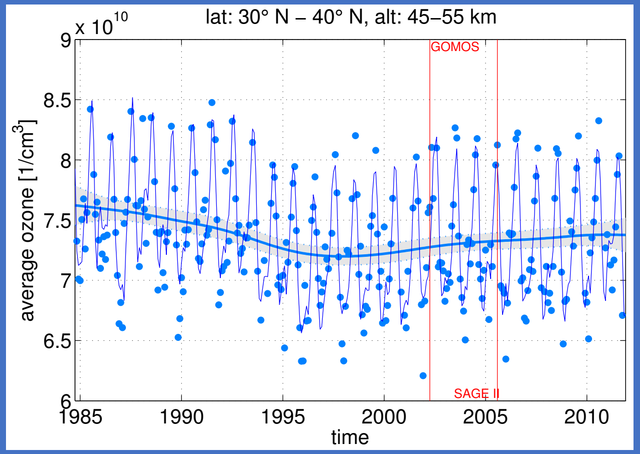
# Ozone trends in the stratosphere determined by Dynamic Linear Model

E. Kyrölä
Finnish Meteorological Institute, Helsinki, Finland

1. Motivation
2. Data sources
3. Regresssion vs. DLM
4. DLM on SAGE-CCI-OMPS 1984-2018
5. Summary

# Motivation



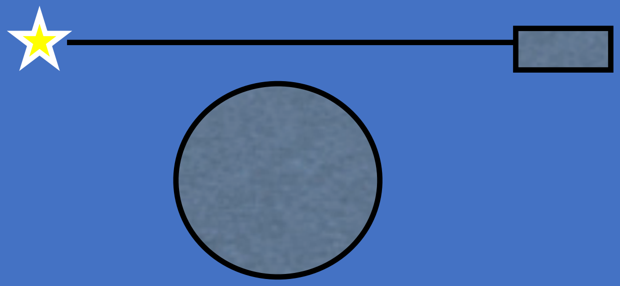

Kyrölä et al., ACP, 2013; GOMOS-SAGE II

How ozone abundance evolves in time

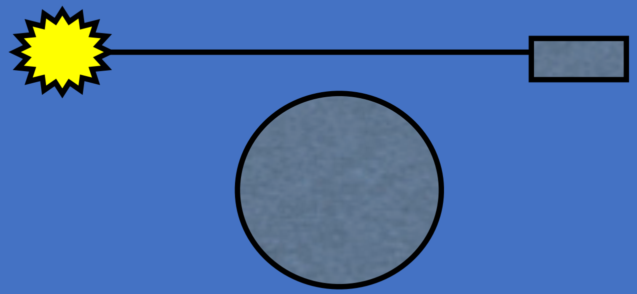
Seasonal cycles + solar, QBO … variations

Proxies, are they”correct”

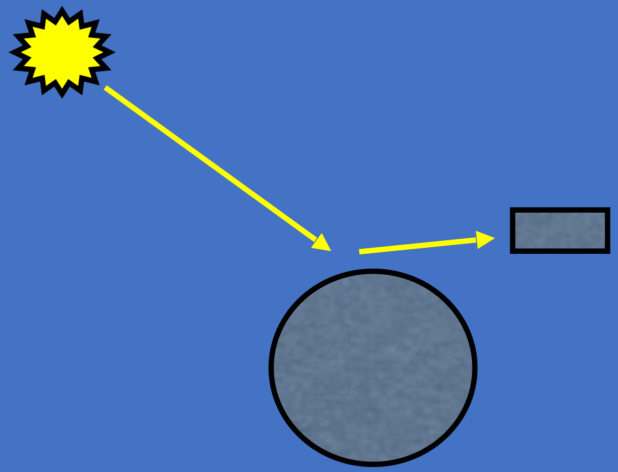
# Ozone profiles from limb viewing instruments



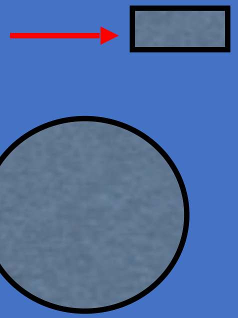
GOMOS, MSX



SAGE, HALOE, POAM,
SCIAMACHY, ACE-MAESTRO/FTS

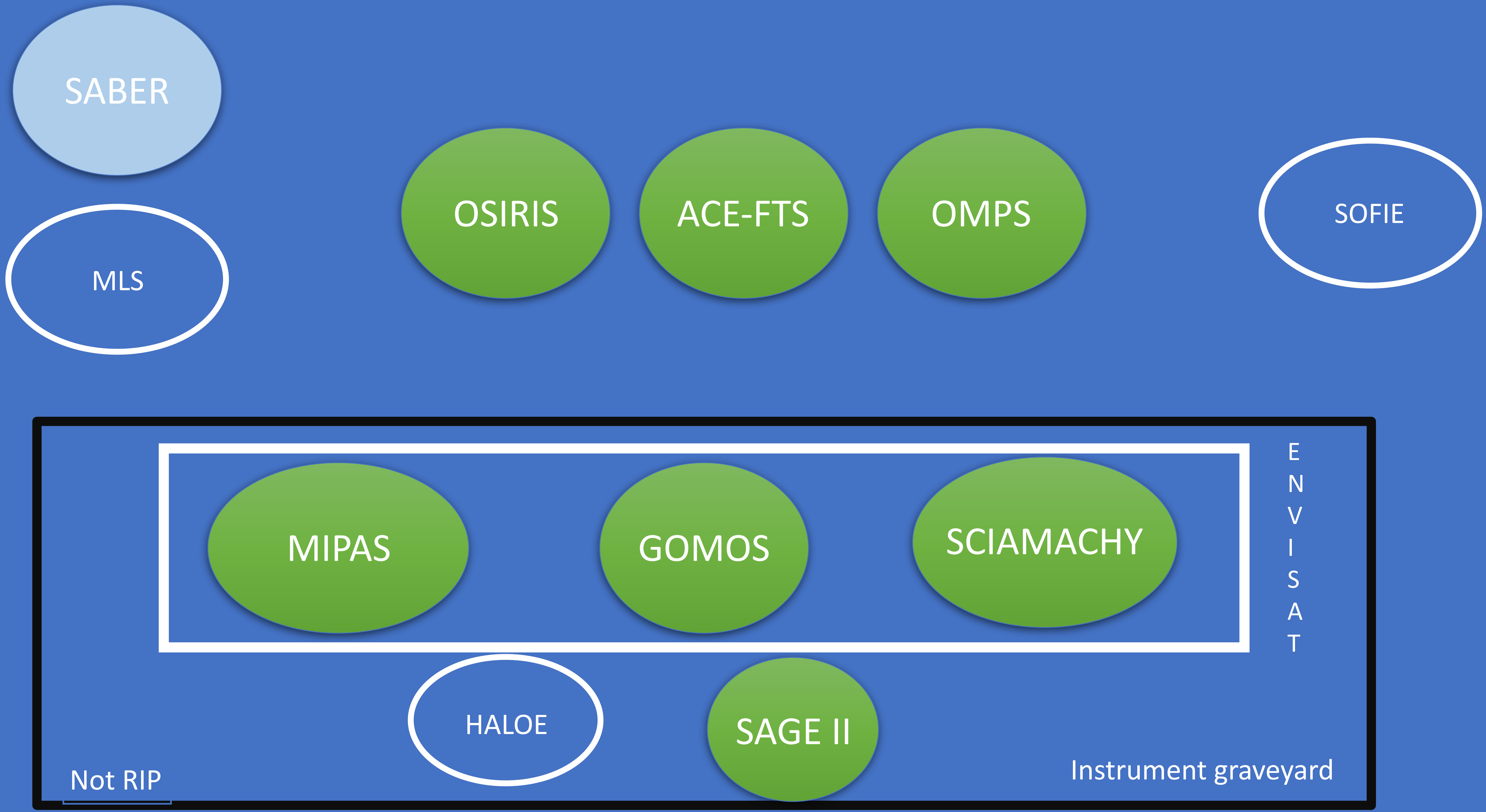


OSIRIS, SCIAMACHY,
OMPS, GOMOS



MIPAS, SMR, MLS, SABER,
SMILES, HIRDLS, LIMS

# Data used in this work



# Time series used in this work

- GOMOS-SAGE II 1984-2011, merged: Kyrölä et al., ACP, 2013
- GOMOS 2002-2011
- SAGE II 1984-2005
- SABER 2001-2018
- ESA-CCI: SAGE-CCI-OMPS 1984-2018, merged: Sofieva et al., ACP, 2017

Viktoria Sofieva

# Time series analysis by linear regression

Observations Estimate Average value Linear trend Annual Solar proxy

$$d_{obs}(t_k) \longleftrightarrow d_{est}(t_k) = c + c_{trend}(t_k - t_{def}) + c_{ann1} cos(wt_k) + \ldots + c_{sol} S(t_k) + \ldots$$

Notice: All c-coefficients are constant except the trend coefficient, which changes at deflection time.

Solution: Find the minimum of $\sum_k (d_{obs}(t_k) - d_{est}(t_k))^2$

Carl Friedrich Gauß (1777–1855)

Adrien-Marie Legendre
1752-1833

# Dynamic Linear Model (DLM)-approach

$$d_{obs}(t_k) \longleftrightarrow d_{est}(t_k) = x_{level}(t_k) + cos(wt_k)x_{ann1}(t_k) + \ldots + S(t_k)x_{sol}(t_k)$$

$$= \sum_j H_j(t_k)x_j(t_k)$$

Observation operator

State

States evolve in time

$$\mathbf{x}(t_k) = \mathbf{M}(t_k)\mathbf{x}(t_{k-1}) + \mathbf{W}$$

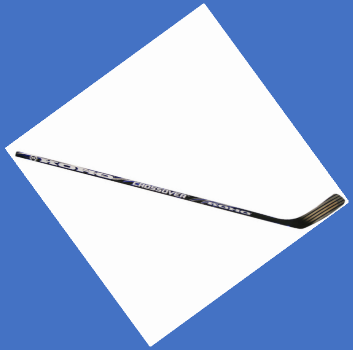
Level state has a hidden state, trend

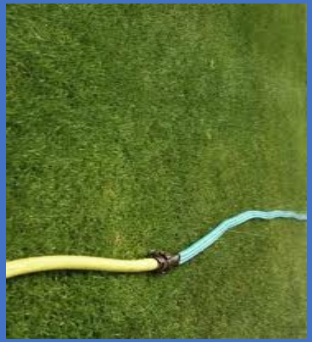
$$x_{level}(t_k) = x_{level}(t_{k-1}) + x_{trend}(t_{k-1})$$

$$x_{trend}(t_k) = x_{trend}(t_{k-1}) + w_{trend}$$

| State | H | M |
|---|---|---|
| Level | 1 | 1 |
| Trend | 0 | 1 |
| Annual | cos(wt) | 1 |
| Solar | S(t) | 1 |

# Linear regression vs Dynamic Linear Model



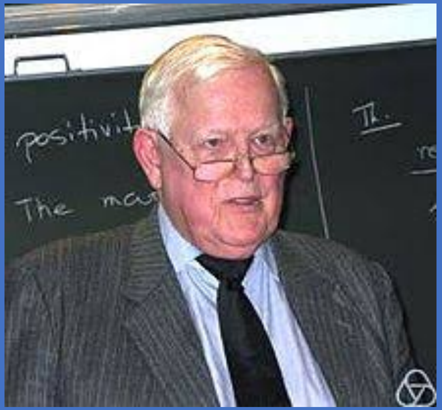


| Linear regression | | DLM |
| --- | --- | --- |
| Trend by a constant term or by several constant terms changed at inflexion points | ⟷ | Trend continuous, smooth function of time |
| Proxies with constant amplitudes | ⟷ | Proxies with time dependent amplitudes (states) |
| Only one common time dependent uncertainty | ⟷ | Every state with individual uncertainty |
| Overdetermined problem | ⟷ | Underdetermined problem |
| Simple scheme | ⟷ | Complicated scheme |

# States can be solved by the Kalman process

$$\mathbf{x}(t_k) = \mathbf{x}^a(t_k) + \mathbf{K}(t_k)(\mathbf{d}_{obs}(t_k) - \mathbf{H}\mathbf{x}^a(t_k))$$

$$\mathbf{x}^a(t_k) = \mathbf{M}(t_k)\mathbf{x}(t_{k-1})$$



**Rudolf Emil Kálmán[**
19.5.1930 – 2.7. 2016)

- Basic problem: more unknowns than observations. Solution: Assuming that at every time step the prior value for a DLM-state is given by the value from the previous step.
- The difference between the predicted observation (from prior states) and the real observed value is used to refresh the state values. **The observability and uncertainty of each state are used to distribute the misfit back to the states**.
- Much more details: Laine et al., ACP, 2014.

Marko Laine

Thomas Bayes
1701 – 7.4. 1761)

# DLM toolbox

The DLM state and parameter estimation (Kalman + Bayes) is implemented in Matlab by M. Laine.

https://github.com/mjlaine/dlm (code) and https://mjlaine.github.io/dlm/ (docs).

State space equations

$$p(\boldsymbol{y}_t|\boldsymbol{x}_t,\boldsymbol{\theta}):\ \boldsymbol{y}_t = \mathbf{F}_t\boldsymbol{x}_t + \boldsymbol{v}_t, \quad \boldsymbol{v}_t \sim N_p(\mathbf{0}, \mathbf{V}_t)$$
$$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\boldsymbol{\theta}):\ \boldsymbol{x}_t = \mathbf{G}_t\boldsymbol{x}_{t-1} + \boldsymbol{w}_t, \quad \boldsymbol{w}_t \sim N_q(\mathbf{0}, \mathbf{W}_t)$$

Model components

| | | |
|---|---|---|
| $t$ | time index | $t = 1, \dots, N$ |
| $\boldsymbol{y}_t$ | observations at time $t$ | $p$ vector |
| $\boldsymbol{x}_t$ | hidden model states at time $t$ | $q$ vector |
| $\mathbf{F}_t$ | observation operator | $p \times q$ matrix |
| $\mathbf{G}_t$ | model operator | $q \times q$ matrix |
| $\mathbf{V}_t$ | observation error covariance matrix | $p \times p$ matrix |
| $\mathbf{W}_t$ | model error covariance matrix | $q \times q$ matrix |
| $\boldsymbol{\theta}$ | static parameter vector | $k$ vector |

Kalman filter recursion

**Step 1**

$$p(\boldsymbol{x}_t|y_t,\boldsymbol{\theta}) = N(\overline{\boldsymbol{x}}_t, \overline{C}_t)$$
$$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1},\boldsymbol{\theta}) = N(\widehat{\boldsymbol{x}}_t, \widehat{C}_t)$$

Kalman filter likelihood

**Step 4**

$$-2\log p(\boldsymbol{y}_{1:N}|\boldsymbol{\theta}) = const + \sum_{t=1}^{N}\left[(\boldsymbol{y}_t - F_t\widehat{\boldsymbol{x}}_t)^T C_{y,t}^{-1}(\boldsymbol{y}_t - F_t\widehat{\boldsymbol{x}}_t) + \log(|C_{y,t}|)\right]$$

Kalman smoother recursion

**Step 2** $\quad p(\boldsymbol{x}_t|\boldsymbol{y}_{1:N},\boldsymbol{\theta}) = N(\widetilde{\boldsymbol{x}}_t, \widetilde{C}_t)$

Simulation smoother

**Step 3** $\quad$ sample: $p(\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N},\boldsymbol{\theta})$

MCMC algorithm

**Step 5** $\quad$ sample: $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:N})$

Monte Carlo sampling

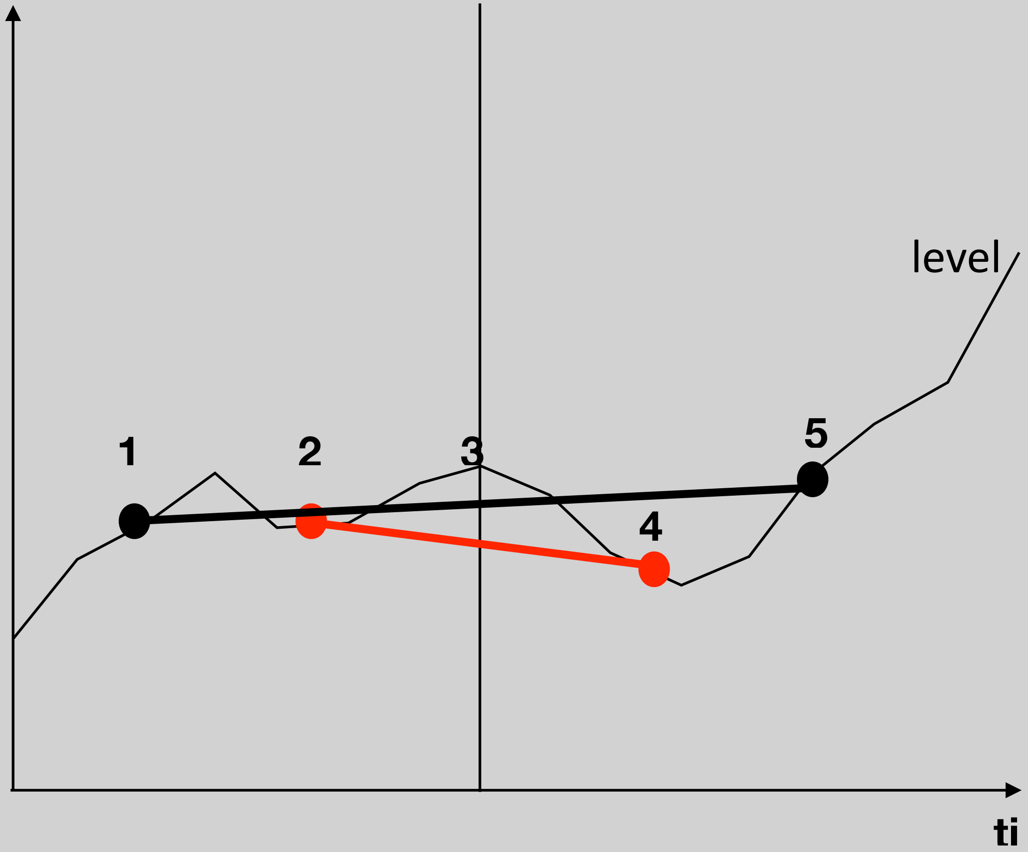
**Step 6**

sample: $p(\boldsymbol{x}_{1:N},\boldsymbol{\theta}|\boldsymbol{y}_{1:N})$
sample: $p(\boldsymbol{x}_{1:N}|\boldsymbol{y}_{1:N})$

Posterior predictive inference of trends

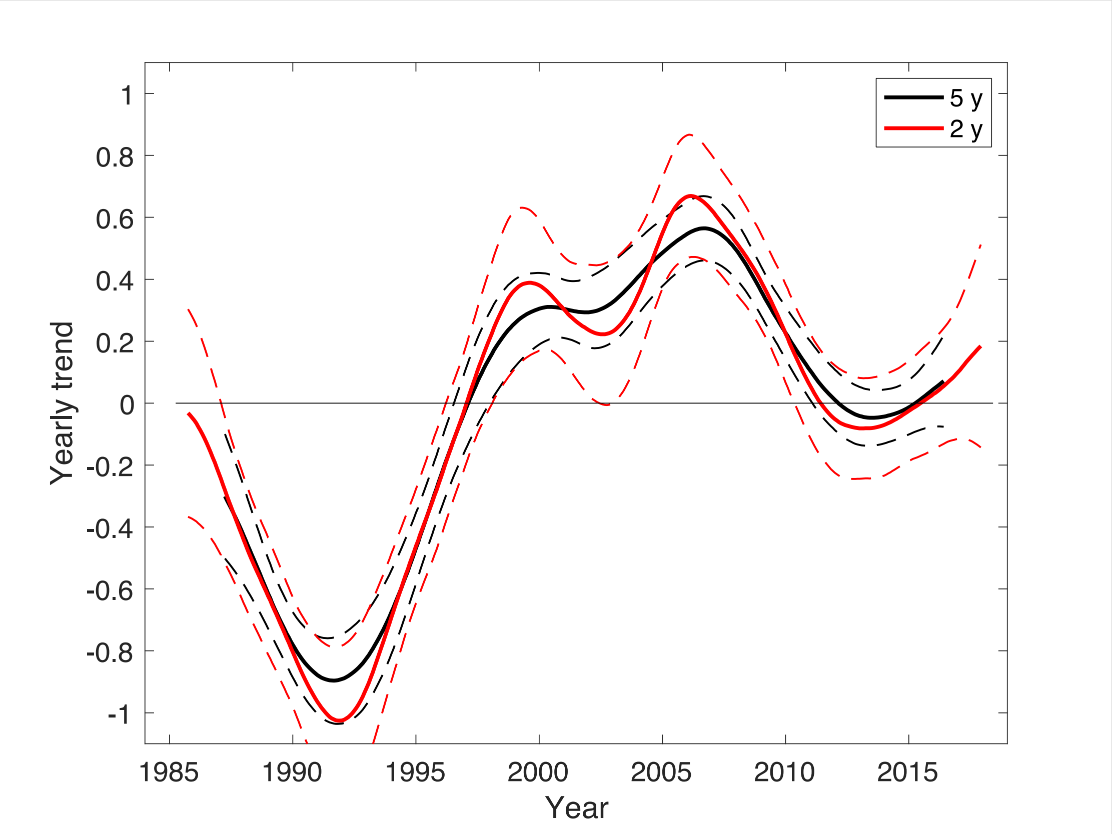
**Step 7** $\quad$ sample: $p(\text{trend}(\boldsymbol{x}_{1:N})|\boldsymbol{y}_{1:N})$

# Trend is calculated from a discrete gradient of the mean level



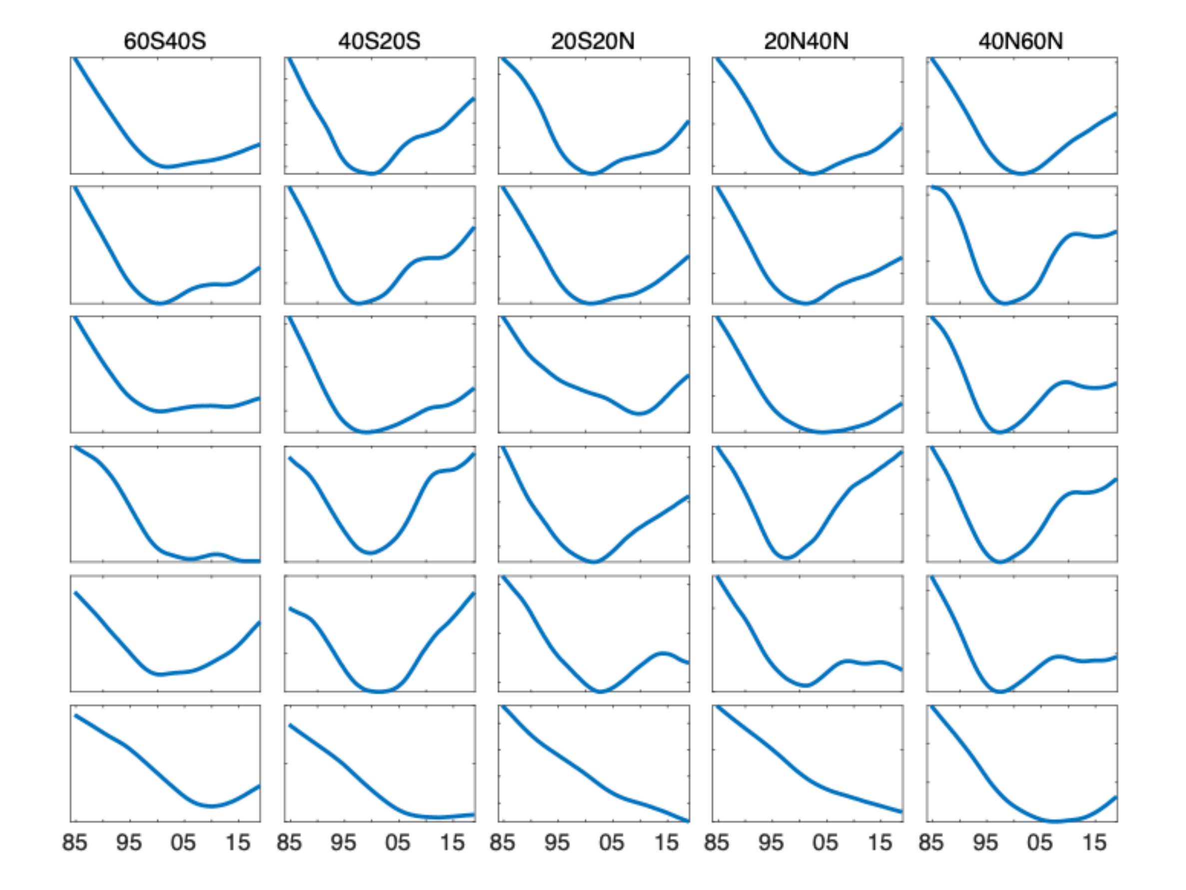

**trend10(t3)=<y(t5)-y(t1)>/10**

**trend5(t3)=<y(t4)-y(t2)>/5**

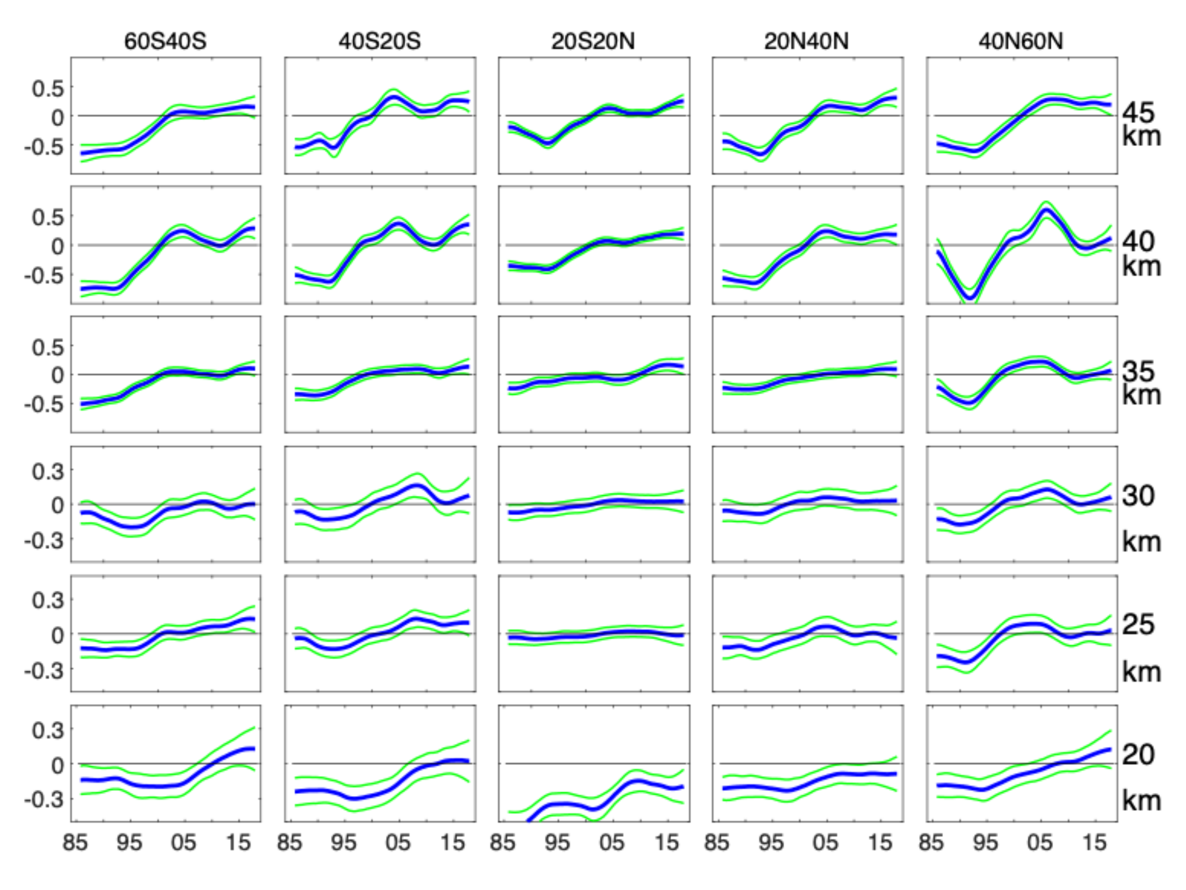


# DLM analysis of SAGE-CCI-OMPS data set 1984-2018

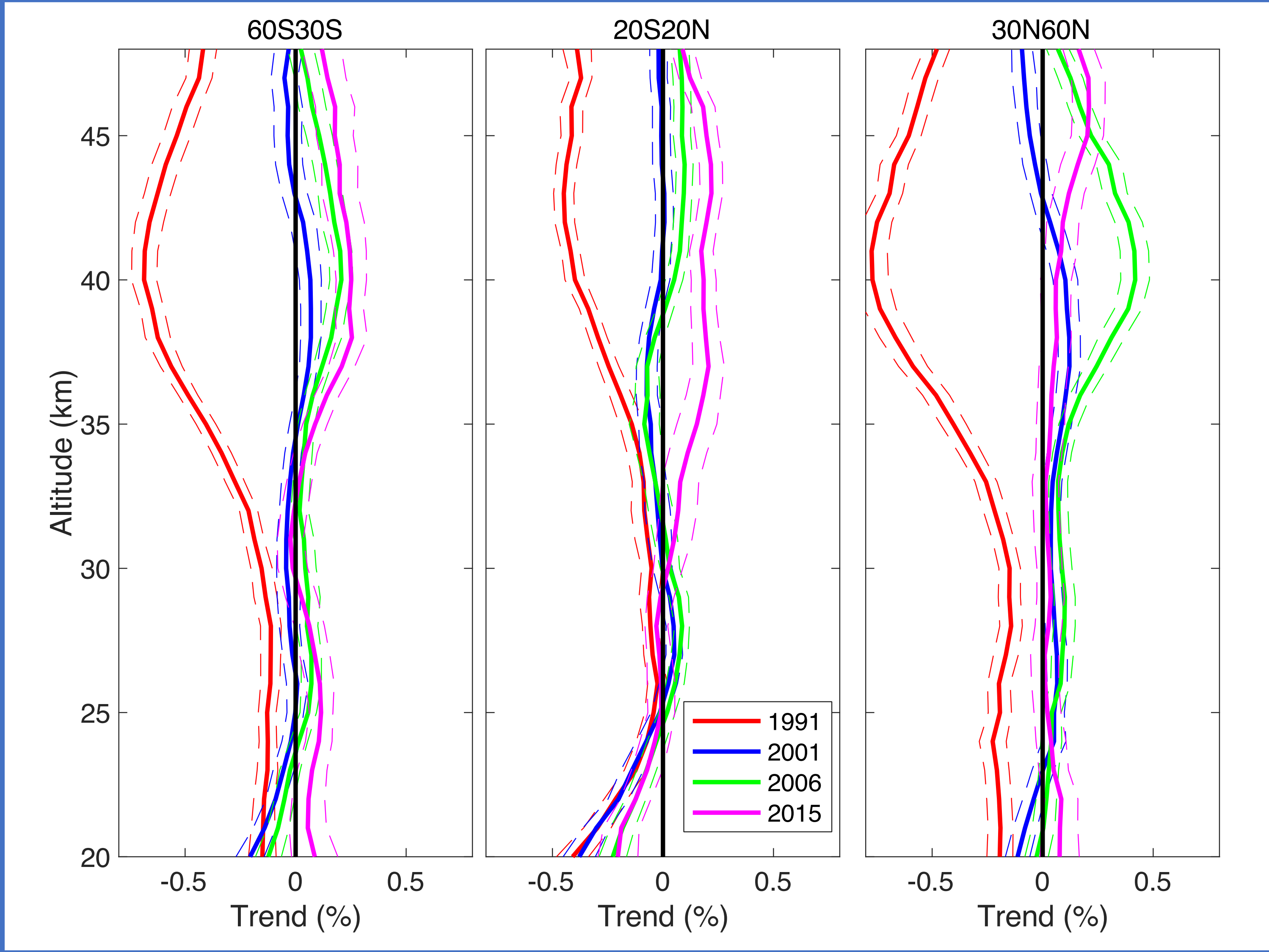
Ozone levels



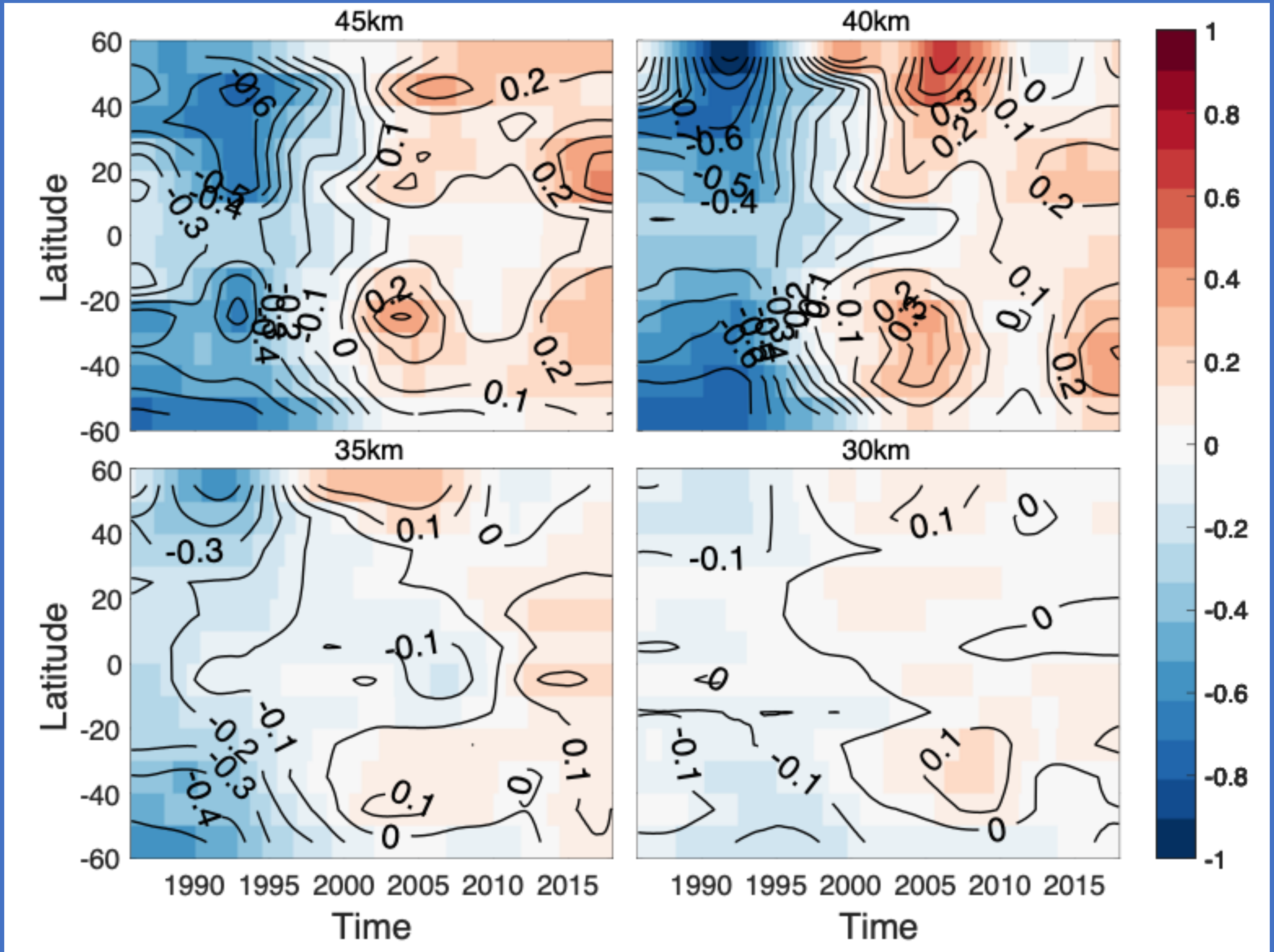

Ozone yearly trends



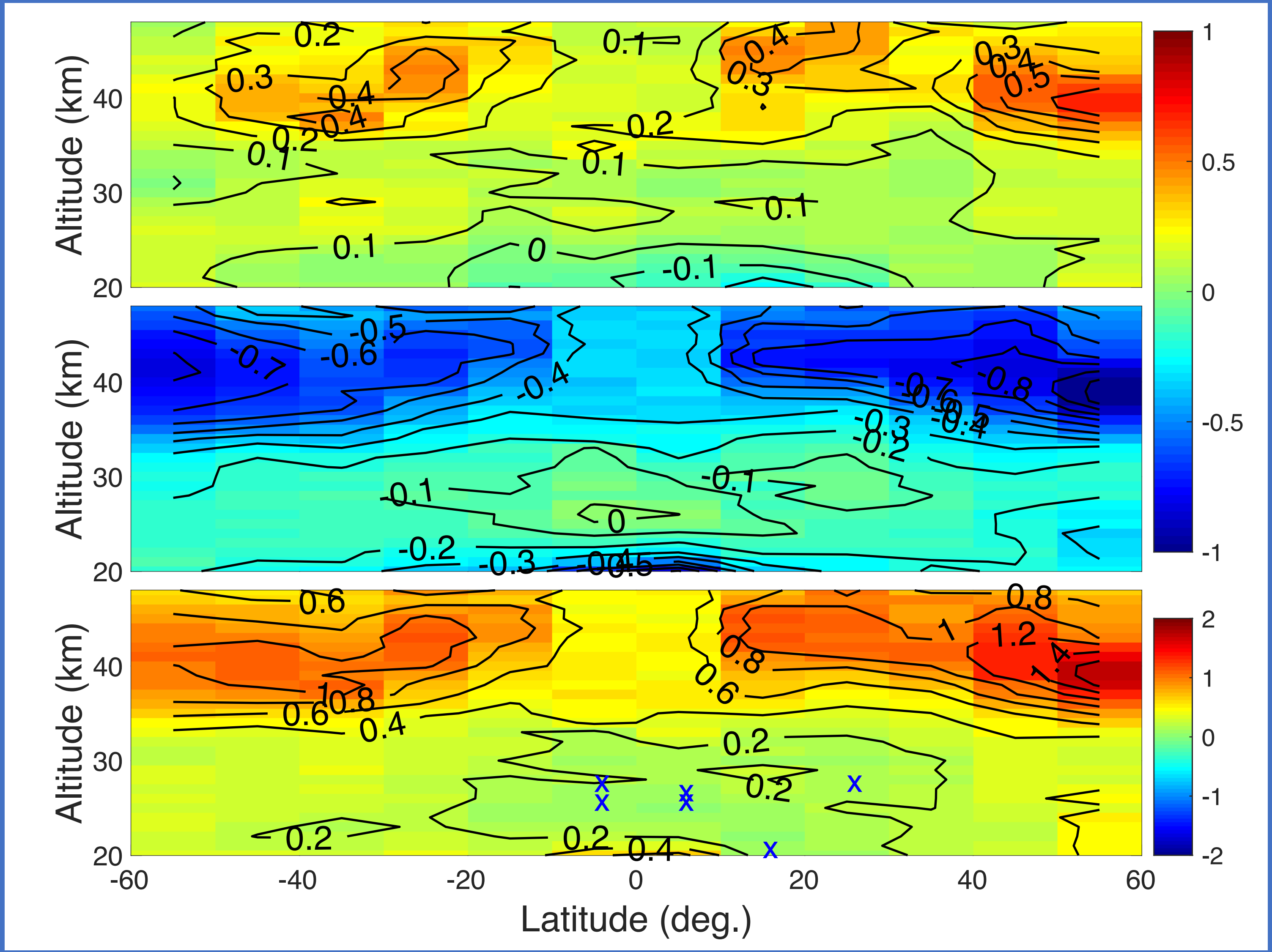
# Trends with wide latitudinal bands at selected years



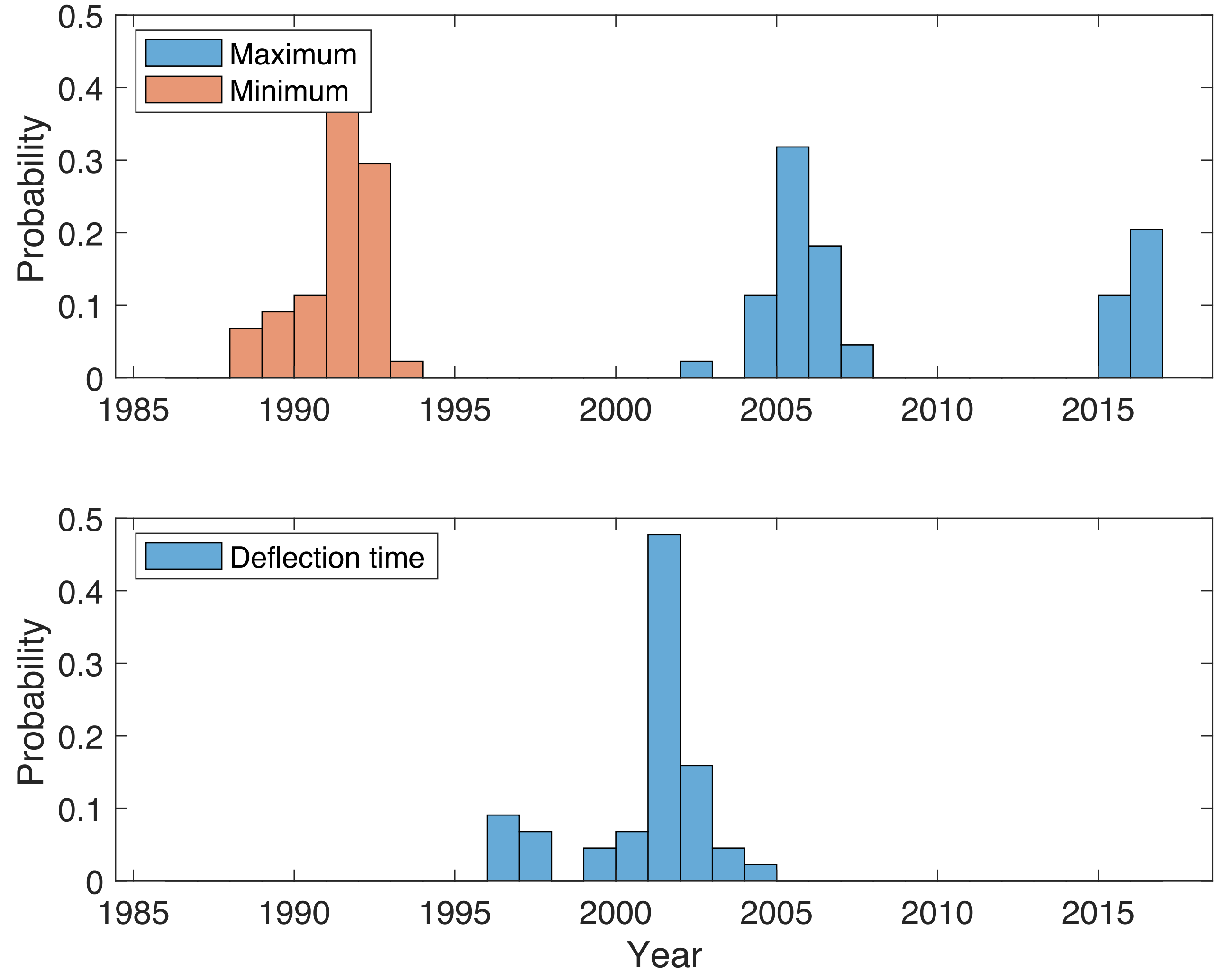
# Trends at 4 altitudes



# All time max, min and max-min of trend values



# Trend minimum and maximum timings. Deflection times.
# 35-45 km 20N-60N.




Deflection time
here= time of first
positive trend

# Conclusions

DLM-method offers an interesting new way to analyse time series. It does not assume linear trend, which in closed  system (here atmosphere) is only a short time approximation.

DLM results are sensitive to prior estimates of uncertainties

Trend estimates are, however, robust

In this work trend, annual and semi-annual cycles and autoregression were allowed to vary significantly. Other proxies had sharp prior uncertainties.

To relax additional states (like solar), methods to validate DLM-results are needed. Better understanding of the proxy couplings are needed, too.