ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Modelling trends in satellite observation time series using dynamic regression

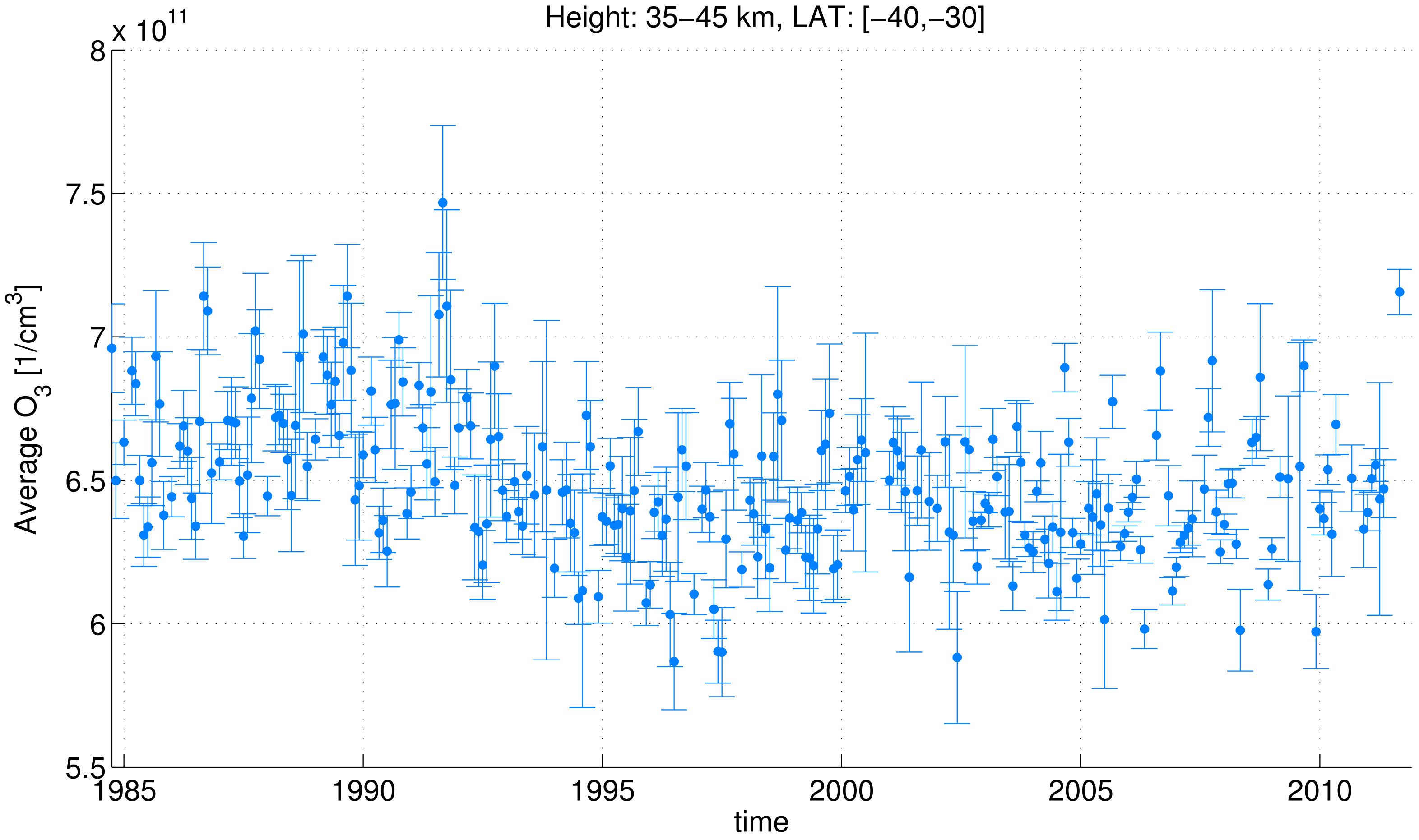
Marko Laine
Erkki Kyrölä

Finnish Meteorological Institute

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Time series analysis

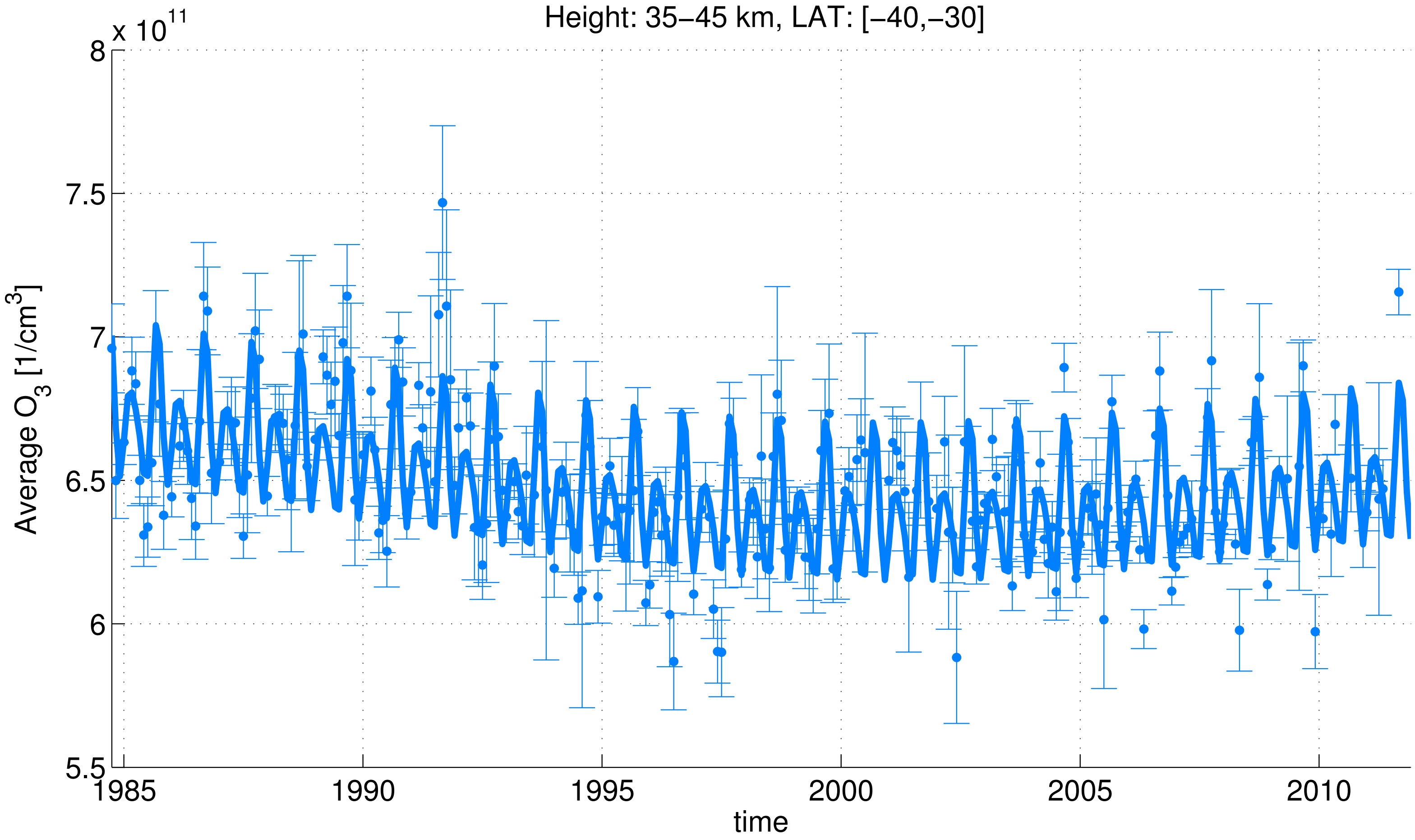
Components of ozone time series …



ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

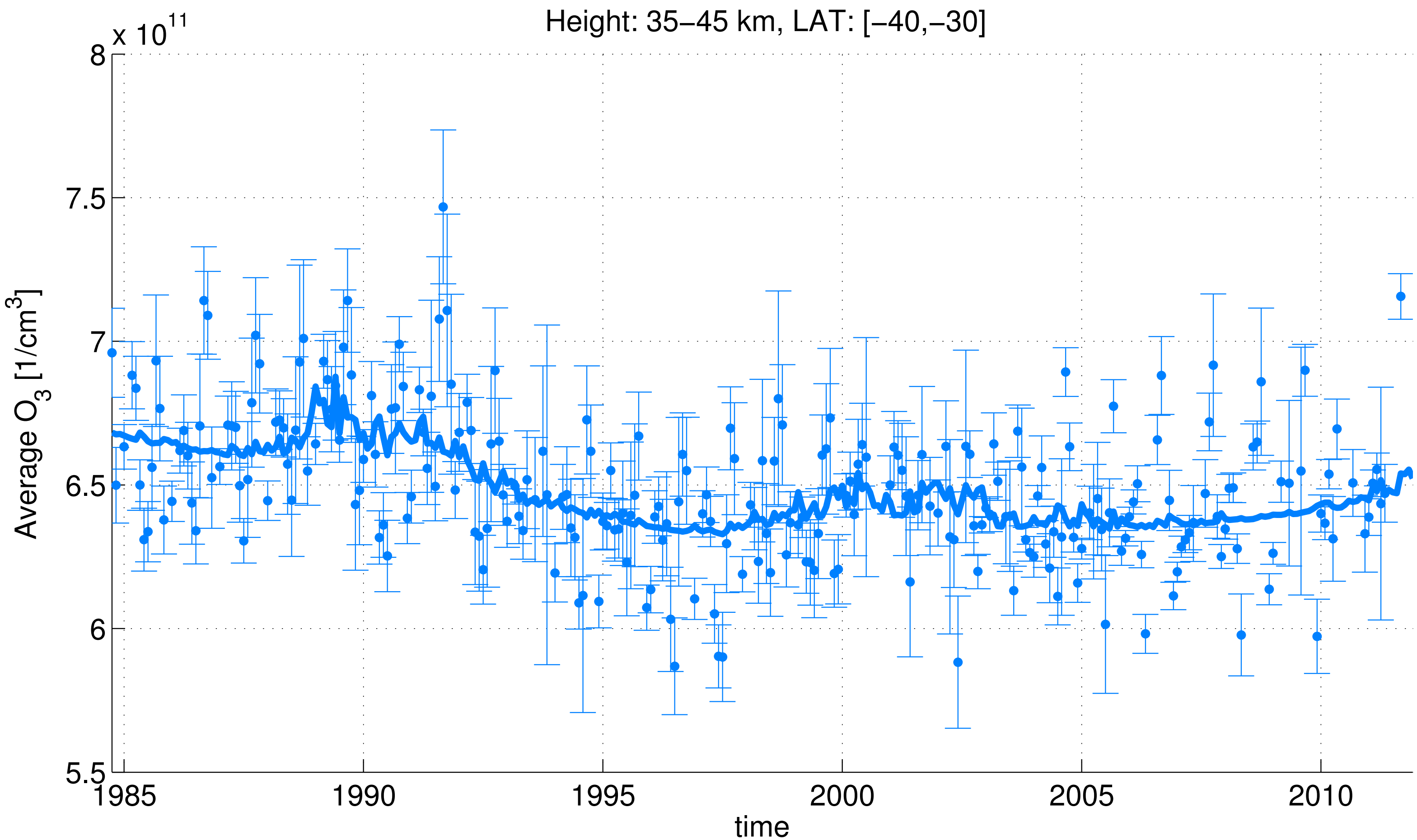
Time series analysis

Seasonal variation



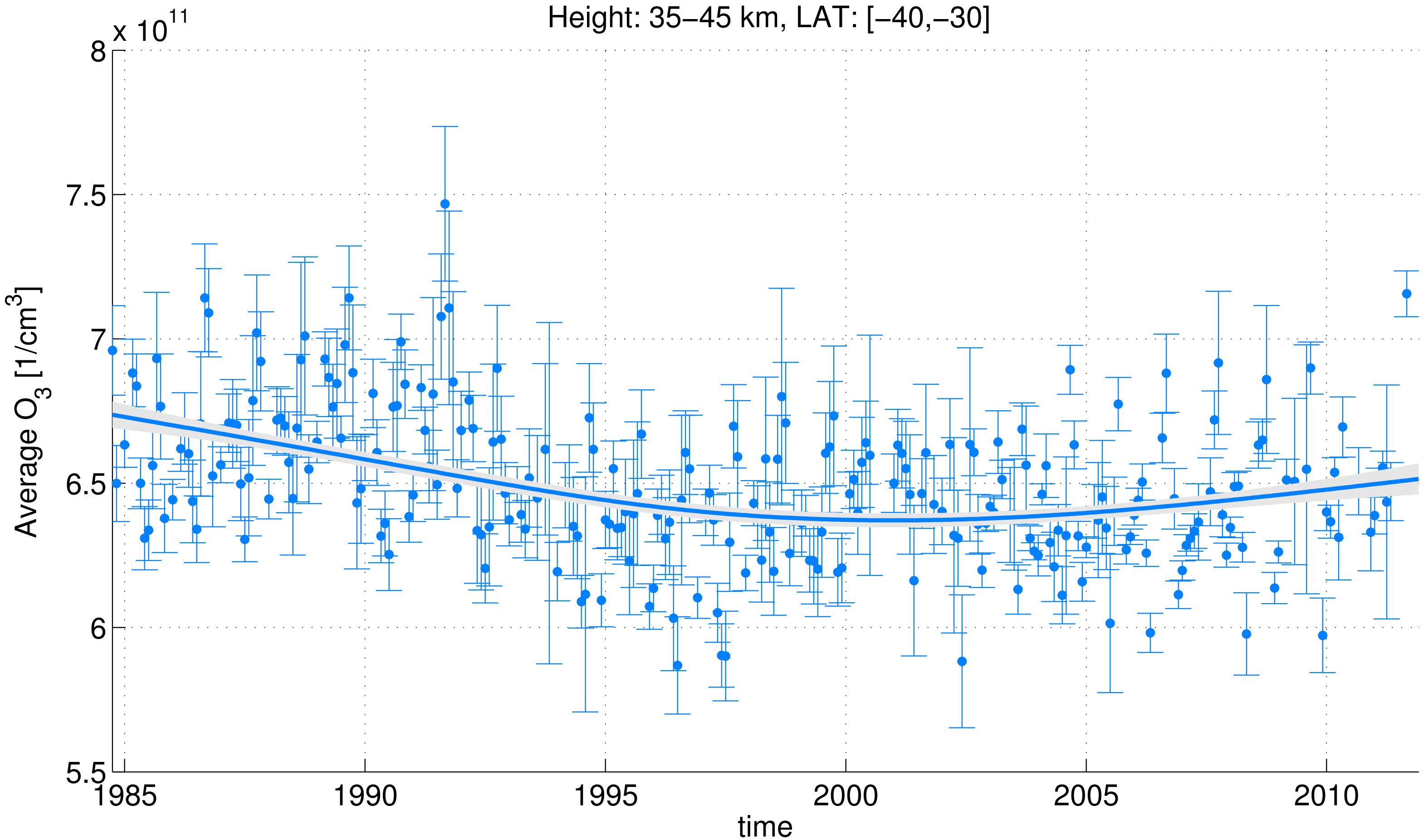
Time series analysis

Solar effect



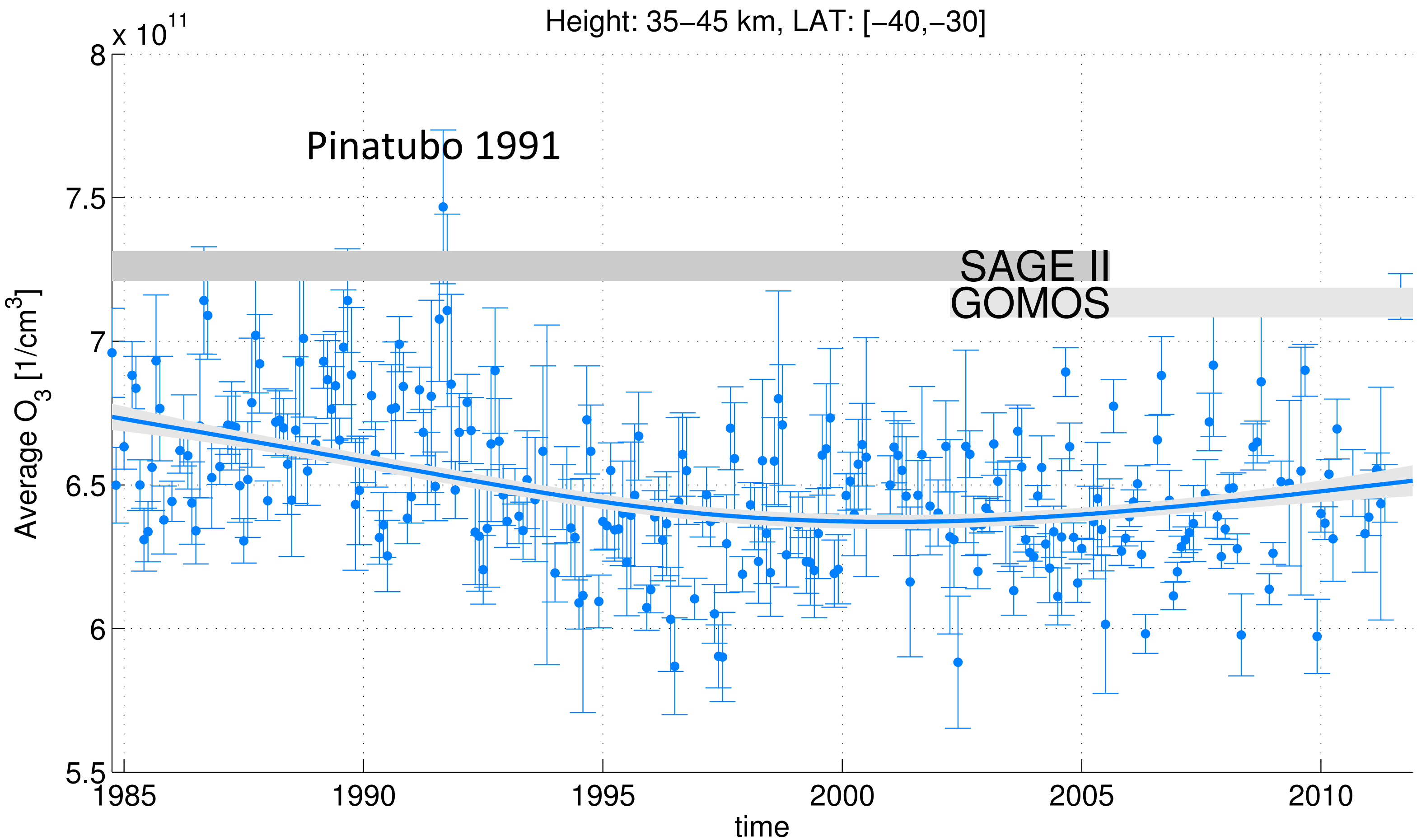
Time series analysis

Change in the background level



Time series analysis

Data quality, different instruments, volcanic events



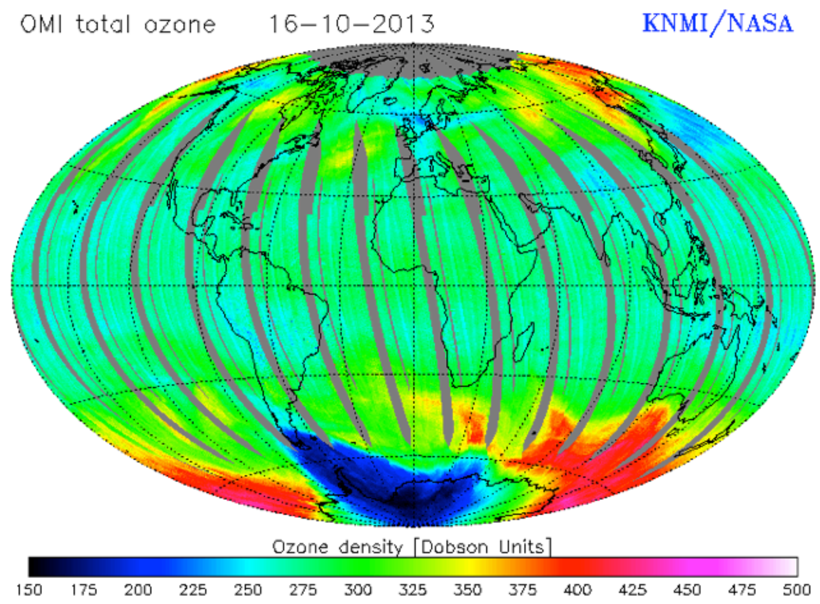
Challenges in satellite time series

- Non-stationary behaviour driven by external forcing.
- Persistent long range correlations.
- Long memory, slower than exponential decay of the autocorrelation function.
- Long time behaviour hard to estimate from short series.

- Non uniform sampling in space and time.
- Data combined from different instruments.
- Missing values.
- Instrument bias, aging, ...

Stratospheric ozone

- Stratospheric ozone is the primary UV radiation shield.
- The ozone layer prevents harmful solar energy (the biologically damaging ultraviolet radiation) from reaching the Earth's surface.
- The concentration at ozone layer around 20 km is only 2-8 ppm.
- In 1985 scientist found "ozone hole" over the Antarctica, which was attributed to human produced chlorofluorocarbons (CFC) emissions.
- Ozone levels have been decreasing ~4% per decade.



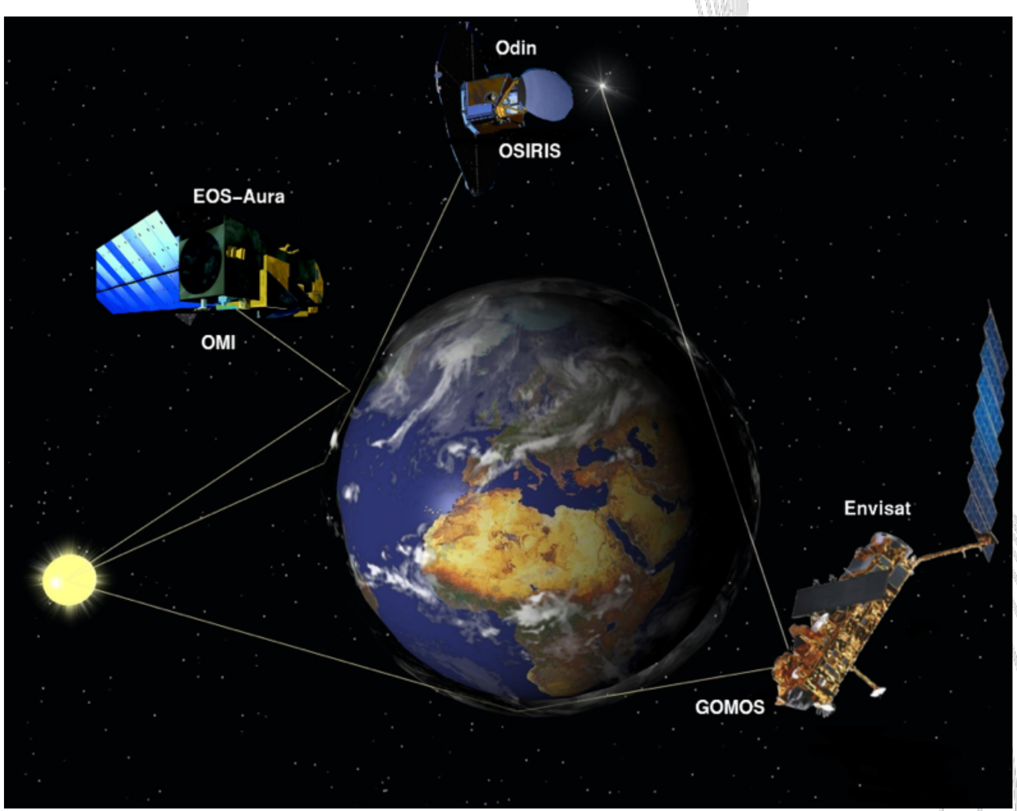
Recovery of ozone after the Montreal treaty for CFC compounds

- The Montreal Protocol in 1987 agreed on a gradual ban of the CFC emissions.
- "Perhaps the single most successful international agreement to date".
- CFC concentrations have indeed levelled out or begun to decrease slowly in recent years.
- Because of the reservoir effect of existing CFCs in the atmosphere, it may be 50 years before ozone levels recover.

Monitoring ozone from space

- In March 2002 European Space Agency ESA launched Envisat satellite, that had GOMOS instrument specially designed (partly at FMI) to observe vertical ozone profiles. The Envisat mission ended May 2012.

- Other satellite instruments capable of producing ozone profiles are, for example, SAGE II/ ERBS (1984-2005), OMI/Aura (2004-).



Picture by Seppo Hassinen FMI

Satellite observation time series

- There are now almost 30 years of ozone profile data.
- One important ongoing task is the production of homogenized data sets of ozone and other essential climate variable for climate scientists.
- The topic in this talk is the statistical analysis of observed variability and trends in satellite observations.

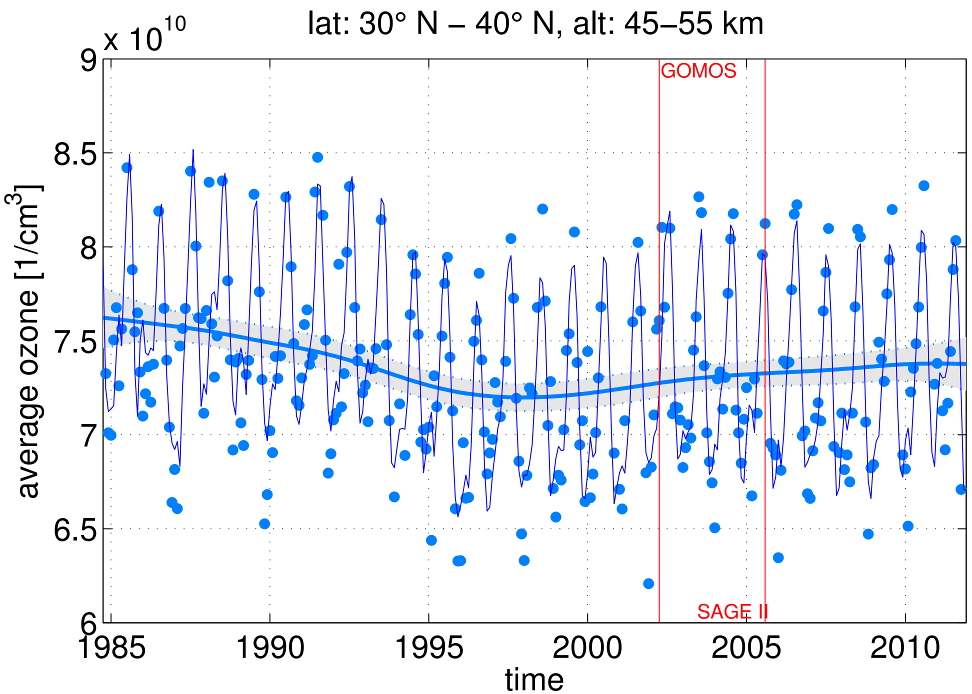
- Kyrölä, E., Laine, M., Sofieva, V., Tamminen, J., Päivärinta, S.-M., Tukiainen, S., Zawodny, J., Thomason, L. (2013): Combined SAGE II-GOMOS ozone profile data set for 1984–2011 and trend analysis of the vertical distribution of ozone *Atmos. Chem. Phys.*, 13.
- Laine, M., Latva-Pukkila, N., Kyrölä, E. (2014): Analysing time-varying trends in stratospheric ozone time series using state the space approach, *Atmos. Chem. Phys.*, 14.

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Dynamic linear model (DLM) for trend analysis

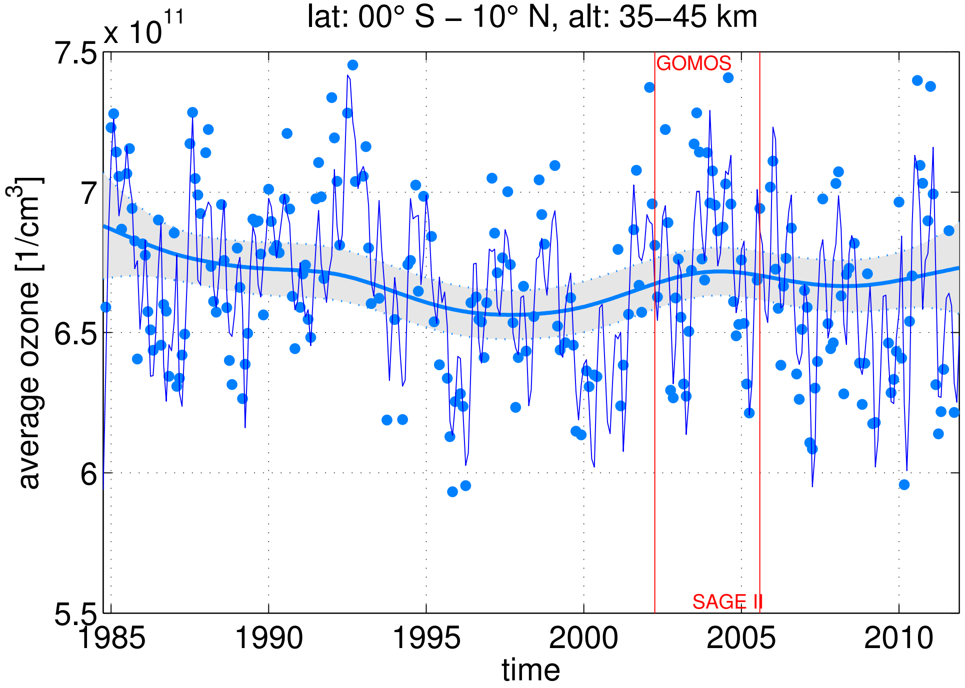
- We apply dynamic regression by DLM to study trends in climatic time series.
- By process description of the model components (trend, seasonality, proxies).
- Structural parameters are estimated by Markov chain Monte Carlo (MCMC).

- Use hierarchical statistical model for uncertainties in data, process, and parameters.
- With verifiable statistical assumptions.



What is trend?

- Trend is a change in the background level of the process.
- We are interested in (smooth) long term (decade) change attributed to ozone recovery.
- Need to filter out seasonality, external variation driven by known phenomena, correlated random noise.



ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Dynamic linear model as hierarchical statistical model

$y_t = F_t x_t + v_t$ $\quad v_t \sim N(0, V_t)$

$x_t = G_t x_{t-1} + w_t$ $\quad w_t \sim N(0, W_t)$

- y_t: observations,
- x_t: hidden model states,
- F_t: observation operator,
- G_t: model operator,
- v_t :observation uncertainty,
- w_t :model uncertainty.

- Observation model: $p(y_t | x_t, \theta)$
- Process model: $p(x_t | x_{t-1}, \theta)$
- Parameter model: $p(\theta)$

- θ: structural and variance parameters in F_t, G_t, V_t, and W_t.

- Bayes formula:

$$p(x_{1:n}, \theta | y_{1:n}) \propto \prod_{t=1}^{n} p(y_t | x_t, \theta) p(x_t | x_{t-1}, \theta) p(\theta)$$

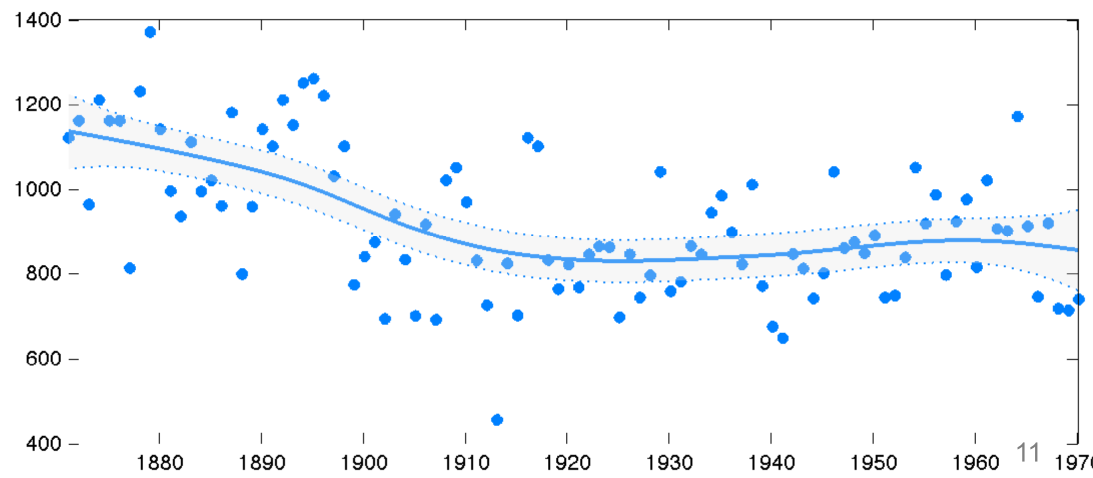
Simple example: spline smoothing

$y_t = \mu_t + \varepsilon_{obs}$,
$\mu_t = \mu_{t-1} + \alpha_t + \varepsilon_{level}$,
$\alpha_t = \alpha_{t-1} + \varepsilon_{trend}$,

$\varepsilon_{obs} \sim N(0, \sigma^2_{obs})$, observations
$\varepsilon_{level} \sim N(0, \sigma^2_{level})$, local level
$\varepsilon_{trend} \sim N(0, \sigma^2_{trend})$, local trend

$$G = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad F = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad x_t = \begin{bmatrix} \mu_t & \alpha_t \end{bmatrix}^T, \theta = \begin{bmatrix} \sigma^2_{\text{obs}} & \sigma^2_{\text{level}} & \sigma^2_{\text{trend}} \end{bmatrix}.$$

When $\sigma_{level} = 0$, this is cubic spline smoothing with smoothness parameter $\lambda = \sigma^2_{trend}/\sigma^2_{obs}$.

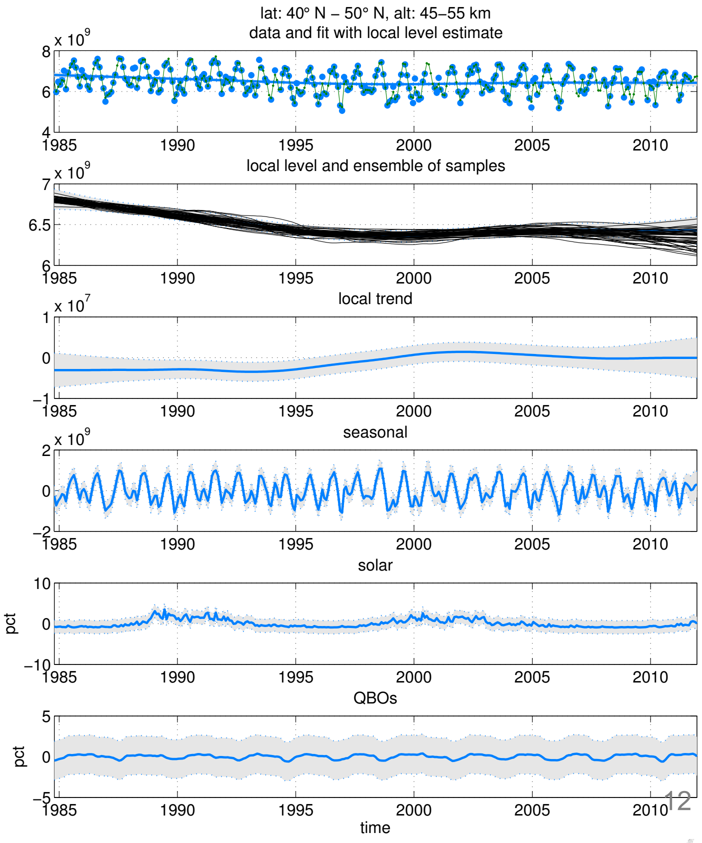


General model for trend, seasonality, AR error, proxies

$$y_t = \mu_t + \gamma_t + \beta_t X_t + \eta_t + \varepsilon_{\mathrm{obs},t}$$

μ_t: background level, the trend,
γ_t: seasonal effect,
β_t: coefficient for proxy covariates X_t,
η_t: autoregressive error term,
$\varepsilon_{\mathrm{obs},t}$: observation uncertainty.

All model components defined by suitable model operator G_t and can depend on time index t.



The system matrices

$$x_t = \begin{bmatrix} \mu_t & \alpha_t & \psi_{t,1} & \psi^*_{t,1} & \psi_{t,2} & \psi^*_{t,2} & \eta_t & \beta_{1,t} & \beta_{2,t} \end{bmatrix}^T$$

$$G = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos\left(\frac{\pi}{6}\right) & \sin\left(\frac{\pi}{6}\right) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\sin\left(\frac{\pi}{6}\right) & \cos\left(\frac{\pi}{6}\right) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos\left(\frac{\pi}{3}\right) & \sin\left(\frac{\pi}{3}\right) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\sin\left(\frac{\pi}{3}\right) & \cos\left(\frac{\pi}{3}\right) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \rho & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$F_t = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & \mathrm{solar}(t) & \mathrm{qbo}(t) \end{bmatrix}$$

$$W = \mathrm{diag}\begin{bmatrix} 0 & \sigma^2_\alpha & \sigma^2_\psi & \sigma^2_\psi & \sigma^2_\psi & \sigma^2_\psi & \sigma^2_\eta & \sigma^2_{\beta,1} & \sigma^2_{\beta,2} \end{bmatrix}$$

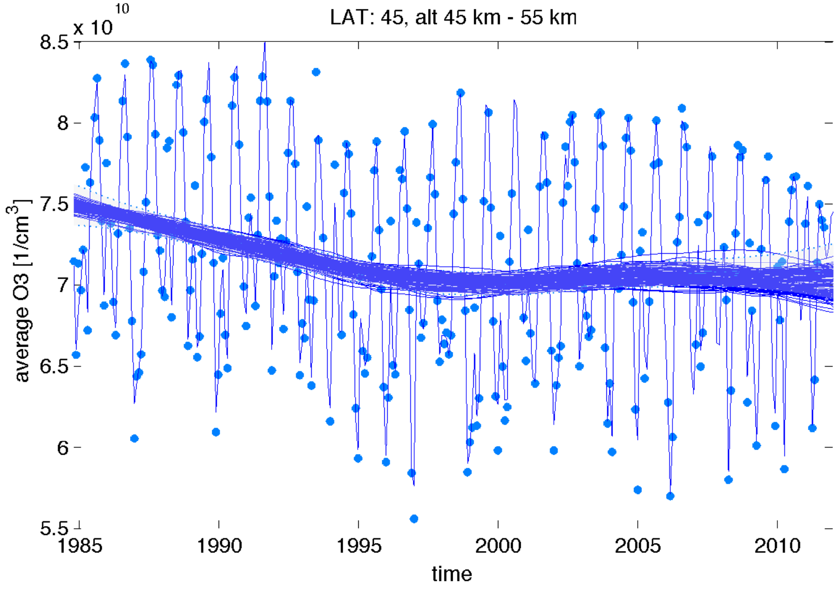
$$\theta = \begin{bmatrix} \sigma_\alpha & \sigma_\psi & \sigma_\eta & \sigma^2_{\beta,1} & \sigma^2_{\beta,2} & \rho \end{bmatrix}^T$$

Full sampling for trend statistics

- Kalman formulas give marginal distributions $p(x_t|y_{1:n}, \theta)$.
- We can simulate model states from $p(x_{1:n}|y_{1:n}, \theta)$.
- Need MCMC to simulate from

$$p(x_{1:n}|y_{1:n}) = \int p(x_{1:n}|y_{1:n}, \theta)\, d\theta.$$

- Then we get uncertainty estimates for trend related statistics.



Flowchart for estimation

Model components

t	time index	$t = 1, \ldots, N$
$\boldsymbol{y}_t$	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

State space equations

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

Kalman filter recursion

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

Kalman filter likelihood

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

Kalman smoother recursion

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

Simulation smoother

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

MCMC algorithm

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

Monte Carlo sampling

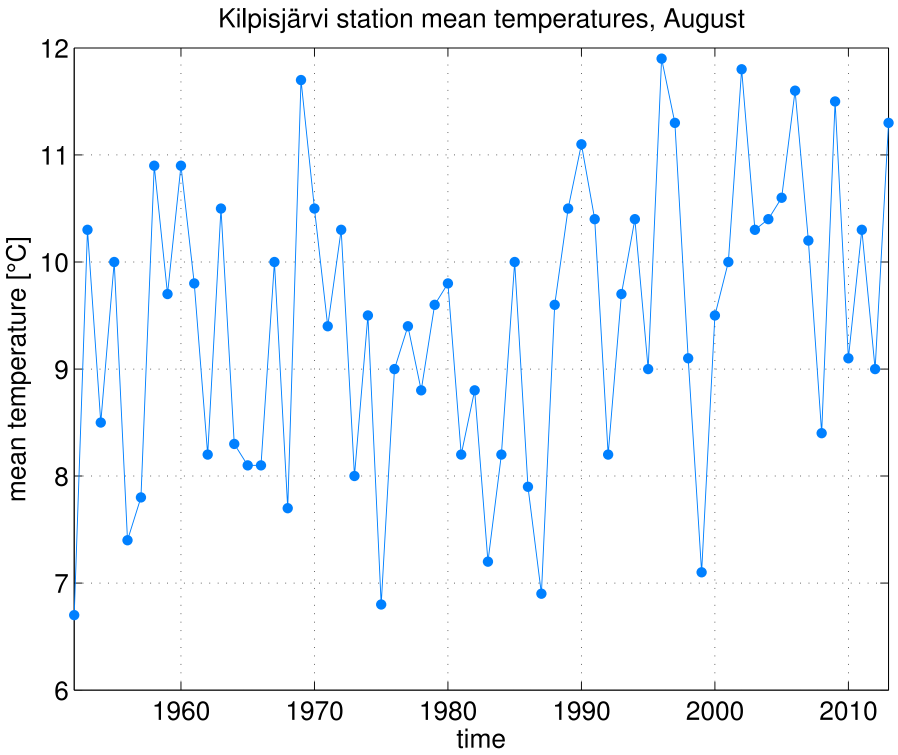
sample: $p(\boldsymbol{x}_{1:N},\boldsymbol{\theta}|\boldsymbol{y}_{1:N})$

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

Posterior predictive inference of trends

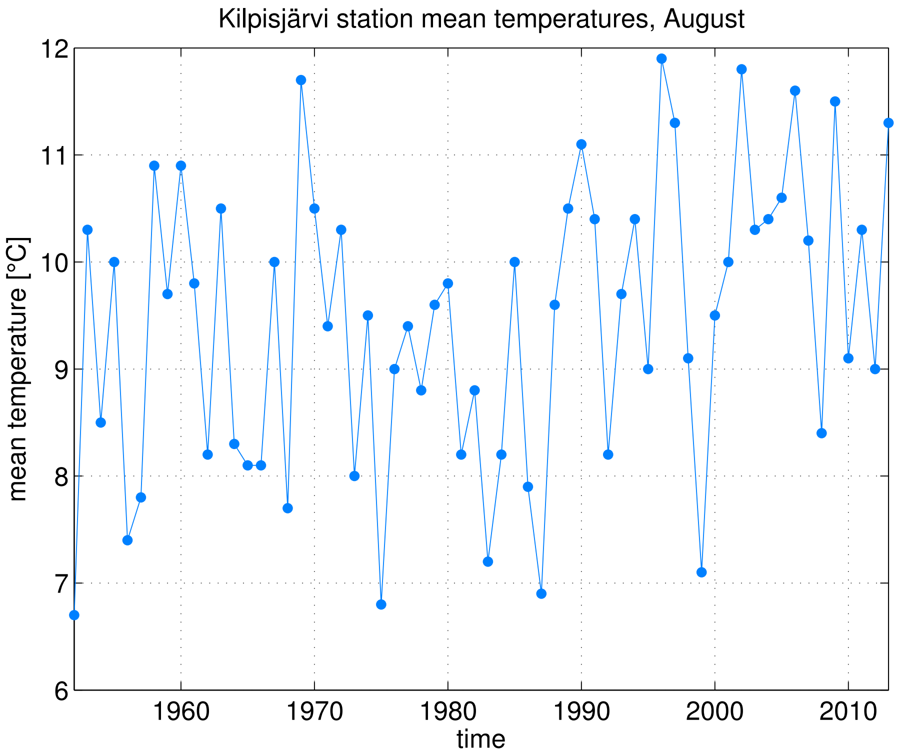
sample: $p(\text{trend}(\boldsymbol{x}_{1:N})|\boldsymbol{y}_{1:N})$

Kilpisjärvi (69°2’54”N, 20°47’42”E) temperatures



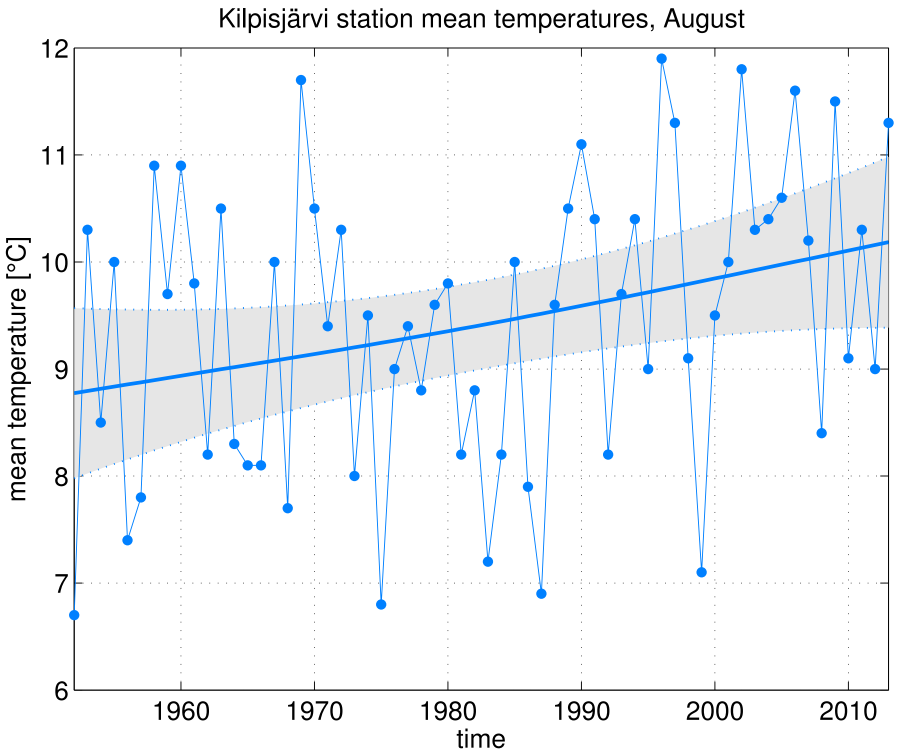
Kilpisjärvi (69°2’54”N, 20°47’42”E) temperatures

- Monthly mean temperatures in August at Kilpisjärvi.



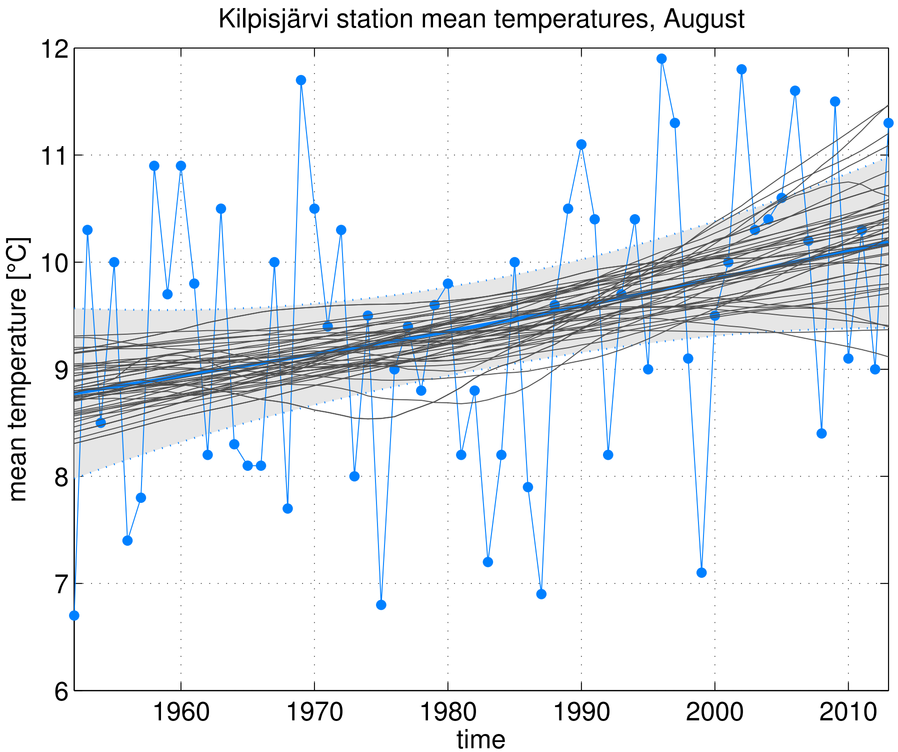
Kilpisjärvi (69°2’54”N, 20°47’42”E) temperatures

- Monthly mean temperatures in August at Kilpisjärvi.
- Fitted DLM model.



Kilpisjärvi (69°2’54”N, 20°47’42”E) temperatures

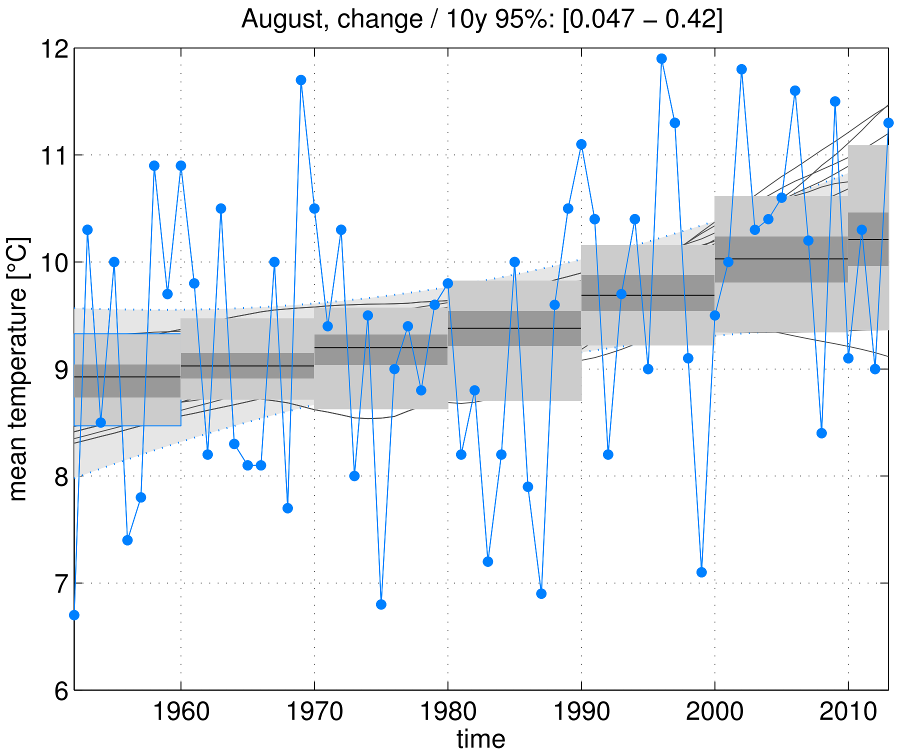
- Monthly mean temperatures in August at Kilpisjärvi.
- Fitted DLM model.
- Sample from the background level.



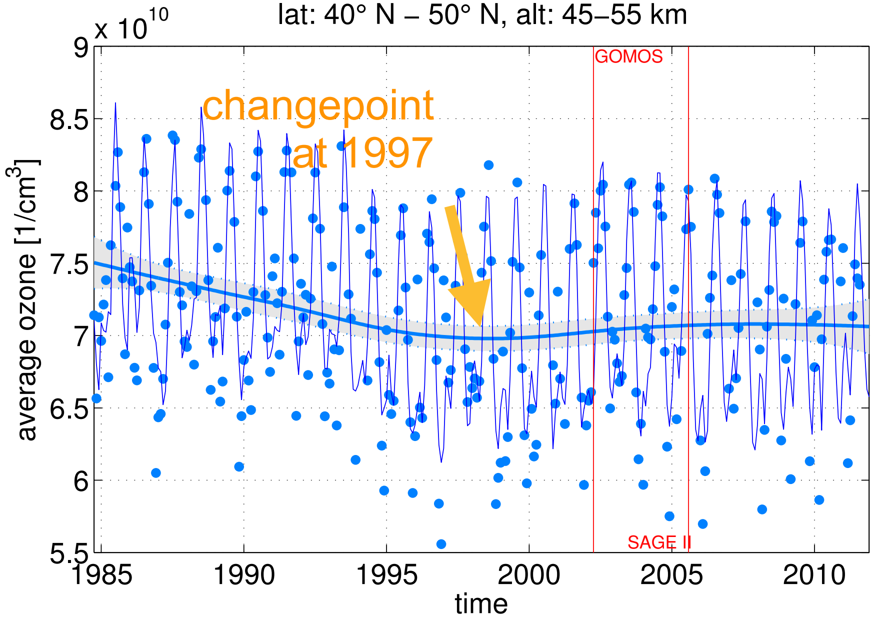
ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

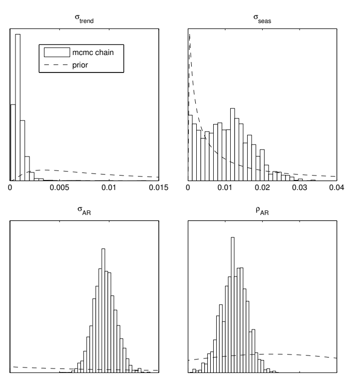
Kilpisjärvi (69°2’54”N, 20°47’42”E) temperatures

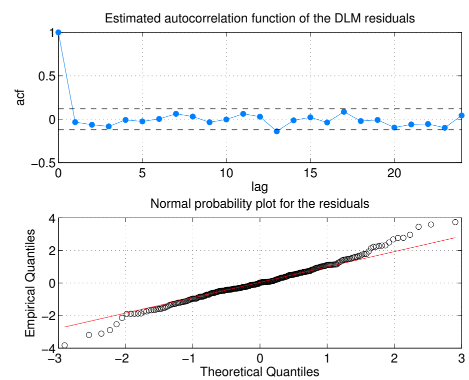
- Monthly mean temperatures in August at Kilpisjärvi.
- Fitted DLM model.
- Sample from the background level.
- Estimated decadal averages.

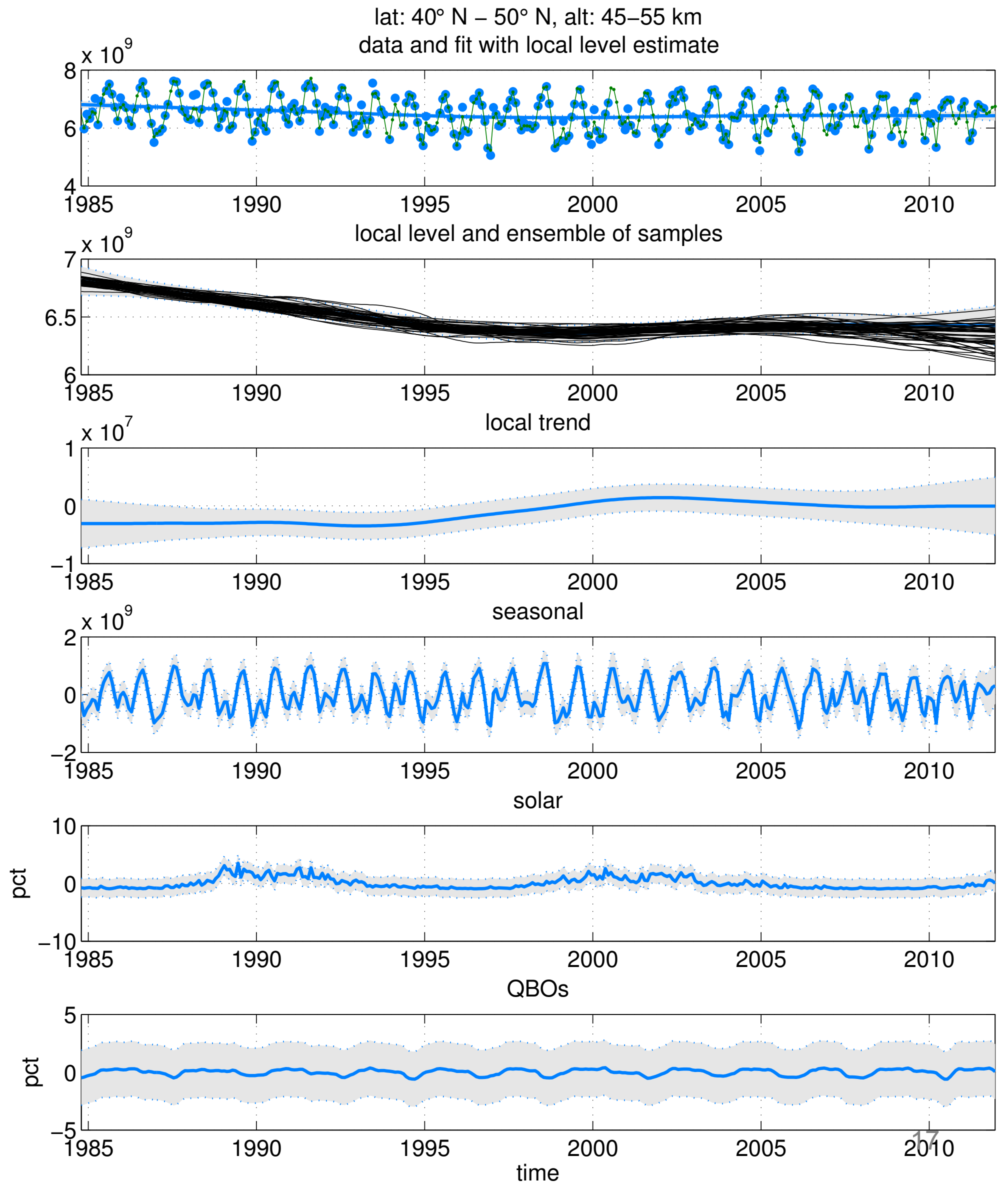


Stratospheric ozone







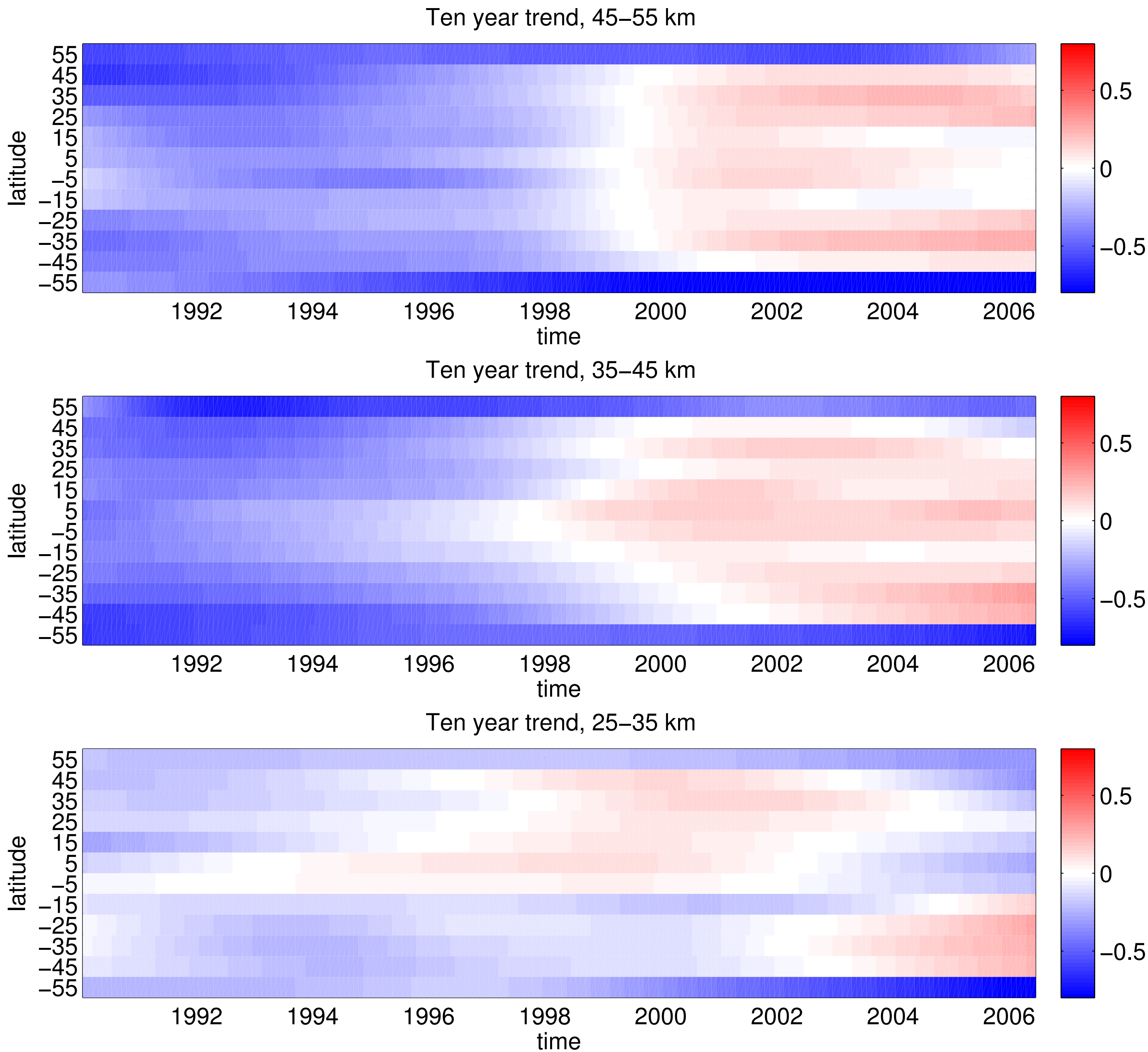


Laine, et al 2013

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

- Combined trend analysis of stratospheric ozone from SAGE II and GOMOS instruments.

- In mid-latitudes, at 35 to 55 km, we see ~ %5 / decade increase after around 1997-2000.



Thank You!

- A. C. Harvey. Forecasting, structural time series and the Kalman filter. Cambridge University Press, 1990.
- Kyrölä, E., Laine, M., Sofieva, V., Tamminen, J., Päivärinta, S.-M., Tukiainen, S., Zawodny, J., Thomason, L.: Combined SAGE II-GOMOS ozone profile data set for 1984–2011 and trend analysis of the vertical distribution of ozone, *Atmos. Chem. Phys.*, 13, 10645–10658, 2013.
- M. Laine, N. Latva-Pukkila, E. Kyrölä: Analysing time-varying trends in stratospheric ozone time series using the state space approach, *Atmos. Chem. Phys.*, 14, 9707–9725, 2014.
- Matlab toolbox for DLM calculations for time series at http://helios.fmi.fi/~lainema/dlm.