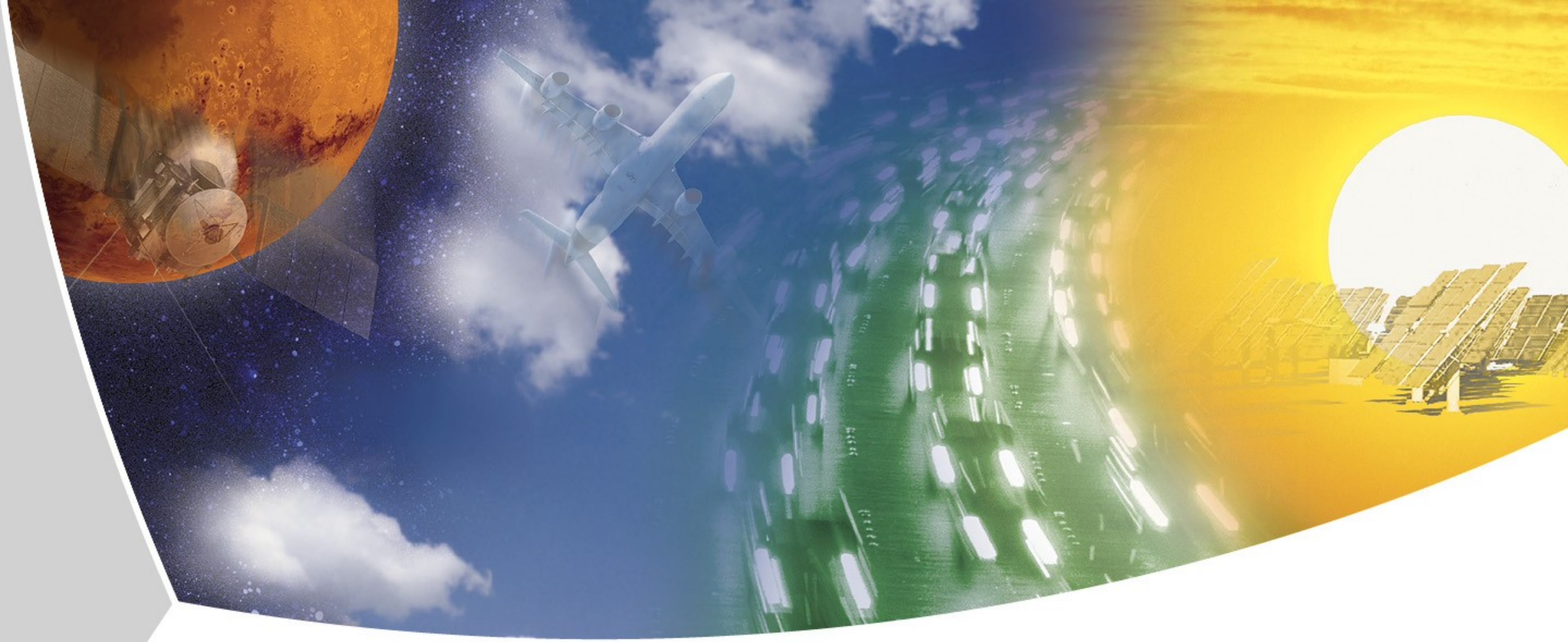


# Nonlinearity in Regression Analyses of Atmospheric Ozone


*Rudolf Deckert*

*Deutsches Zentrum für Luft- und Raumfahrt e.V.*

*Institut für Physik der Atmosphäre*

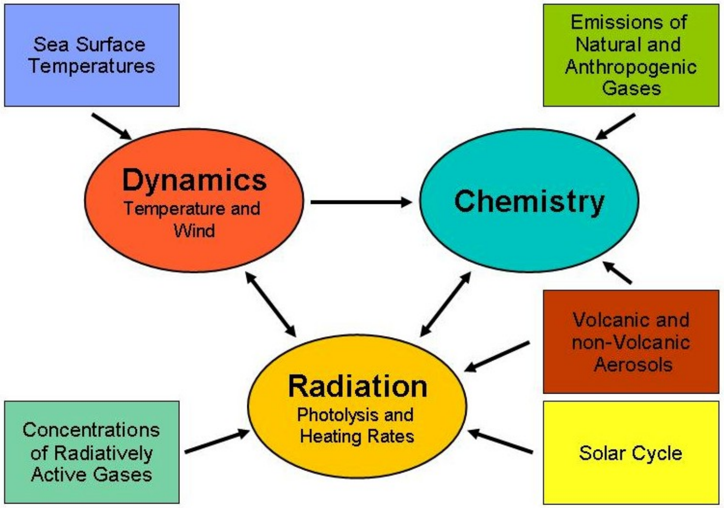

Deutsches Zentrum für Luft- und Raumfahrt e.V.
in der Helmholtz-Gemeinschaft

Potsdam, 30 Sept. 2007

# Introduction

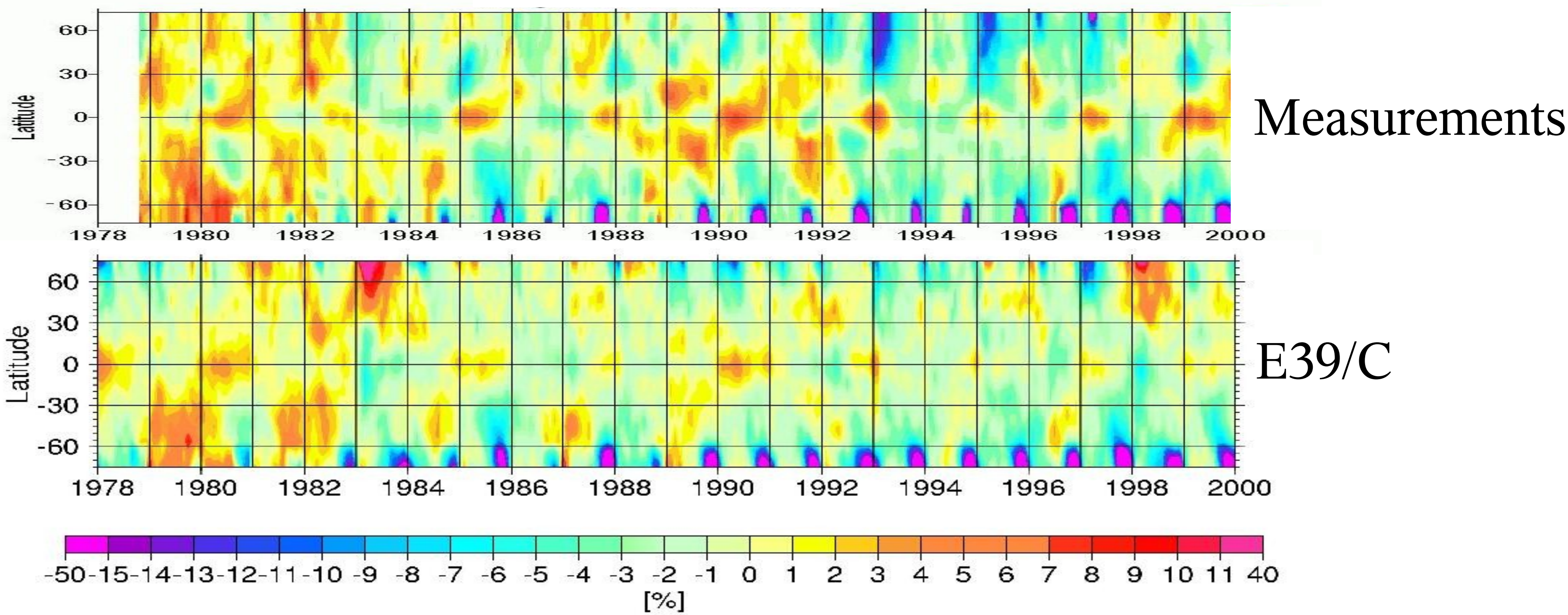
- Chemistry-climate model ECHAM39/CHEM (E39/C)
  - Multi-decadal transient simulations
  - Dynamics and chemistry fully coupled via radiation

- Useful for investigations of ozone interannual variability



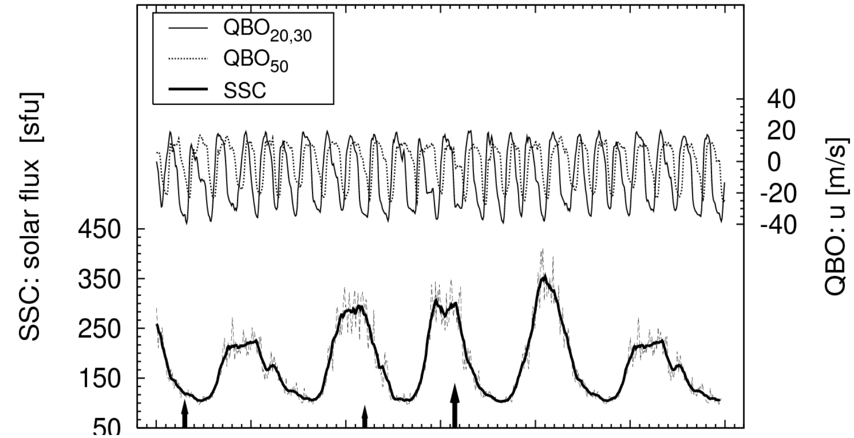
# Introduction

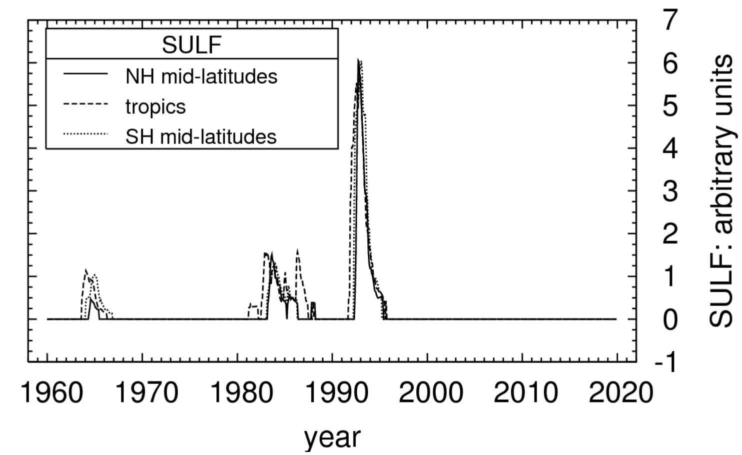
- Similar evolution of total column ozone in measurements and E39/C simulation = deterministic variability
- Analysis of E39/C time series worthwhile

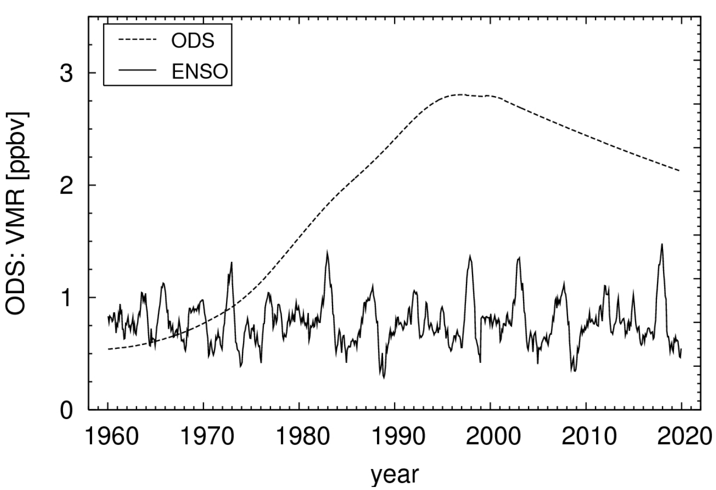


# Introduction

- Deterministic variability associated with E39/C boundary conditions

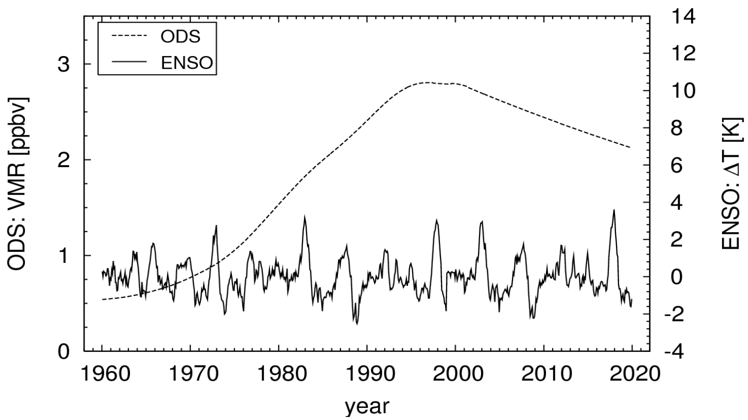


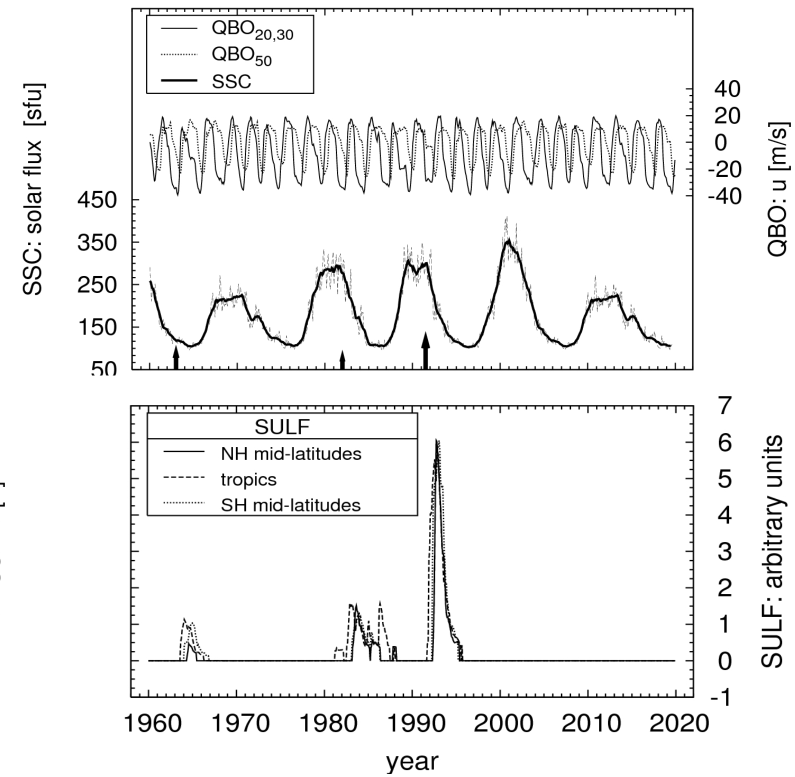



# Introduction

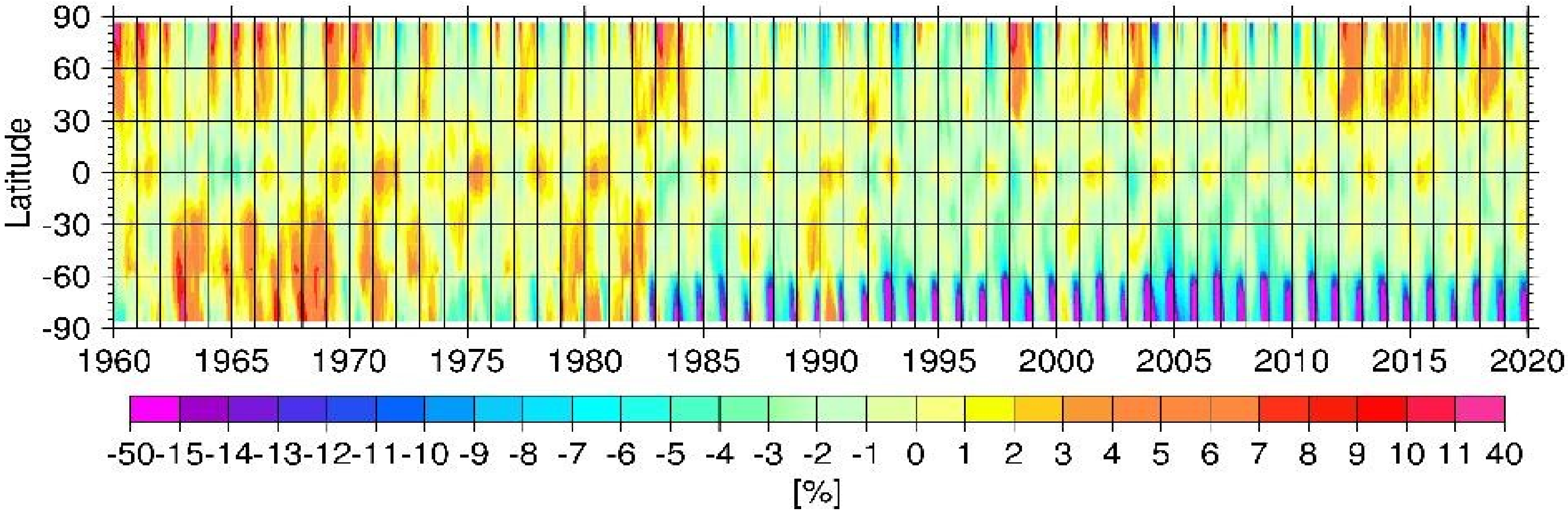
- Multiple regression analysis of E39/C total column ozone
  - Attribution of deterministic interannual variability
  - E39/C boundary conditions as predictors



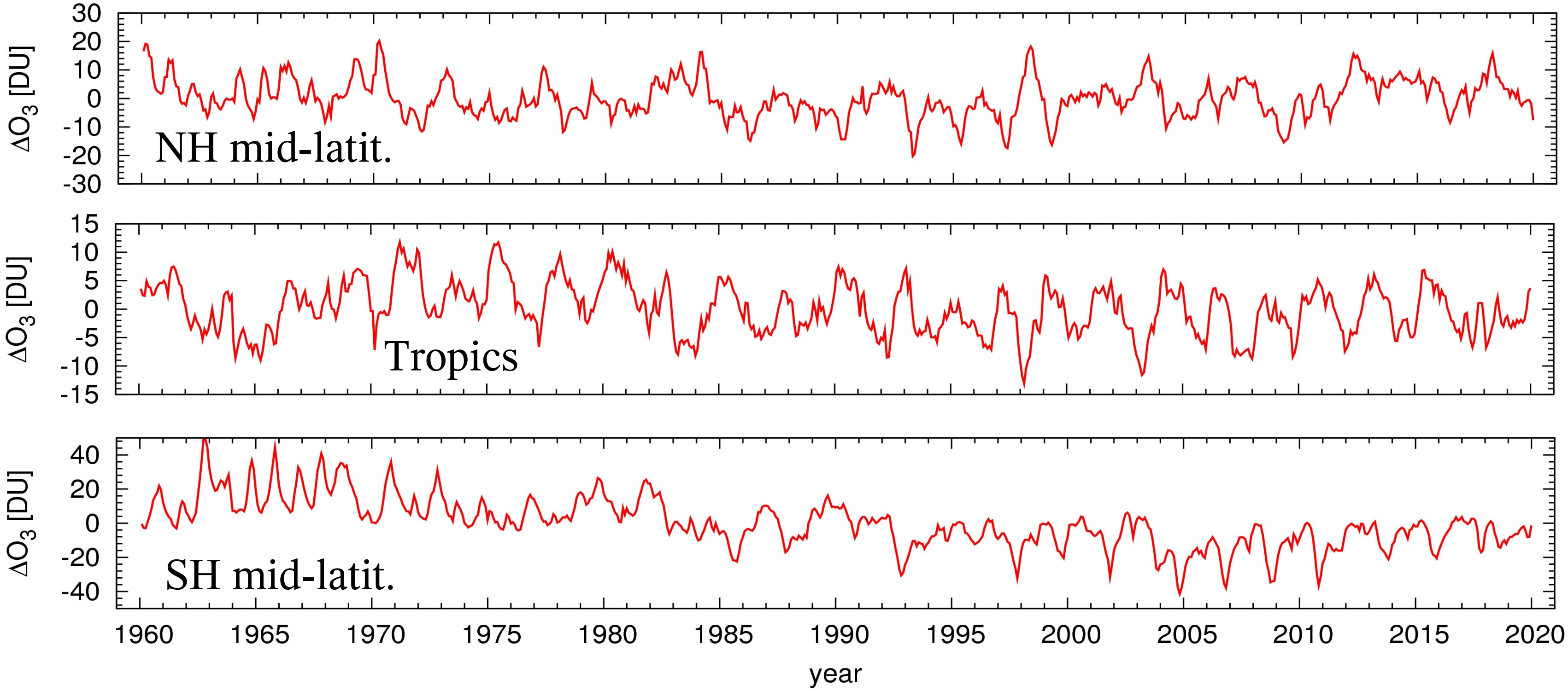
# Data

- E39/C total column ozone (*Dameris et al., 2005, 2006; WMO, 2007*)
  - Ensemble mean of three realisations
  - Monthly zonal means
  - Period 1960 to 2019



# Data

- Multiple-regression analysis of three meridional means
  - NH mid-latitudes (60°N to 35°N)
  - Tropics (15°N to 15°S)
  - SH mid-latitudes (35°S to 60°S)



# Regression model

- Linear combination of deterministic predictors $\beta_k$

$$Y_t = x_1\beta_{1,t} + x_2\beta_{2,t} + \cdots + x_K\beta_{K,t} + a_t$$

- Sinusoidal expansion of parameters $x_k$

$$x_k = x_{k,0} + \underbrace{x_{k,1}\cos\left(\frac{2\pi}{12}t + x_{k,2}\right)}_{\text{annual}} + \underbrace{x_{k,3}\cos\left(\frac{2\pi}{6}t + x_{k,4}\right)}_{\text{semi-annual}}$$
$$+ \underbrace{x_{k,5}\cos\left(\frac{2\pi}{4}t + x_{k,6}\right)}_{\text{four-monthly}} + \underbrace{x_{k,7}\cos\left(\frac{2\pi}{3}t + x_{k,8}\right)}_{\text{quarterly}}$$

- Iterative weighting

# Regression model

- Residuals autocorrelated: autoregressive stochastic process

$$a_t = \epsilon_t \underbrace{+\Phi_1 a_{t-1} + \Phi_2 a_{t-2} + \cdots + \Phi_n a_{t-n}}_{\mathrm{AR}(n)}$$

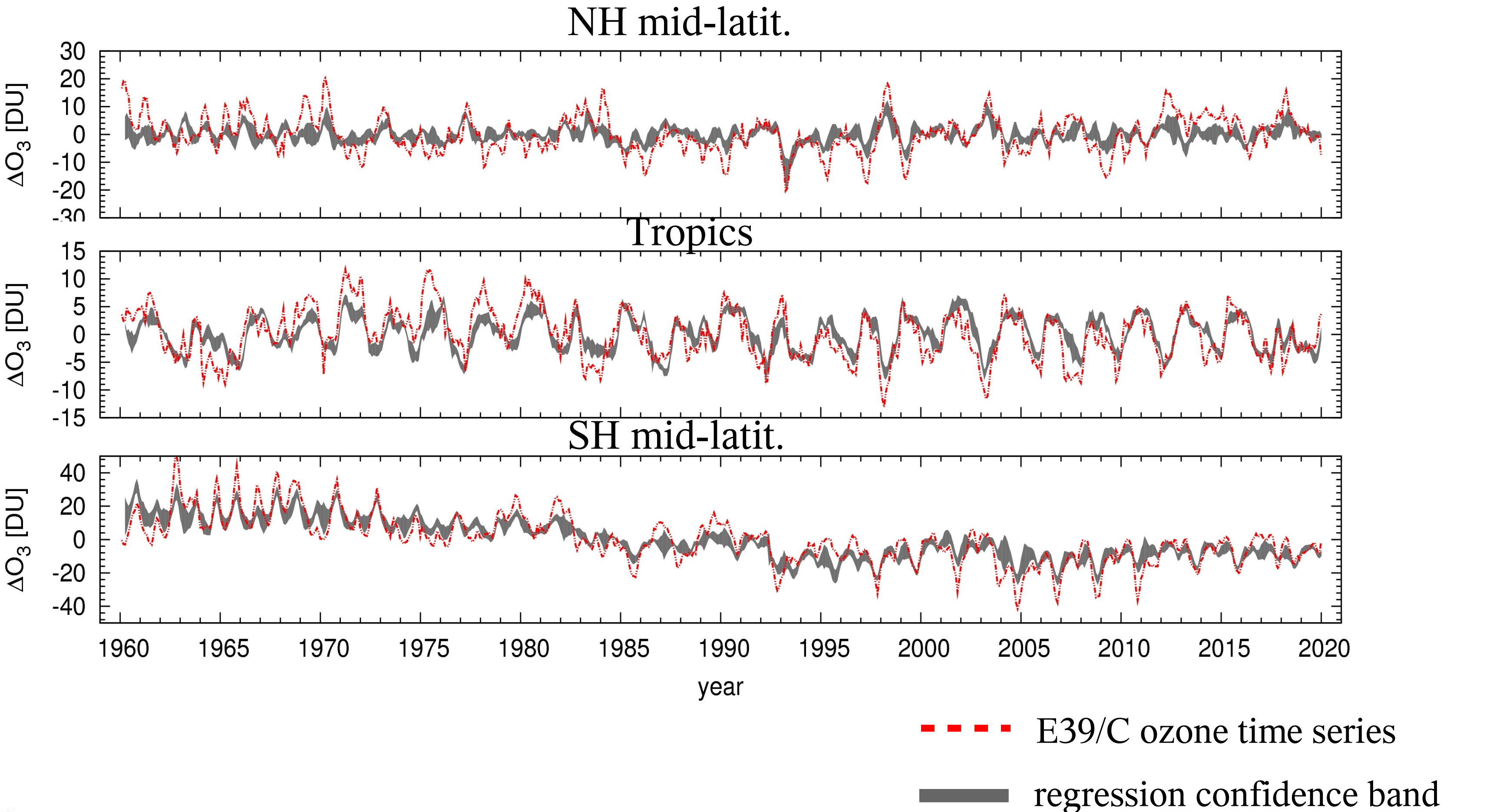
- Combined deterministic/autoregressive model is nonlinear

$$Y_t - \Phi_1 Y_{t-1} = x_1 \beta_{1,t} - \underbrace{\Phi_1 x_1}_{\text{parameter product}} \beta_{1,t-1} + x_2 \beta_{2,t} - \underbrace{\Phi_1 x_2}_{\text{parameter product}} \beta_{2,t-1} + \epsilon_t$$

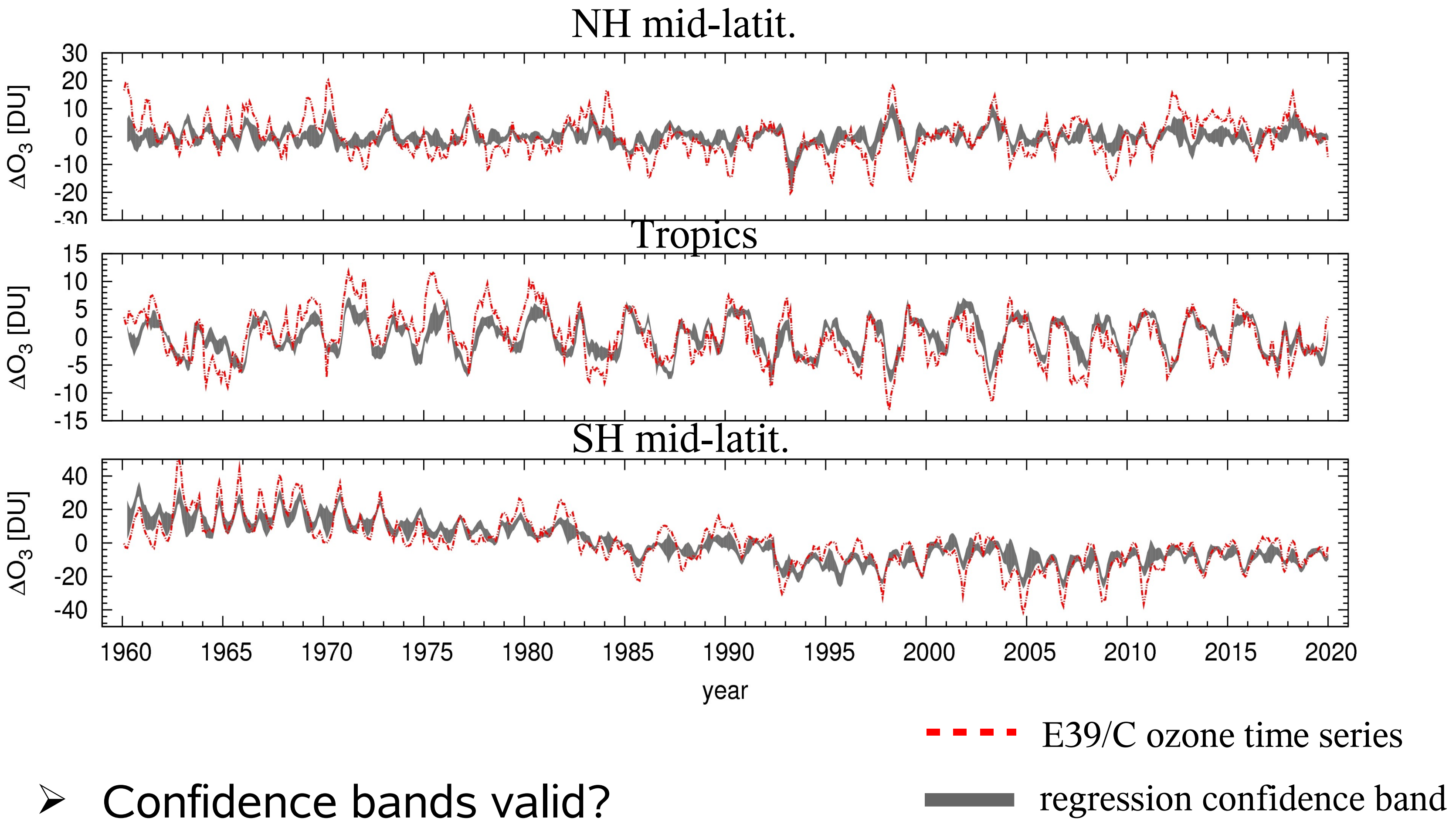
- Use nonlinear minimisation: Levenberg-Marquart algorithm (LMDIF, *Garbow et al., 1980; More et al., 1980*)
- Iterative linear minimisation popular: Cochrane-Orcutt method (*Cochrane and Orcutt, 1949; Thejll and Schmith, 2005*)

# Statistical inference

- Example of statistical inference: 95% confidence band for the combined regression line



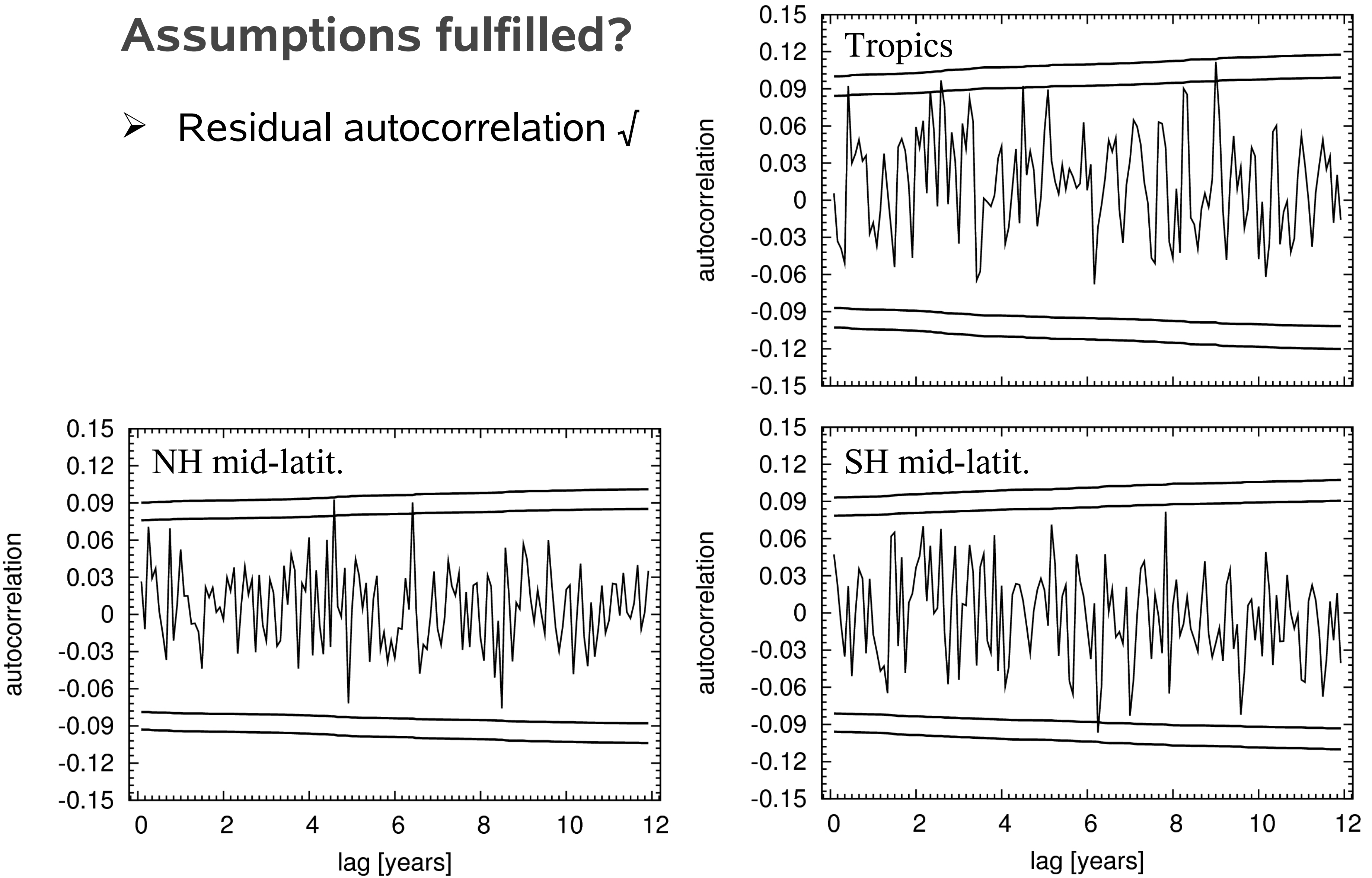
# Statistical inference




- ➢ Confidence bands valid?
- ➢ = Regression assumptions fulfilled?

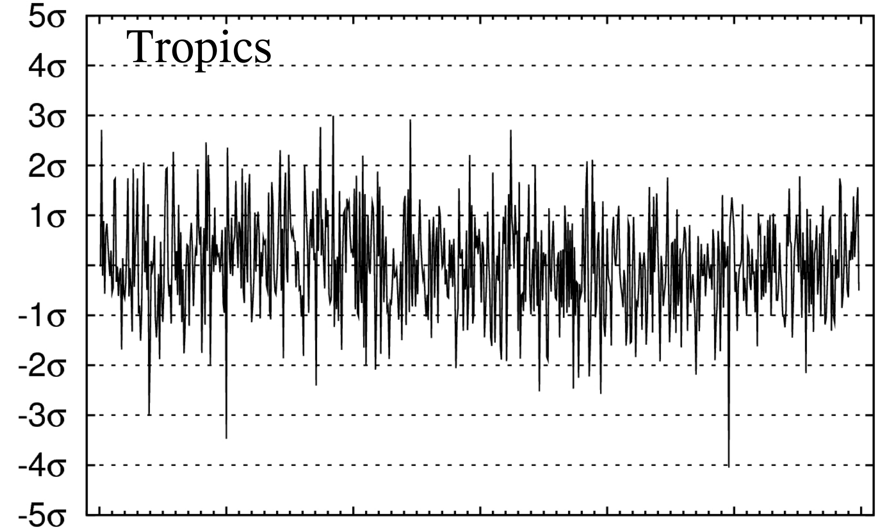
# Assumptions fulfilled?

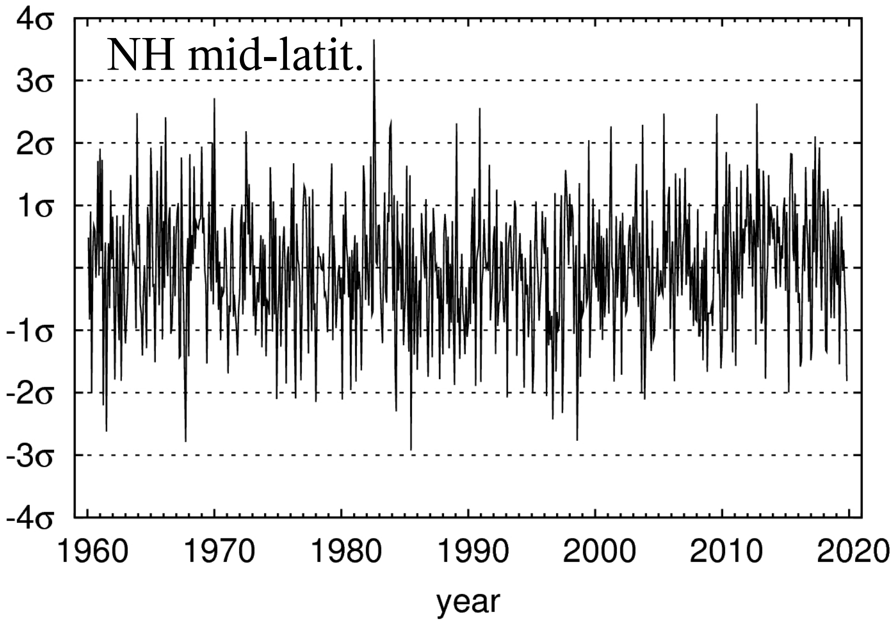
➢ Residual autocorrelation √

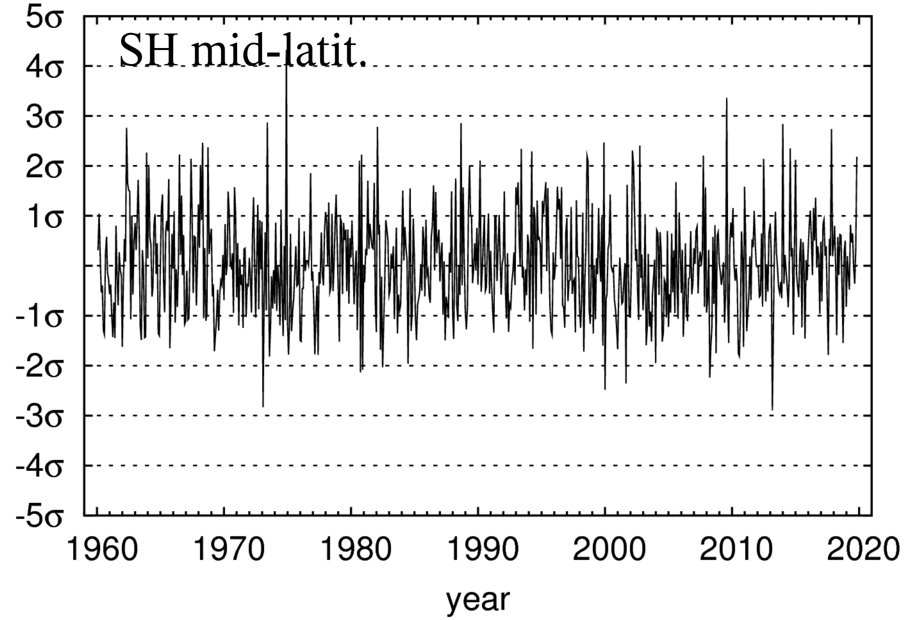


# Assumptions fulfilled?

- Residual distribution √



# Assumptions fulfilled?

- Parsimony √
  - Each parameter significant at least at 95% level
  - Parameter confidence intervals do not include zero
  - Pairwise correlation of deterministic predictors not extreme

- Hence: **standard statistical inference justifiable**

# Assumptions fulfilled?

- Parsimony √
  - Each parameter significant at least at 95% level
  - Parameter confidence intervals do not include zero
  - Pairwise correlation of deterministic predictors not extreme

- Hence: **standard statistical inference justifiable**

But...

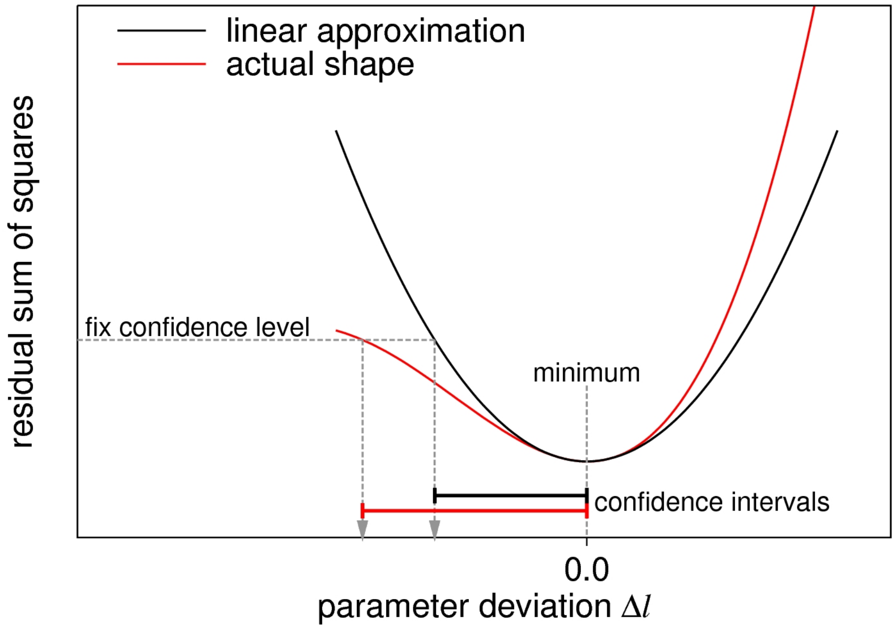
# Distortion problem

- Nonlinearity may distort the error surface
  - = Deviation from pure quadratic shape
  - May affect linear statistical inference
  - Could question Cochrane-Orcutt method

$$Y_t - \Phi_1 Y_{t-1} = x_1 \beta_{1,t} - \underbrace{\Phi_1 x_1}_{\text{parameter product}} \beta_{1,t-1} + x_2 \beta_{2,t} - \underbrace{\Phi_1 x_2}_{\text{parameter product}} \beta_{2,t-1} + \epsilon_t$$

# Distortion problem

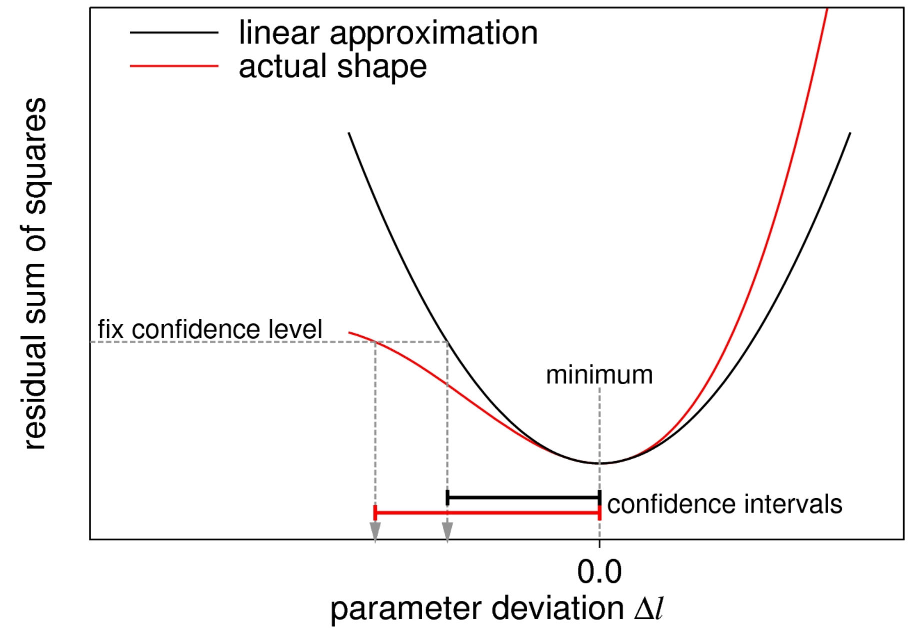
- Nonlinearity may distort the error surface
  - = Deviation from pure quadratic shape
  - May affect linear statistical inference
  - Could question Cochrane-Orcutt method



# Distortion problem

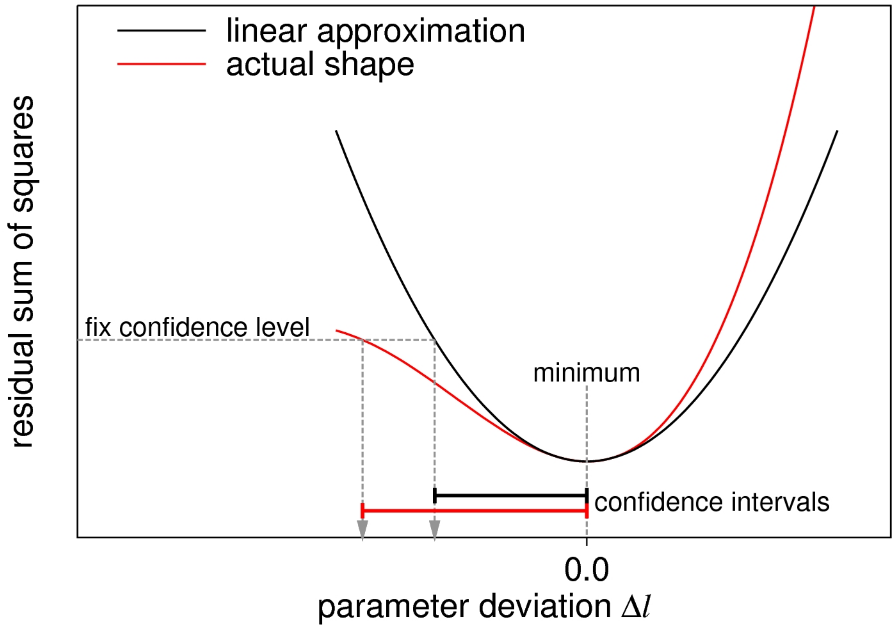
- Econometrics and others: severity of distortion may depend on
  - Strength of autocorrelation (~0.85 for monthly ozone data)
  - Length of time series
  - Type of data

- Probably not been examined for atmospheric-ozone data



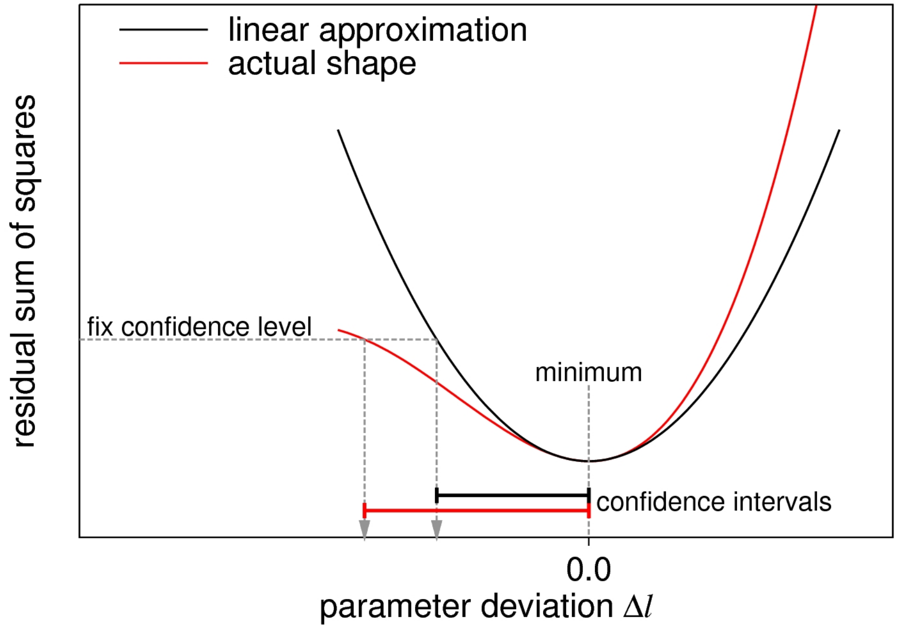
# Examination method

- Approach: check error surface profiles for non-quadratic shape (*Bates and Watts, 1988*)



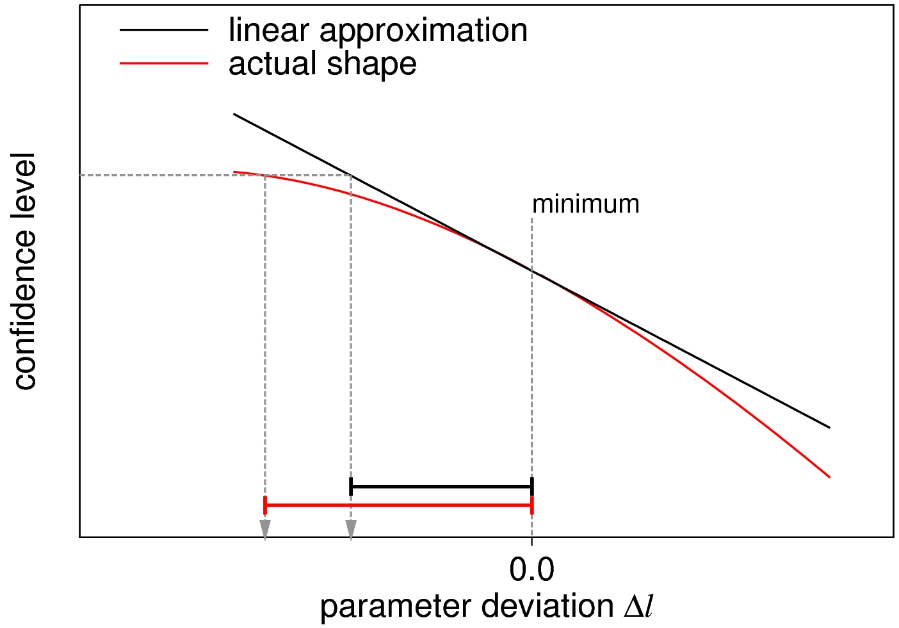
# Examination method

- Calculate profile trace function $t_{x_k}(\Delta l) = RSS_{x_k}(\Delta l) - RSS_{min}$
  - Start at error surface minimum
  - Fix a single parameter away from its optimum value
  - Minimise over all other parameters
  - Repeat for increasing deviations from optimum value
  - Do for all parameters



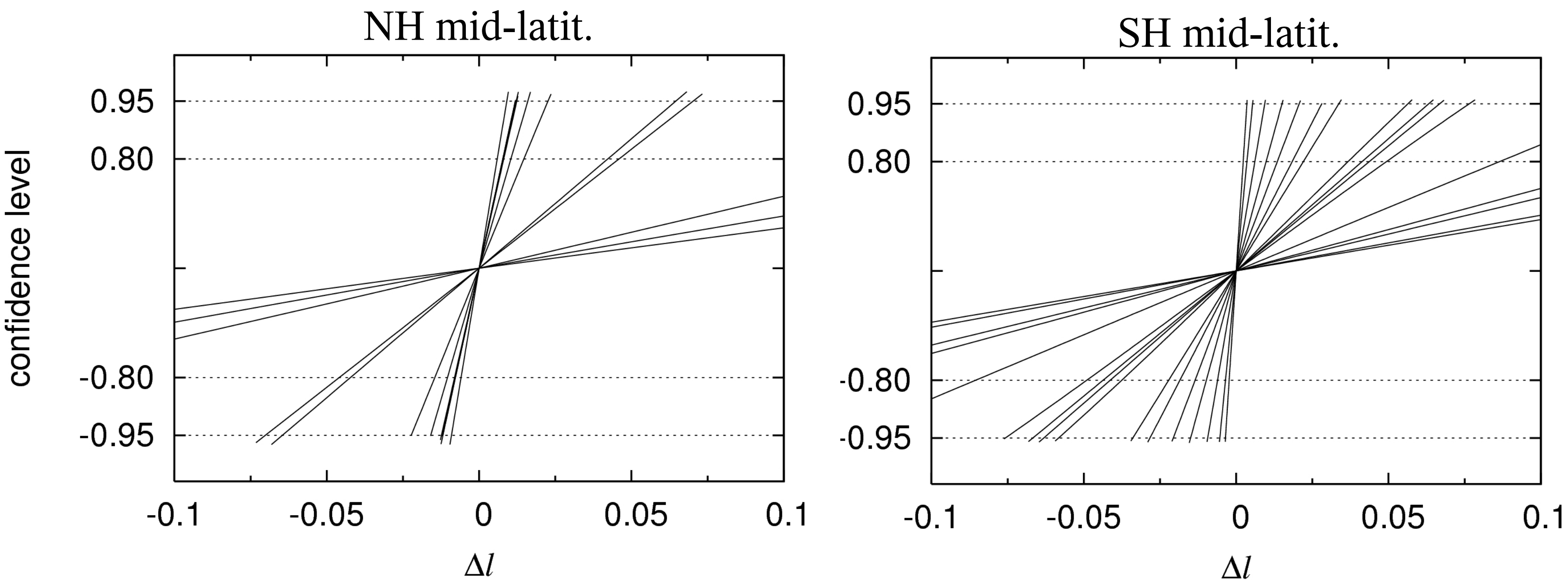
# Examination method

- Going to display $\sqrt{t_{x_k}(\Delta l)}$ with ordinate scaled as parameter confidence level
  - Nonlinear distortion easier to see
  - Parameter confidence intervals straightforward



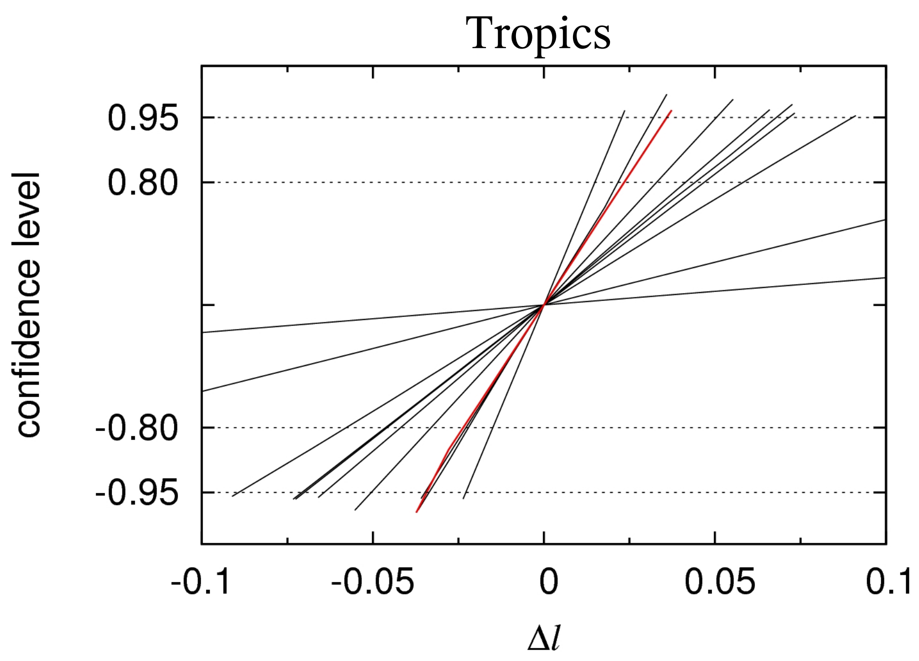
# Results

- NH and SH mid-latitudes / small confidence intervals
  - No obvious distortion



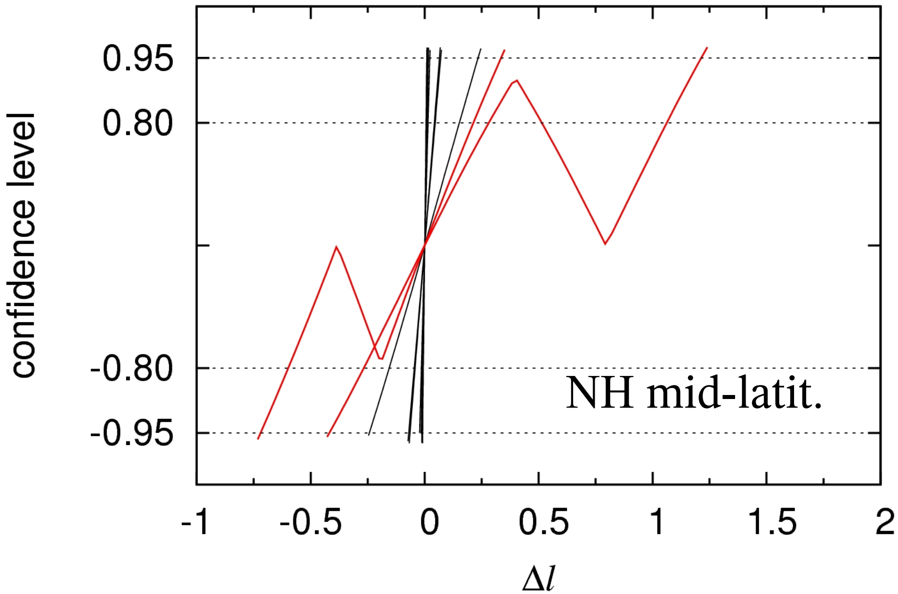
# Results

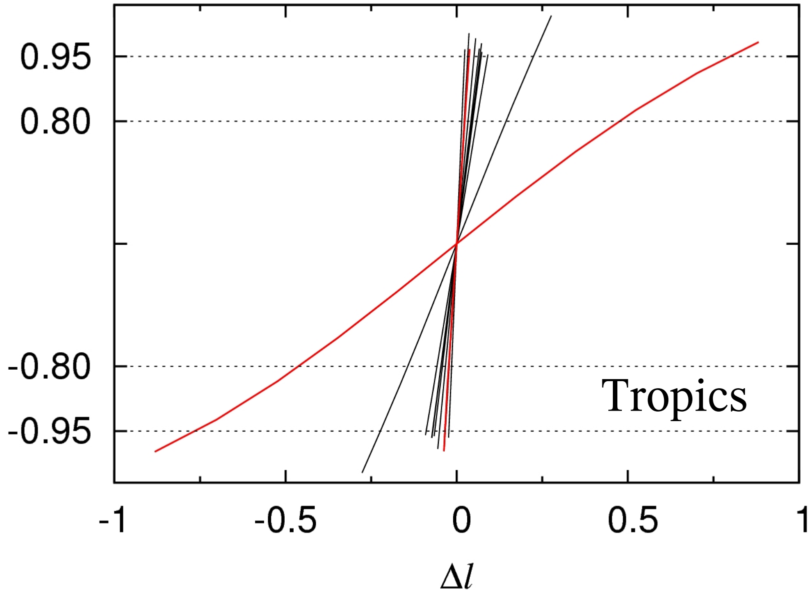
- Tropics / small confidence intervals
  - Small distortion
  - Amplitude of yearly expansion of QBO predictor



# Results

- NH mid-latitudes / large confidence intervals
  - Severe distortion
  - Phase of yearly expansions of ODSs and ENSO predictors
- Tropics / large confidence intervals
  - Small distortion
  - Phase of yearly expansion of QBO predictor



## Results

- Distortion vanishes in case of linear seasonal expansion

$$x_k = x_{k,0} + x_{k,1} \cos\left(\frac{2\pi}{12} t + x_{k,2}\right)$$

$$x_k = x_{k,0} + x_{k,1} \cos\left(\frac{2\pi}{12}\right) + x_{k,2} \sin\left(\frac{2\pi}{12}\right)$$

- **Hence:** distortion **not** due to nonlinearity from combined deterministic/autoregressive model

$$Y_t - \Phi_1 Y_{t-1} = x_1 \beta_{1,t} - \underbrace{\Phi_1 x_1}_{\text{parameter product}} \beta_{1,t-1} + x_2 \beta_{2,t} - \underbrace{\Phi_1 x_2}_{\text{parameter product}} \beta_{2,t-1} + \epsilon_t$$

- Remark: nonlinear seasonal expansion causes weaker pairwise parameter correlation than linear expansion

# Summary and conclusion

- Combined deterministic/autoregressive regression models are nonlinear
  - Iterative linear minimisation popular: Cochrane-Orcutt method
- Nonlinearity may distort the error surface, affecting
  - Validity of linear inference
  - Validity of Cochrane-Orcutt method

- Conduct multiple-regression analysis of E39/C total ozone
  - Nonlinear minimisation
- Profile trace function doesn't show distortion effect from deterministic/autoregressive combination
  - Linear inference justifiable
  - Cochrane-Orcutt method appropriate
  - Not necessarily valid for other time series
- Linear versus nonlinear seasonal expansion