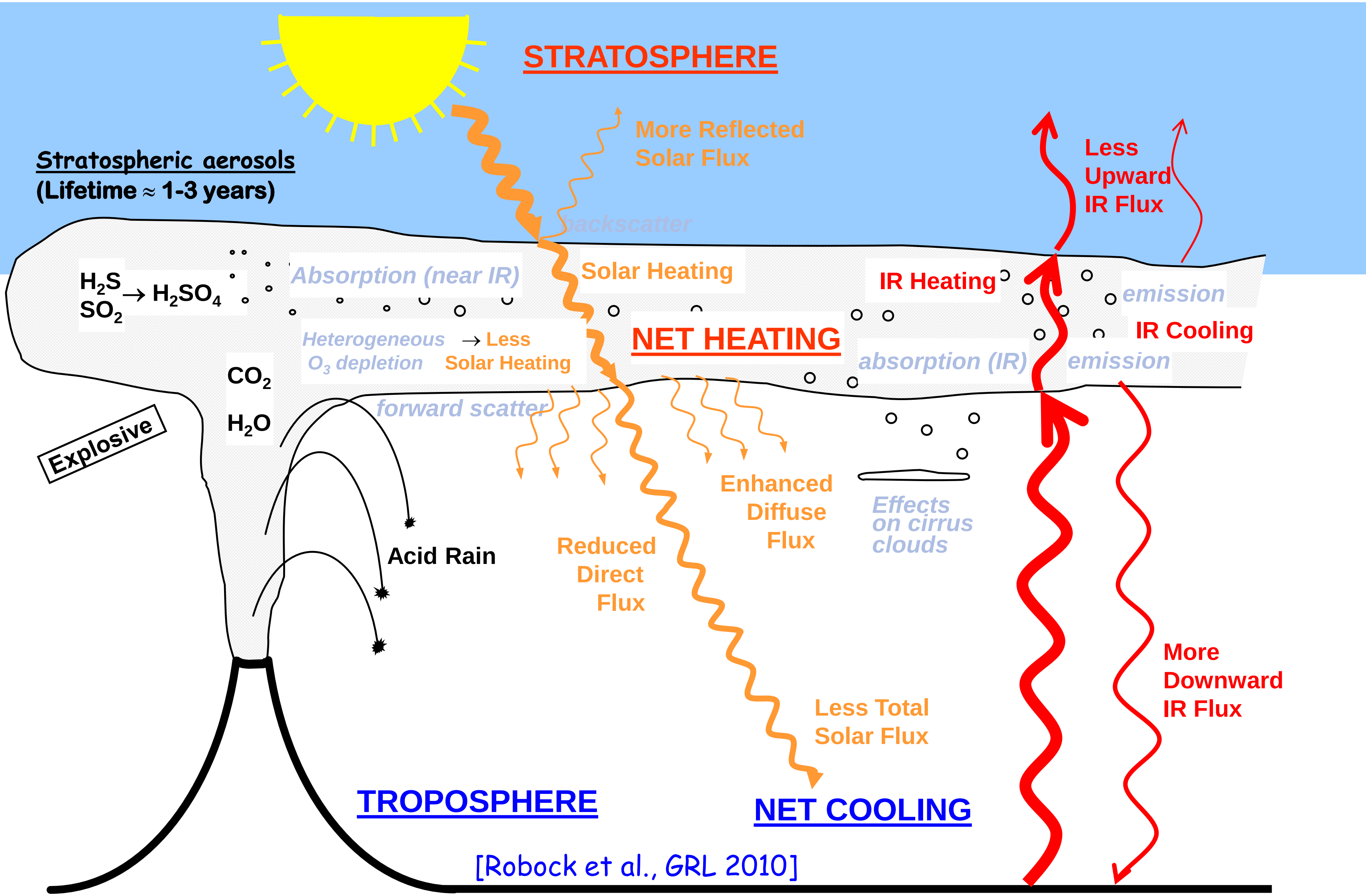






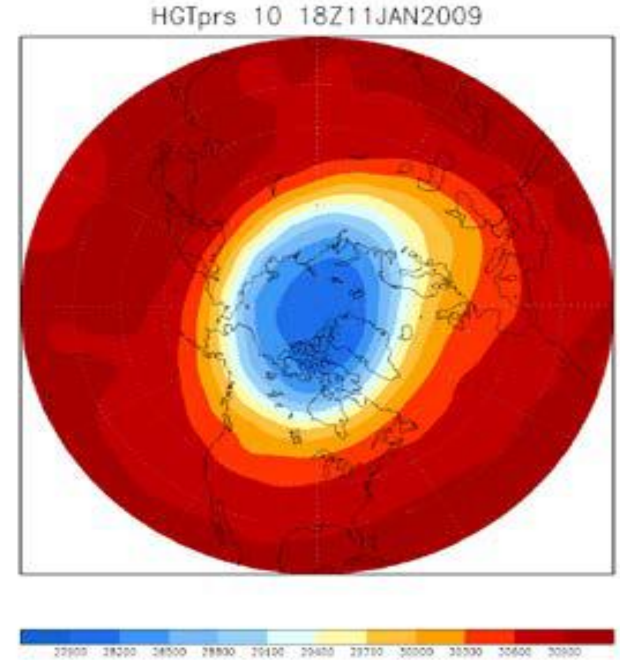
- El Niño enhances the tropical upwelling [Randel et al., 2009; Calvo et al. 2010; Konopka et al., 2016; Diallo et al., 2018].
- QBO$_W$ reduces tropical upwelling but enhances horizontal transport.
- QBO$_E$ enhances tropical upwelling and anomalous cold tropopause [Plumb & Bell, 1982; Trepte et al., 1992]

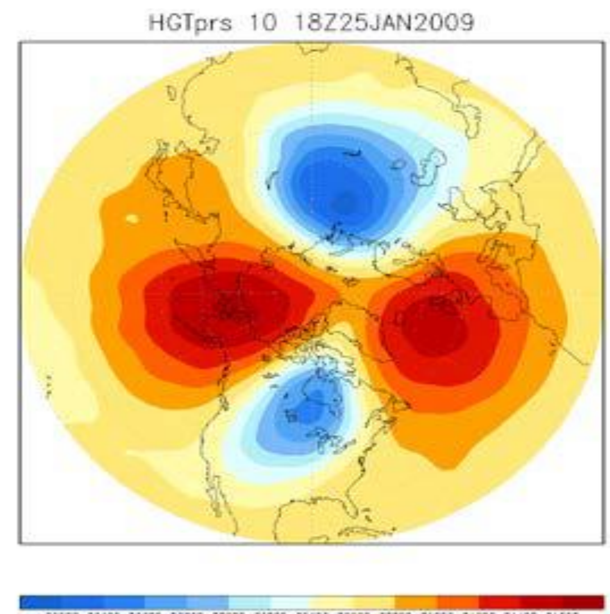
Radiative forcing of the Junge Layer and BDC

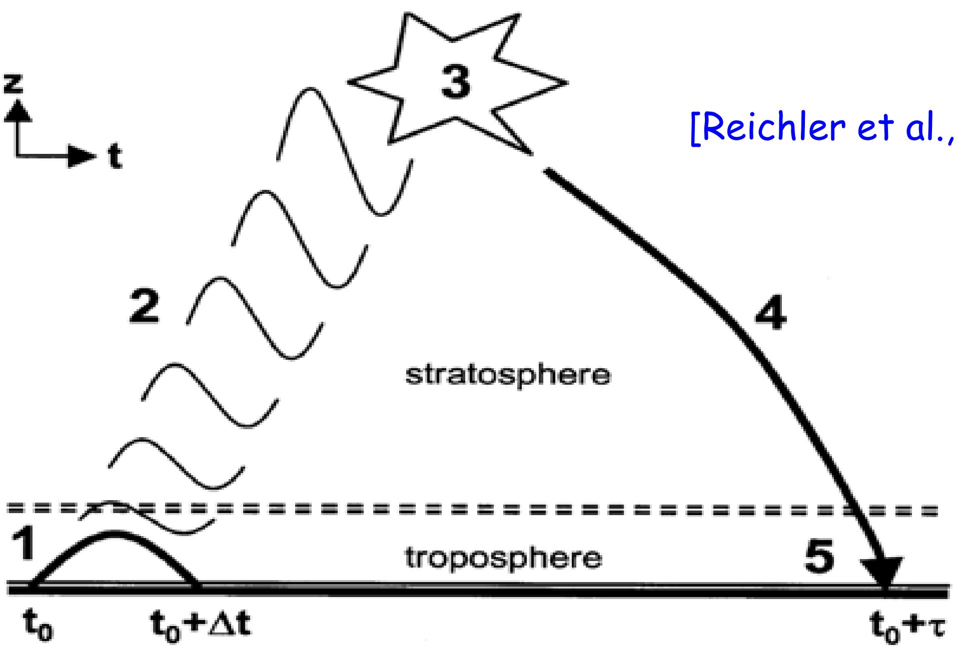


Stratos. Sudden Warmings & its tropospheric impact

- **Wintertime stratosphere** can undergo **large** and **rapid changes** known as **Sudden Warmings**.
- These are **characterized** by **dramatic changes** in **high-latitude wind** and **temperature**.
- In the **troposphere**, they have been **associated** with a **change** in path of **North-Atlantic weather systems**.







[Reichler et al., 2005]

The North Atlantic Oscillation (NAO)

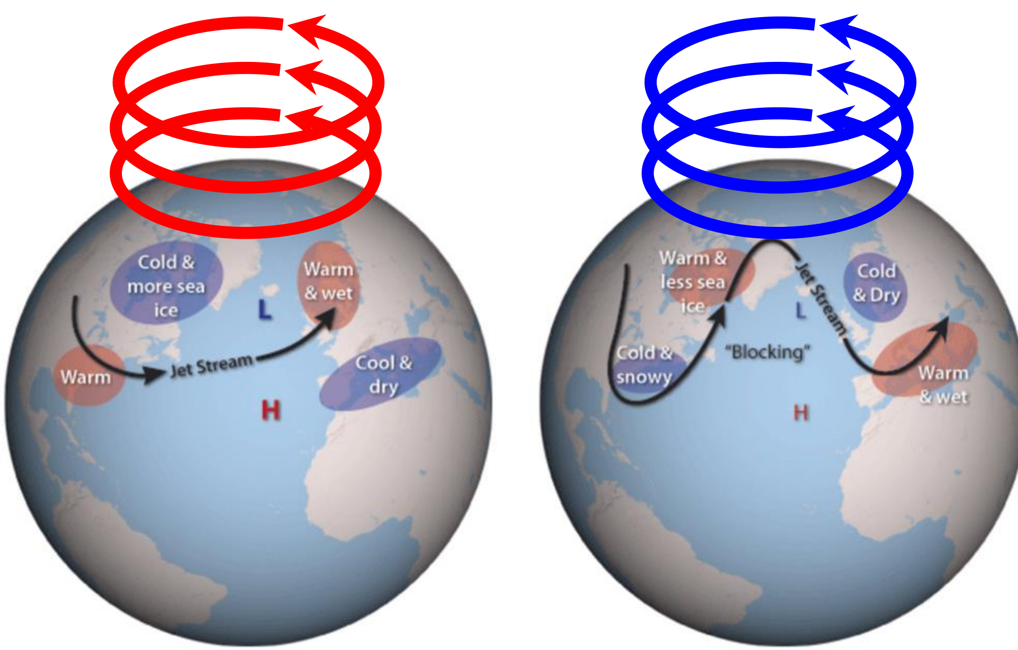
- **Variation** of the **pressure gradient** between **Iceland** and **Azores**.
- **Leading mode of large-scale atmospheric circulation patterns over Europe.**
- **Control** the **strength/direction** of **westerly winds** & **storm track locations**.
- **Phases linked to weather regimes: Cyclonic (NAO+) or Blocking (NAO-).**

Positive NAO phase

Cyclonic regime

Negative NAO phase

Blocking regime

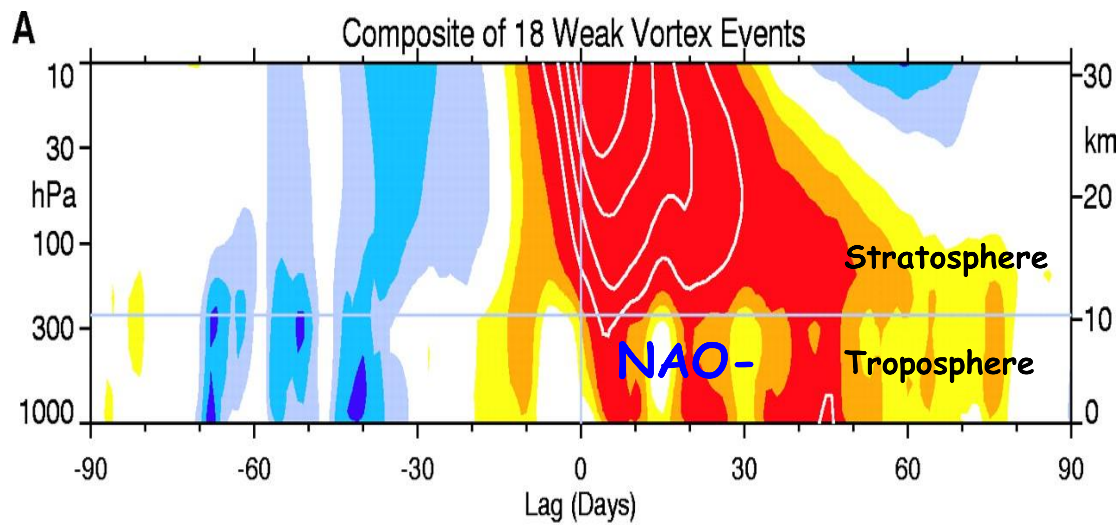


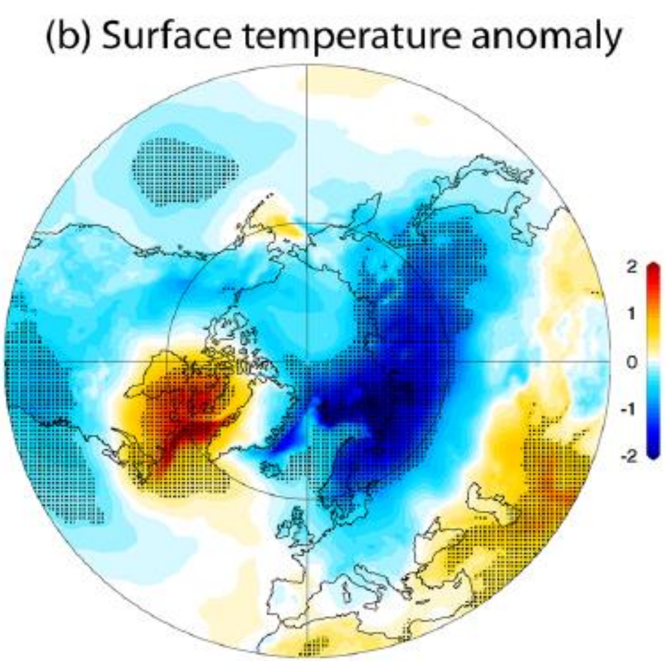
[Cassou, Nat. 2008]
[Grams et al. Nclim 2017]

Weather extremes related to Stratospheric Variability

- **Stratospheric variability** modulates the **European weather**
- **Winter weather extremes** (low temperatures, snow, etc.) are much more common during **weak vortex events**.
- **Atlantic blocking** occurs almost exclusively during **weak vortex events**.
- **Strong winds** and **ocean wave** events are much more common during **strong vortex events**.
- **Composites** of the **60 days sfc T anomalies** (K) after **historical SSWs** in the JRA-55 reanalyses.

[Baldwin & Dunkerton, Sci. 2001]





[Butler et al., 2017]

4. Statistical models

“Essentially, all models are wrong, but some are useful.”

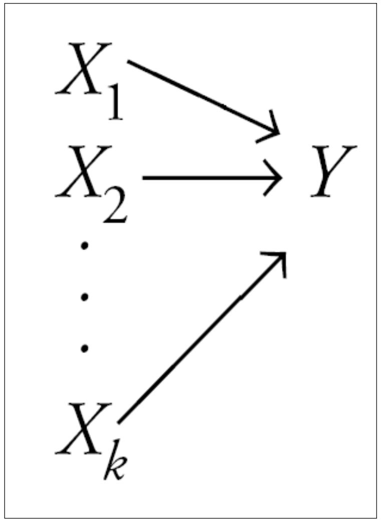
General idea

Simple regression considers the relation between a single explanatory variable and response variable

$$X \to Y$$

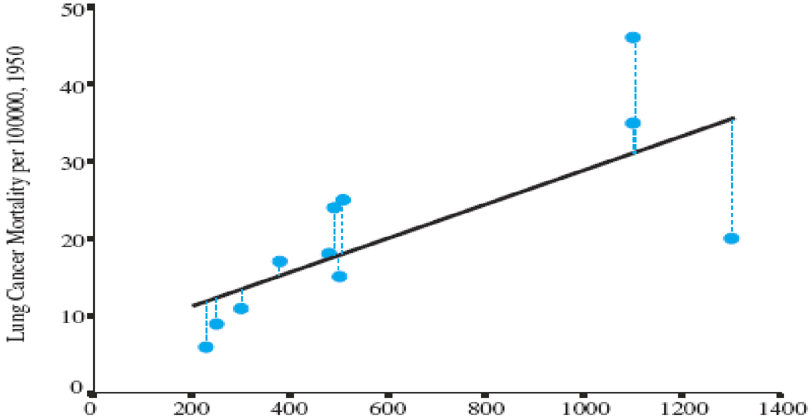
Multiple regression simultaneously considers the influence of multiple explanatory variables on a response variable Y

The intent is to look at the independent effect of each variable **Xi**, while "adjusting out" the influence of potential confounders **Xj (j#i)**



General idea

- A simple regression model (one independent variable) fits a regression *line* in 2-D space



- A multiple regression model with two explanatory variables fits a regression plane in *3*-D space

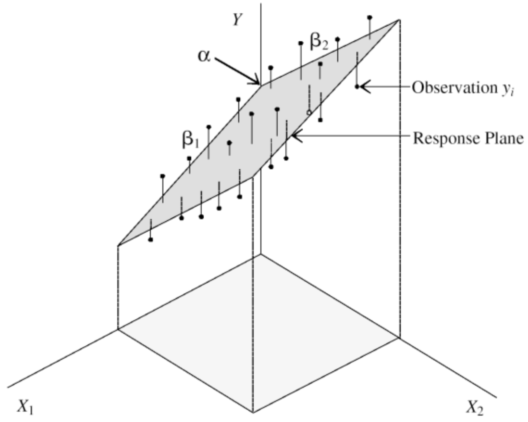
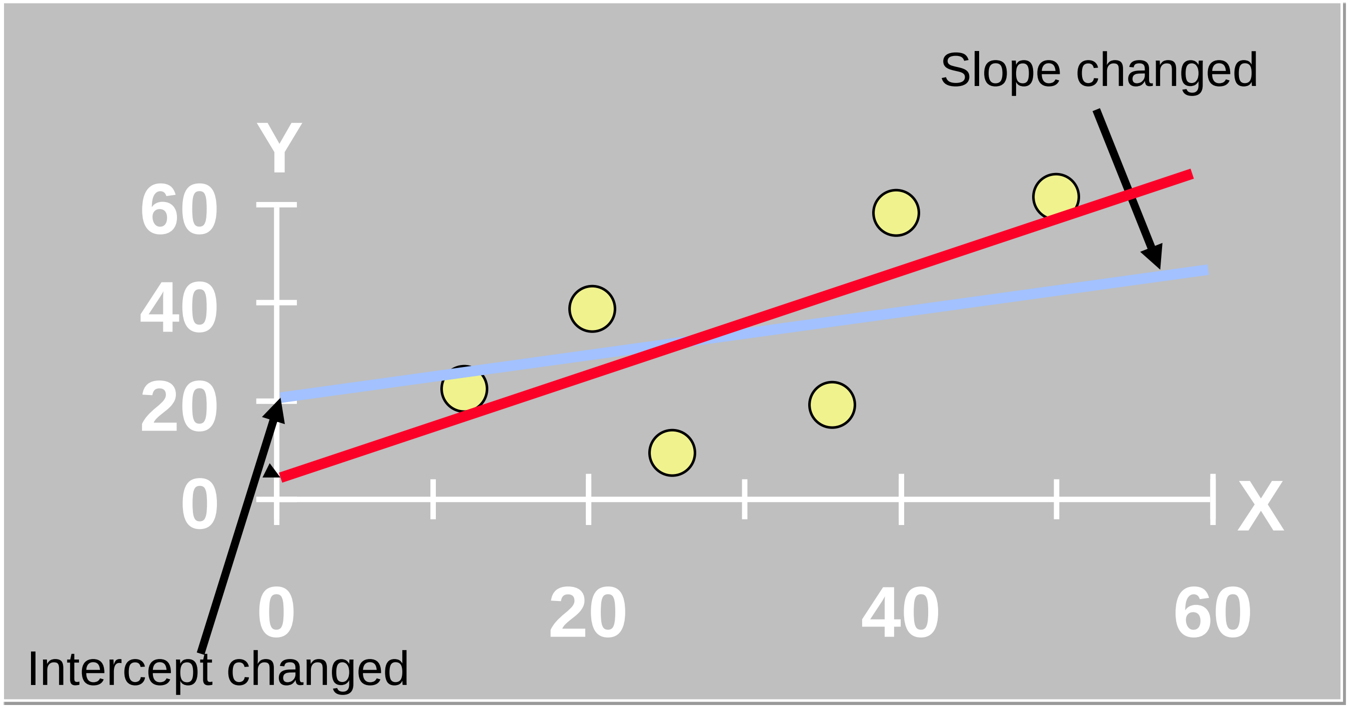


FIGURE 15.1 Three-dimensional response plane.

Thinking Challenge

How would you draw a line through the points?

How do you determine which line 'fits best'?



Simple linear regression model

- For **n** given observations, a linear regression model can be written as

Observation Y-Intercept

Observation Slope

Random Error

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Dependent (Response) Variable

Independent (Explanatory) Variable

Simple linear regression model

- The previous formulas can transform into matrix notation

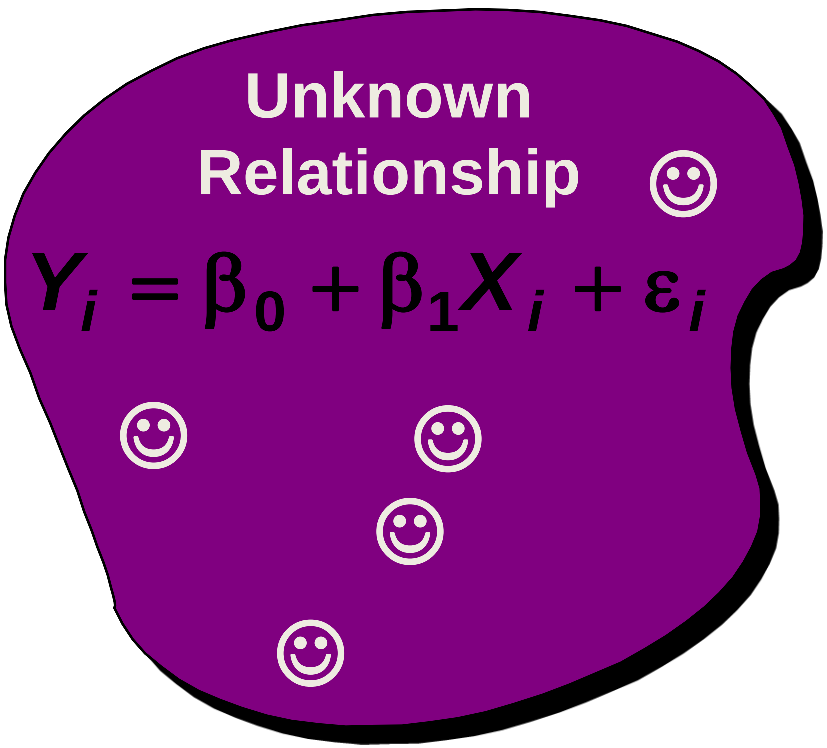
$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ & . \\ & . \\ & . \\ 1 & X_n \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \text{et} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

Observation & sample regression model

- To estimate the parameters, we randomly sample the sample the observations as following:

Observations



$\beta_0 \quad \beta_1 \quad \sigma^2$

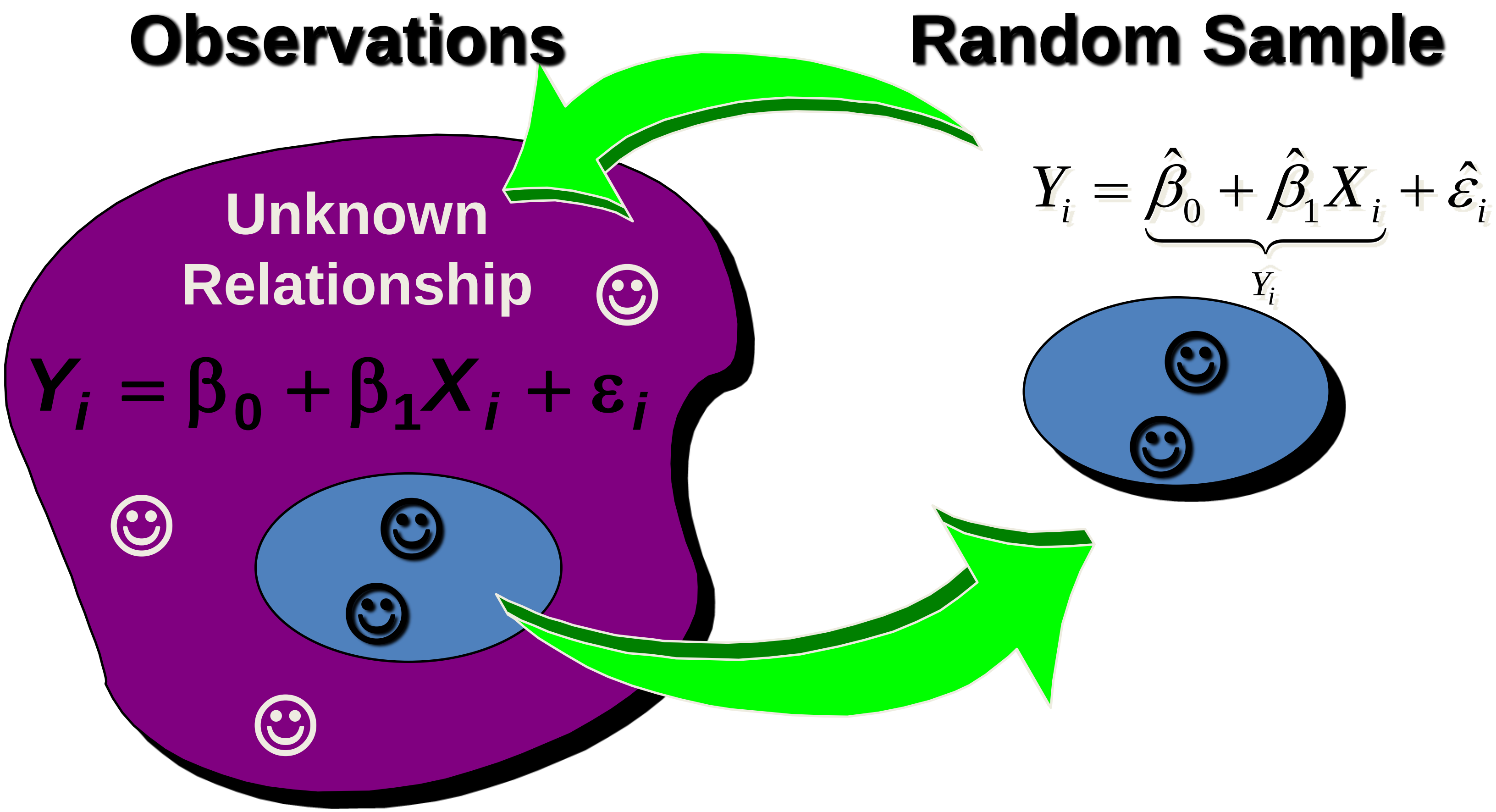
$\hat{\beta}_0 = E(\beta_0)$

$\hat{\beta}_1 = E(\beta_1)$

$\sigma^2 = Var(\varepsilon_i)$

Observation & sample regression model

- To estimate the parameters, we randomly sample the sample the observations as following:



Least squares method

- **Least squares method** (LS) is used to minimize the sum of the squared errors (SSE).

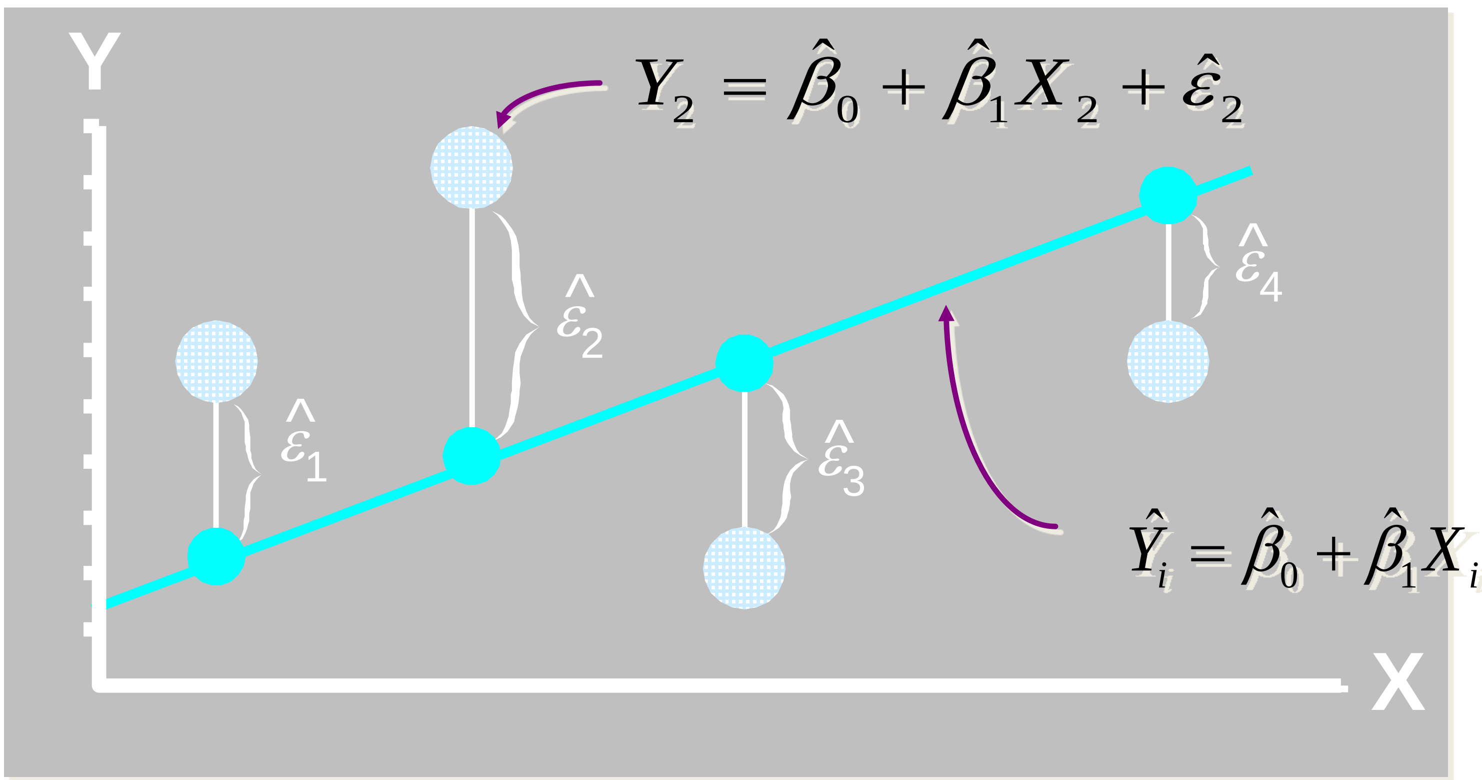
$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

- “**Best Fit**” means difference between actual **Y** values and predicted **Yhat** values are a **minimum**. *But* positive differences off-set negative. So square errors!

Least squares method

- Graphycal representation of LS method.

$$\text{LS minimizes } \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \hat{\varepsilon}_1^2 + \hat{\varepsilon}_2^2 + \hat{\varepsilon}_3^2 + \hat{\varepsilon}_4^2$$



Estimating Parameters

- Parameters are estimated using LS method, which minimizes the SSE.

$$\left(\beta_0, \beta_1\right) = \underset{(\beta_0,\beta_1)\in\mathbb{R}^2}{Arg\min}\left[\sum_{i=1}^{n}\varepsilon_i^2\right]$$

$$= \underset{(\beta_0,\beta_1)\in\mathbb{R}^2}{Arg\min}\sum_{i=1}^{n}\left(Y_i - \beta_0 - \beta_1 X_i\right)^2$$

$$= \min \mathrm{F}\left(\beta_0,\beta_1\right)$$

- Solve the equation system

$$\begin{cases} \left.\dfrac{\partial F(\beta_0, \beta_1)}{\partial \beta_0}\right|_{\beta_0=\beta_0,\beta_1=\beta_1} = 0, \\ \left.\dfrac{\partial F(\beta_0, \beta_1)}{\partial \beta_1}\right|_{\beta_0=\beta_0,\beta_1=\beta_1} = 0, \end{cases}$$

Estimating Parameters

- Prediction equation

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

- Sample slope

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum\left(x_i - \overline{X}\right)^2} = \frac{Cov(X,Y)}{Var(X)}$$

- Sample Y - intercept

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X},$$

Coeficcient of determination

- It is a measure of the **overall quality** of the regression.

- Specifically, it is the percentage of **total variation** exhibited in the Yi data that is accounted for by the sample regression line.

  - The sample mean of Y: $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$
  - Total variation in Y : $= \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2$
  - Residual (unaccounted) variation in Y: $= \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n} \hat{\varepsilon}_i^2$

Coeficcient of determination

$$R^2 = \frac{\text{variation accounted for by x variables}}{\text{total variation}}$$

$$= 1 - \frac{\text{variation not accounted for by x variables}}{\text{total variation}}$$

$$= 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}_i\right)^2}$$

Accuracy of the coeficcient estimates

- The **standard error** of an estimator reflects how it varies under repeated sampling. We have

$$SE(\hat{\beta}_1) = \frac{\sigma}{\sqrt{\underbrace{\sum_{i=1}^{n}(X_i - \overline{X})^2}_{\text{how close together x values are}}}}; \quad SE(\hat{\beta}_0) = \sigma\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}}}$$

$$\sigma^2 = \frac{1}{n-2}\overbrace{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}^{\text{how large the errors are}}$$

- These standard errors can be used to compute **confidence intervals**. A 95% confidence interval is defined as a range of values such that with 95% probability, the range will contain the true unknown value of the parameter. It has the form

$$\hat{\beta}_1 \pm 2.SE(\hat{\beta}_1)$$

Mutiple linear regression model

- The previous formulas can transform into matrix (linear alegebra) notation

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & X_{11} & . & . & . & X_{1p} \\ 1 & X_{22} & . & . & . & X_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & X_{n1} & . & . & . & X_{np} \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_p \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

response vector design matrix slope vector error vector

Mutiple linear regression model

- What mutiple linear regresion means graphycally?

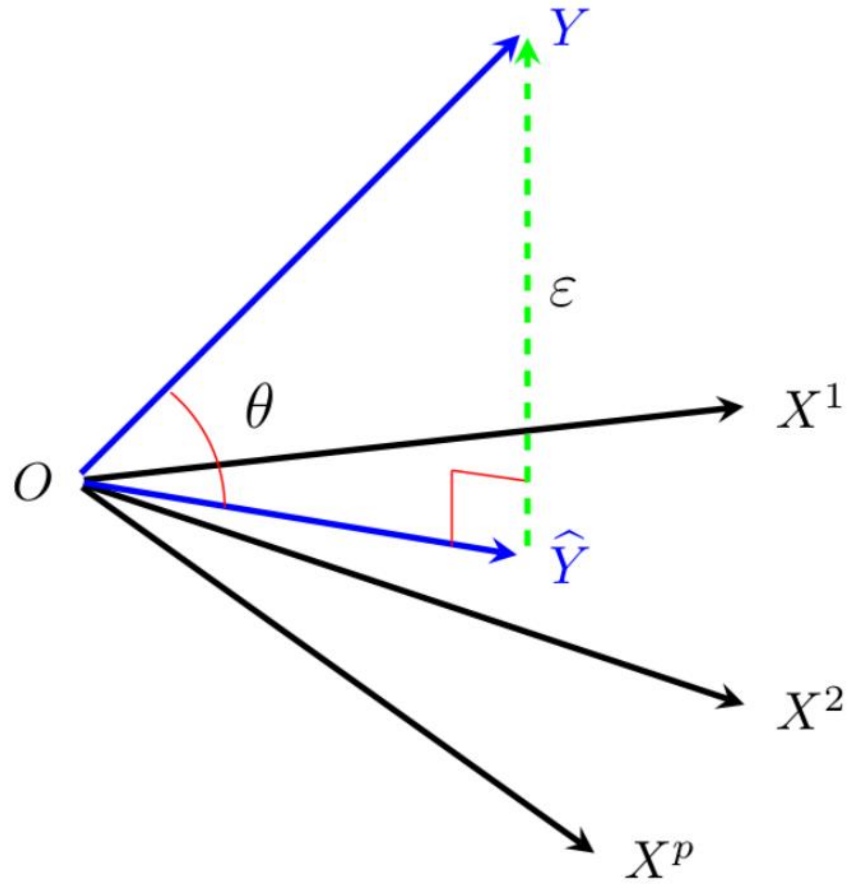


FIGURE 1 – Géométriquement, la régression est la projection $\widehat{Y}$ de Y sur l'espace vectoriel $\text{Vect}\{\mathbf{1}, X^1, \ldots, X^p\}$; de plus $R^2 = \cos^2(\theta)$.

Least squares method

- In order to estimate β, we take a least squares approach that is analogous to what we did in the simple linear regression case. That is, we want to minimize

$$\min_{\beta_{i=1,..,n} \in \mathbb{R}^2} \left[\sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_{i,1} - . \quad . \quad . \quad - \beta_p X_{i,p} \right)^2 \right]$$

over all possible values of the intercept and slopes. It is a fact that this is minimized by setting

$$\beta = \left(X^T X \right)^{-1} X^T Y$$

Estimating Parameters

- The **fitted values** are

$$Y = X\beta = X\left(X^T X\right)^{-1} X^T Y = HY$$

- The **residuals** are

$$\varepsilon = Y - Y = \left(I - X\left(X^T X\right)^{-1} X^T\right)Y = \left(I - H\right)Y$$

- The **error standard deviation** is estimated as

$$\sigma^2 = \frac{1}{n-1-p} \quad \overbrace{\sum_{i=1}^{n} \varepsilon_i^{2}}^{\text{how large the errors are}}$$

Application to mean age of air

- To disentangle the impact of the natural variability (areasols, QBO, ENSO) on BD-circulation, the monthly zonal mean age of air from CLaMS and TRACZILLA simulations, as function of latititude and altitude is analysed using a multiple linear regression model. This model yields:

$$
\begin{aligned}
\mathrm{AoA}(t,\phi,z) = a(\phi,z)\cdot t + C(t,\phi,z) & \quad \text{Trend + Seasonal cycle} \\
+ b_1(\phi,z)\cdot \mathrm{qbo}(t-\tau_{\mathrm{qbo}}) & \quad \text{Quasi-biennial Oscillation} \\
+ b_2(\phi,z)\cdot \mathrm{enso}(t-\tau_{\mathrm{enso}}) & \quad \text{El Nino Southern Oscillation} \\
+ b_3(\phi,z)\cdot \mathrm{aod}(t-\tau_{\mathrm{aod}}) & \quad \text{Volcanic aerosols} \\
+ \varepsilon(t,\phi,z) & \quad \text{Residual}
\end{aligned}
$$

Application to mean age of air

- In this case **p**=5 and **n**=12 months x **N**year.

$$\boldsymbol{Y} = \boldsymbol{\beta}\,\boldsymbol{X} + \boldsymbol{\varepsilon}$$

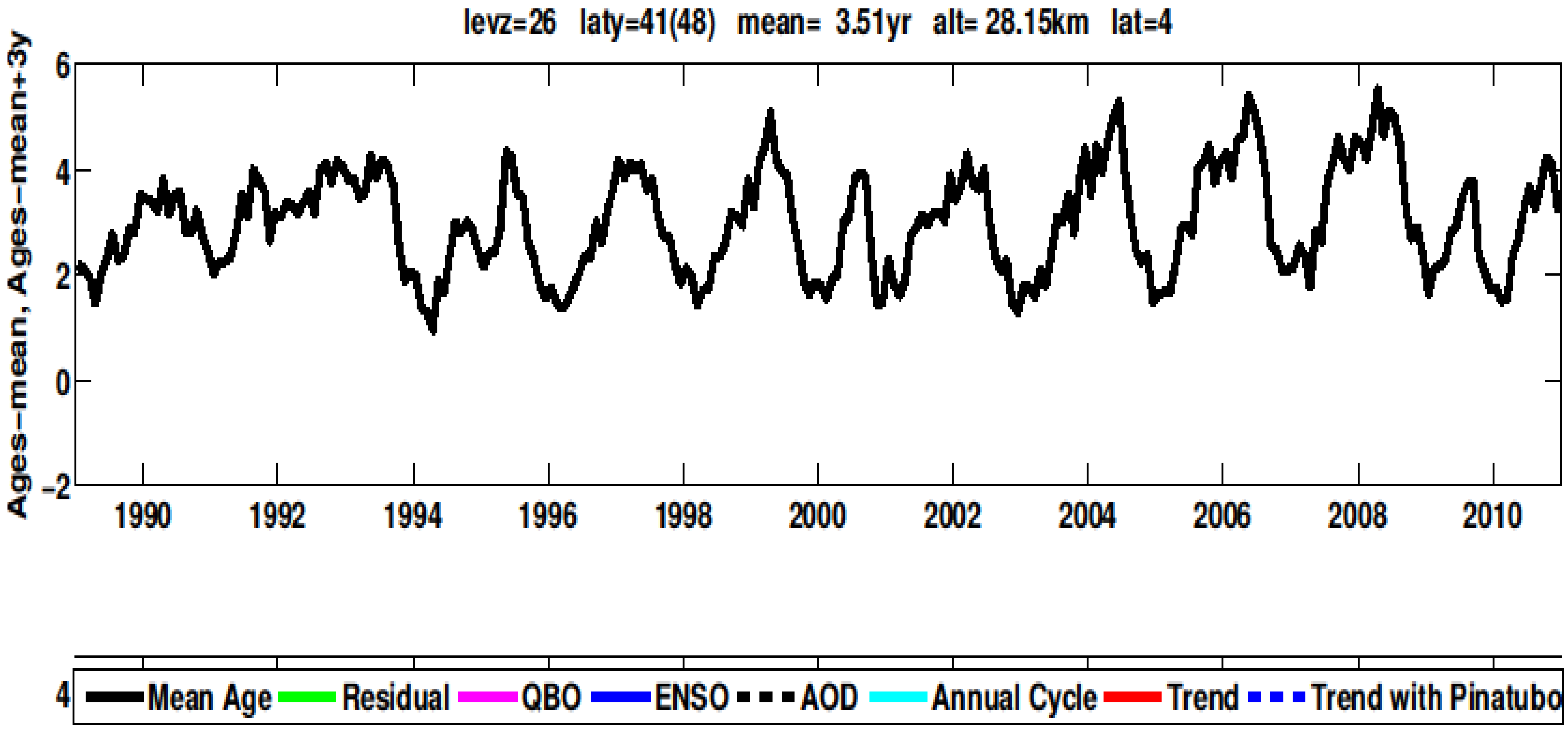
$$\min_{\beta_{i=1,..,n}\in\mathbb{R}^2}\left[\sum_{i=1}^{n}\left(age_i - d_j - a.t_i - b_1.qbo_i - b_2.enso_i - b_3.ado_i\right)^2\right] = \min_{\beta_{i=1,..,n}\in\mathbb{R}^2}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

$$j = 1 + [i-1,12]$$

- Instead of **15** coefficients to estimate we can reduce the system to **4** and use an trick to calculate the **11** coefs.

Application to mean age of air

- Time serie of the mean age of air at 28 km & 4N



Application to mean age of air

- Time serie of the mean age of air and different components of the best fit derived from the regression model.