



Analysing time-varying trends in stratospheric ozone time series using the state space approach

M. Laine¹, N. Latva-Pukkila^{1,2}, and E. Kyrölä¹

¹Finnish Meteorological Institute, Earth Observation Unit, P.O. Box 503, 00101 Helsinki, Finland

²Department of Mathematics and Statistics, P.O. Box (MaD), 40014 University of Jyväskylä, Jyväskylä, Finland

Correspondence to: M. Laine (marko.laine@fmi.fi)

Received: 11 July 2013 – Published in Atmos. Chem. Phys. Discuss.: 6 August 2013

Revised: 4 August 2014 – Accepted: 14 August 2014 – Published: 16 September 2014

Abstract. We describe a hierarchical statistical state space model for ozone profile time series. The time series are from satellite measurements by the Stratospheric Aerosol and Gas Experiment (SAGE) II and the Global Ozone Monitoring by Occultation of Stars (GOMOS) instruments spanning the years 1984–2011. Vertical ozone profiles were linearly interpolated on an altitude grid with 1 km resolution covering 20–60 km. Monthly averages were calculated for each altitude level and 10° wide latitude bins between 60° S and 60° N. In the analysis, mean densities are studied separately for the 25–35, 35–45, and 45–55 km layers. Model variables include the ozone mean level, local trend, seasonal oscillations, and proxy variables for solar activity, the Quasi-Biennial Oscillation (QBO), and the El Niño–Southern Oscillation (ENSO).

This is a companion paper to Kyrölä et al. (2013), where a piecewise linear model was used together with the same proxies as in this work (excluding ENSO). The piecewise linear trend was allowed to change at the beginning of 1997 in all latitudes and altitudes. In the modelling of the present paper such an assumption is not needed as the linear trend is allowed to change continuously at each time step. This freedom is also allowed for the seasonal oscillations whereas other regression coefficients are taken independent of time. According to our analyses, the slowly varying ozone background shows roughly three general development patterns. A continuous decay for the whole period 1984–2011 is evident in the southernmost latitude belt 50–60° S in all altitude regions and in 50–60° N in the lowest altitude region 25–35 km. A second pattern, where a recovery after an initial decay is followed by a further decay, is found at northern latitudes from the equator to 50° N in the lowest altitude region (25–35 km) and between 40° N and 60° N in the

35–45 km altitude region. Further ozone loss occurred after 2007 in these regions. Everywhere else a decay is followed by a recovery. This pattern is shown at all altitudes and latitudes in the Southern Hemisphere (10–50° S) and in the 45–55 km layer in the Northern Hemisphere (from the equator to 40° N). In the 45–55 km range the trend, measured as an average change in 10 years, has mostly turned from negative to positive before the year 2000. In those regions where the “V” type of change of the trend is appropriate, the turning point is around the years 1997–2001. To compare results for the trend changes with the companion paper, we studied the difference in trends between the years from 1984 to 1997 and from 1997 to 2011. Overall, the two methods produce very similar ozone recovery patterns with the maximum trend change of 10 % in 35–45 km. The state space method (used in this paper) shows a somewhat faster recovery than the piecewise linear model. For the percent change of the ozone density per decade the difference between the results is below three percentage units.

1 Introduction

Time series constructed from satellite remote sensing observations provide important information about variability and trends in atmospheric chemical composition. Many satellite time series provide global coverage of the measurement and some of the data sets run from the 1980s. The analysis of trends, both natural and anthropogenic, is complicated by natural variability and forcing affecting stratospheric chemical compositions. In this study, the recovery of stratospheric ozone from the depletion caused by

chlorofluorocarbon (CFC) compounds is studied using a statistical time series model.

Slow background changes in stratospheric ozone are easily masked by both seasonal and irregular natural variabilities. Thus, the requirements are stringent for the stability of ozone observations. Self-calibrating occultation instruments are good candidates for such a task. The observations analysed in this work consist of satellite measurements by the SAGE II and GOMOS instruments, operational during 1984–2005 and 2002–2012, respectively. The data set used here spans the years 1984–2011. Vertical ozone profiles were linearly interpolated on an altitude grid with 1 km resolution covering 20–60 km. Monthly averages were calculated for each altitude level and 10° wide latitude bins between 60° S and 60° N. Combining the observations from different instruments having different measurement principles is a challenge. Kyrölä et al. (2013) explain the data set and its construction in more detail. Here the analysis is done with both the original 1 km vertical spacing and by calculating mean densities over 10 km intervals.

There is a wealth of literature concerning the analysis of atmospheric time series. A good reference to stratospheric ozone time series regression analysis is SPARC (1998). A recent study that reviews the challenges and problems in trend analysis of climatic time series was published by Bates et al. (2012), and a general trend analysis reference is Chandler and Scott (2011). For state space and functional analysis of atmospheric time series of similar type to that performed here, see Lee and Berger (2003) and Meiring (2007).

This paper studies the feasibility and practical implementation of a state space approach for atmospheric time series analysis by defining a dynamic linear model (DLM) for stratospheric ozone time series. “Dynamic” means here that the regression coefficients can evolve in time. This makes it possible to describe and analyse smooth changes in the average background behaviour of ozone. Model variables include the ozone mean level, local trend, seasonal oscillations, and proxy variables for solar activity, the Quasi-Biennial Oscillation (QBO), and the El Niño–Southern Oscillation (ENSO). We do not claim novelty in the presented methods themselves, but argue that they should be more extensively applied in the analysis of climatic time series and provide a simple framework for time series analyses that can be generalized to more comprehensive studies. In this paper, we describe the necessary steps for applying the methods.

A typical feature in atmospheric time series is that they are not stationary but exhibit both slowly varying and abrupt changes in the distributional properties. These are caused by external forcing such as changes in the solar activity or volcanic eruptions. Further, the data sampling is often non-uniform – there are data gaps, and the uncertainty of the observations can vary. When observations are combined from various sources there will be instrument and retrieval method related biases. The differences in sampling lead to uncertainties, too. Straightforward linear regression analysis leaves the

model residuals correlated as not all variability can be explained by a static linear structure. Usually this is compensated by allowing some correlation structure to the model observation error by using, e.g. an autoregressive model. If the residual correlation is not accounted for, the model uncertainty analyses are misleading. A simple autoregressive process can explain some of the unmodelled systematic variations by correlated noise, again confusing the analyses. In conclusion, much care in interpretation is needed for those standard classical statistical time series methods that require stationarity, such as the ARIMA (autoregressive integrated moving average) approach. The more general approach discussed in this paper makes use of dynamic linear models and Kalman filter type sequential estimation algorithms.

State space models, sometimes called hidden Markov models or structural time series models, are well known and documented in time series literature, e.g. Harvey (1990); Hamilton (1994); Migon et al. (2005). Modern computationally oriented references are Durbin and Koopman (2012) and Petris et al. (2009). Here, we review the basic properties relevant to the analysis of atmospheric ozone time series data and explain the necessary steps to fit the model to monthly time series observations and how to assess the uncertainties in the trend estimation.

The structure of this paper is the following. The data sets and the statistical model are described in Sect. 2. Results of the statistical time series analyses are given in Sect. 3 and the paper ends with discussion and conclusions in Sect. 4. Appendix A contains mathematical and computational details.

2 Materials and methods

2.1 Ozone time series from satellite observations

We use a combination of two ozone data sets. The first consists of solar occultation measurements of ozone in the stratosphere and lower mesosphere from the SAGE II instrument (Chu et al., 1989) operational during 1984–2005. The second is the GOMOS instrument (Bertaux et al., 2010) that measured ozone in the stratosphere, mesosphere and lower thermosphere during 2002–2012 using stellar occultations. The individual data sets have been homogenized to form a combined time series from 1984 to 2011. The stability of the SAGE II and GOMOS instruments, the construction of the combined time series, data screening, bias correction, and other issues are discussed in more detail in the companion paper by Kyrölä et al. (2013). Four proxy time series are used for the solar flux, Quasi-Biennial Oscillations and ENSO. Except for ENSO, these proxies are the same as used in the companion paper. For ENSO we use the Multivariate ENSO Index (MEI) from NOAA¹.

¹The MEI index is available from <http://www.esrl.noaa.gov/psd/enso/mei/>.

2.2 Statistical time series model

A general linear state space model with Gaussian errors can be written with an observation equation and a state evolution equation as

$$\mathbf{y}_t = \mathbf{F}_t \mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim N_p(0, \mathbf{V}_t), \quad (1)$$

$$\mathbf{x}_t = \mathbf{G}_t \mathbf{x}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N_q(0, \mathbf{W}_t), \quad (2)$$

where \mathbf{y}_t is a vector of length p containing the observations and \mathbf{x}_t is a vector of length q of unobserved states of the system at time t . The state variables are used to describe the various components of the time series model, such as mean level, trend, seasonality and the effect of proxy variables. Matrix \mathbf{F}_t is the observation operator mapping the unobserved states to the observations and matrix \mathbf{G}_t is the model evolution operator giving the dynamics of the states. In this basic formulation the uncertainties \mathbf{v}_t and \mathbf{w}_t are assumed to be Gaussian, with observation uncertainty covariance \mathbf{V}_t and model error covariance \mathbf{W}_t . Above, $N_p(0, \mathbf{V}_t)$ stands for p -dimensional Gaussian distributions, with vector of zeros as mean and \mathbf{V}_t as the $p \times p$ covariance matrix. The time index t will go from 1 to N , the length of the time series to be analysed. In the following, the matrices defining the model will mostly be time invariant, i.e. $\mathbf{G}_t = \mathbf{G}$, etc., and we will usually drop the time subscript, still retaining it in general formulas that are not specific to this particular time series application.

In this work, we use a DLM to explain variability in the ozone time series with four components: smooth locally linear trend, seasonal effect, effect of forcing via proxy variables, and noise that is allowed to have autoregressive correlation. All components are built using the state space approach.

To describe the trend we start with a simple local level and locally linear trend model that has two hidden states $\mathbf{x}_t = [\mu_t, \alpha_t]^T$, where μ_t is the mean level and α_t is the change in the level from time t to time $t + 1$. In addition we need stochastic terms for the observational error and for the allowed change in the dynamics of the trend and the level. These are defined by Gaussian “ ϵ ” terms below. The system can be written by equations

$$\mathbf{y}_t = \mu_t + \epsilon_{\text{obs}}, \quad \epsilon_{\text{obs}} \sim N(0, \sigma_{\text{obs}(t)}^2), \quad (3)$$

$$\mu_t = \mu_{t-1} + \alpha_{t-1} + \epsilon_{\text{level}}, \quad \epsilon_{\text{level}} \sim N(0, \sigma_{\text{level}}^2), \quad (4)$$

$$\alpha_t = \alpha_{t-1} + \epsilon_{\text{trend}}, \quad \epsilon_{\text{trend}} \sim N(0, \sigma_{\text{trend}}^2). \quad (5)$$

In terms of the state space equations (1) and (2) this becomes

$$\mathbf{G}_{\text{trend}} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{F}_{\text{trend}} = [1 \quad 0],$$

$$\mathbf{W}_{\text{trend}} = \begin{bmatrix} \sigma_{\text{level}}^2 & 0 \\ 0 & \sigma_{\text{trend}}^2 \end{bmatrix}, \quad \mathbf{V}_t = [\sigma_{\text{obs}(t)}^2]. \quad (6)$$

Depending on the choice of variances σ_{level}^2 and σ_{trend}^2 , this will define a smoothly varying background level of the data series that is used to infer changes in atmospheric ozone.

Most atmospheric series exhibit seasonal variability. The seasonality can be modelled using harmonic functions. If the number of cyclic components is s , the full seasonal model has $s/2$ harmonics. For the k th harmony, with $k = 1, \dots, s/2$, we need to add two state variables. With monthly data we have $s = 12$ and the corresponding blocks of the model and observation matrices are

$$\mathbf{G}_{\text{seas}(k)} = \begin{bmatrix} \cos(k2\pi/12) & \sin(k2\pi/12) \\ -\sin(k2\pi/12) & \cos(k2\pi/12) \end{bmatrix},$$

$$\mathbf{F}_{\text{seas}(k)} = [1 \quad 0] \text{ and } \mathbf{W}_{\text{seas}(k)} = \begin{bmatrix} \sigma_{\text{seas}(k)}^2 & 0 \\ 0 & \sigma_{\text{seas}(k)}^2 \end{bmatrix}. \quad (7)$$

Here the state equation matrices are independent of time index t and we have used subscript k to stand for the harmonic component. The rationale behind $\mathbf{G}_{\text{seas}(k)}$ is that if we know the harmonic $u_{t,k} = a_k \cos(k2\pi/12t) + b_k \sin(k2\pi/12t)$ and, as an auxiliary state, its conjugate $u_{t,k}^* = -a_k \sin(k2\pi/12t) + b_k \cos(k2\pi/12t)$, with some constants a_k and b_k , we can update the state with

$$\begin{bmatrix} u_{t+1,k} \\ u_{t+1,k}^* \end{bmatrix} = \mathbf{G}_{\text{seas}(k)} \begin{bmatrix} u_{t,k} \\ u_{t,k}^* \end{bmatrix}, \quad (8)$$

where $\mathbf{G}_{\text{seas}(k)}$ is defined in Eq. (7) and does not depend on time t (e.g. Petris et al., 2009, Sect. 3.2.3). In our case the seasonality can be adequately explained by two harmonics, e.g. by yearly and half-a-year variation ($k = 1$ and $k = 2$), which will increase the number of hidden states to be estimated by four. In addition we need to define the error covariance matrix \mathbf{W}_{seas} for the allowed time-wise variability in the seasonal components. We note that there is no need for a separate seasonal observation error matrix \mathbf{V}_{seas} , as \mathbf{V}_t is used only in the observation equation (1), not in the model equation (2) where the seasonal cycle is defined.

In the previous definitions, the state operator \mathbf{G} , the observation operator \mathbf{F} and the model error covariance \mathbf{W} have been time invariant. The observation uncertainty covariance \mathbf{V}_t defined in Eq. (6) is, in our case, time dependent and it will contain the known observation uncertainties. The inclusion of auxiliary proxy variables is done by augmenting the observation matrix \mathbf{F}_t , making it time dependent. In the following, the stratospheric ozone analysis will utilize four proxy time series explaining parts of the natural variability: one for the solar flux, two proxy variables for QBO, and one for ENSO. This is achieved by adding the following components into the system matrices:

$$\mathbf{G}_{\text{proxy}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{F}_{\text{proxy}(t)} = [z_{1,t} \quad z_{2,t} \quad z_{3,t} \quad z_{4,t}] \text{ and}$$

$$\mathbf{W}_{\text{proxy}} = \begin{bmatrix} \sigma_{z1}^2 & 0 & 0 & 0 \\ 0 & \sigma_{z2}^2 & 0 & 0 \\ 0 & 0 & \sigma_{z3}^2 & 0 \\ 0 & 0 & 0 & \sigma_{z4}^2 \end{bmatrix}, \quad (9)$$

where $z_{1,t}$, $z_{2,t}$, $z_{3,t}$ and $z_{4,t}$ contain the values of the four proxy series at time t . With positive error variances $\sigma_{z_i}^2$ this is an extension of linear regression analysis into one with time-varying coefficients. In our analysis, for simplicity, we set all the elements of the model error covariance matrix $\mathbf{W}_{\text{proxy}}$ to zero to obtain time-invariant regression coefficients for the proxy variables.

To allow autocorrelation in the residuals we use a first-order autoregressive model (AR(1)). This is similar to the Cochrane–Orcutt correction in classical multiple regression (see e.g. Hamilton, 1994) used by Kyrölä et al. (2013). However, in the DLM approach we can estimate the autocorrelation coefficient and the extra variance term together with the other model parameters, not by a separate iteration, as needed, e.g. by the Cochrane–Orcutt method. A first-order autoregressive model for a state component η_t can be written as $\eta_t = \rho\eta_{t-1} + \epsilon_{\text{AR}}$, with $\epsilon_{\text{AR}} \sim N(0, \sigma_{\text{AR}}^2)$, where ρ is the AR coefficient and σ_{AR}^2 is usually called the innovation variance in classical time series literature. In DLM form, we simply have

$$\mathbf{G}_{\text{AR}} = [\rho], \mathbf{F}_{\text{AR}} = [1], \text{ and } \mathbf{W}_{\text{AR}} = [\sigma_{\text{AR}}^2], \quad (10)$$

and both ρ and σ_{AR}^2 can be estimated from the observations.

The next step in DLM model construction is the combination of the selected individual model components into larger model evolution and observation equations. For the model evolution matrix \mathbf{G} and the observation operator \mathbf{F}_t we have

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{\text{trend}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{\text{seas}(1)} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}_{\text{seas}(2)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_{\text{proxy}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_{\text{AR}} \end{bmatrix},$$

$$\mathbf{F}_t = [\mathbf{F}_{\text{trend}} \quad \mathbf{F}_{\text{seas}(1)} \quad \mathbf{F}_{\text{seas}(2)} \quad \mathbf{F}_{\text{proxy}(t)} \quad \mathbf{F}_{\text{AR}}], \quad (11)$$

In our analyses, the state vector \mathbf{x}_t has a total of 11 components:

$$\mathbf{x}_t = [\mu_t \quad \alpha_t \quad u_{t,1} \quad u_{t,1}^* \quad u_{t,2} \quad u_{t,2}^* \quad \beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4 \quad \eta_t]^T, \quad (12)$$

where μ_t is the local mean, α_t is the local trend, i.e. the local change of the mean, $u_{t,k}$ and $u_{t,k}^*$ are the states representing

the two seasonal harmonics, the β are the regression coefficients for the four proxy variables, and η_t is an extra state for the autoregressive component that “remembers” the value of the observation from the previous time step. In this study, the regression coefficients (β) are assumed not to depend on time t . The full model error covariance matrix \mathbf{W} is a diagonal matrix with the corresponding variances at the diagonal,

$$\text{diag}(\mathbf{W}) = \begin{bmatrix} 0 & \sigma_{\text{trend}}^2 & \sigma_{\text{seas}}^2 & \sigma_{\text{seas}}^2 & \sigma_{\text{seas}}^2 & \sigma_{\text{seas}}^2 \\ 0 & 0 & 0 & 0 & \sigma_{\text{AR}}^2 & 0 \end{bmatrix}, \quad (13)$$

where the local level variance and the proxy regression coefficient variances have been set to zero and the four seasonal variances set to be all equal, to correspond to the simplification assumed in the analyses. Lastly, as already stated, the observation error covariance matrix \mathbf{V}_t , will be a 1×1 matrix and equals the known observation uncertainty: $\mathbf{V}_t = \sigma_{\text{obs}(t)}^2$. Next, the analysis will proceed to the specification of the variance parameters and other parameters in the model formulation (e.g. the AR coefficient ρ above), and to the estimation of the model states by state space methods.

2.3 Model parameter estimation

We have two kinds of unknowns, the model state variables \mathbf{x}_t , one vector for each time t , and the auxiliary parameters that define the model error covariance matrix \mathbf{W} and the system evolution matrix \mathbf{G} . At first sight, this might seem to be a vastly underdetermined system, as we have several model parameters for each single observation. However, by the sequential nature of the equations, we can estimate the states by standard recursive Kalman filter formulas.

Implicitly assumed in the state space equations (1) and (2) is that the state at time t is statistically conditionally independent of the history given the previous state at time $t - 1$. When the model equation matrices are known, this Markov property allows sequential estimation of the states given the observations by famous Kalman formulas (see e.g. Rodgers, 2000). We can use a Kalman filter for one-step-ahead prediction of the state and a Kalman smoother for the marginal probability distribution of the state at time t given the whole time series of all observations ($t = 1, \dots, N$). Marginal means that the uncertainty of the states at all other times than t has been accounted for. In time series applications, the Kalman filter output can be used to calculate the model likelihood function, needed in the estimation of the auxiliary parameters, and the Kalman smoother provides an efficient algorithm to estimate the model states and to decompose time series into parts given by the model formulation. Thus, this high-dimensional problem is computationally not much more intensive than a classical static multiple linear regression analysis. Furthermore, in the linear Gaussian case, the probability distributions provided by the Kalman formulas are exact, not approximations. Non-linear models can be approached by linearization of the state equations, and

non-Gaussian error models by, e.g. particle filter algorithms (Doucet et al., 2001). More details on the DLM computations can be found in Appendix A.

Next, we consider the model error covariance matrix \mathbf{W} . If we set all model error variances to zero it will change the DLM model into an ordinary, non-dynamic, multiple linear regression model. By using non-zero variances we can fit a smoothly varying mean level, and the smoothness can be controlled by the size of the variances. A simplification done here is that we assume \mathbf{W} to be diagonal, and even some of the diagonal elements are set to zero. In our case non-zero elements are the variability of the trend, σ_{trend}^2 , and the variances of seasonal variation, $\sigma_{\text{seas}(k)}^2$, i.e. we will set the variance of the level and the four proxy variables to zero. Our motivation for excluding dynamic variation from the proxies is to use an as simple as possible dynamic model that would have similar properties as the piecewise linear model in the companion paper. By studying dynamic changes in the trend and seasonal variation we will then either validate the static linear approach or show that it is not appropriate.

A common procedure to estimate the elements of the model error matrix \mathbf{W} is based on the maximum likelihood method using the likelihood function provided by the Kalman filter. After the estimation, the obtained values could be plugged into the system equations as known constants. However, this plug-in method neglects the uncertainty in the estimates. Instead, we will use an alternative method based on Bayesian analysis (Gamerman, 2006; Petris et al., 2009; Särkkä, 2013) and outline it shortly below in Sect. 2.4.

2.4 Markov chain Monte Carlo analysis for model parameters

The Markov chain Monte Carlo (MCMC) method provides an algorithm to draw samples from the Bayesian posterior uncertainty distribution of model parameters given the likelihood function and the prior distribution (Gamerman, 2006). As the Kalman filter likelihood provides the likelihood of the auxiliary model parameters, marginalized over the unknown model states, we can use it to draw samples from the marginal posterior distribution of these parameters. We will use an adaptive Metropolis algorithm of Haario et al. (2006) for the three unknown variance parameters (σ_{trend} , σ_{seas} , σ_{AR}) in the matrix \mathbf{W} in Eq. (13) as well as for the autoregressive parameter ρ in the system evolution matrix \mathbf{G} (Eq. 10). After obtaining this sample from the posterior distribution, we use sampled parameter values, one by one, to simulate realizations of the model states $\mathbf{x}_{1:N}$ using the Kalman simulation smoother (e.g. Petris et al., 2009). This allows us to account for both the parameter and the model state uncertainties in the trend analysis. Again, more computational details are given in Appendix A.

For Bayesian analysis of the unknown model parameters, prior distributions for the parameters must be specified. We want the variances in the matrix \mathbf{W} to reflect our prior knowl-

edge on the assumed variability in the processes captured by the observations. As noted by Gamerman (2006, Sect. 2.5.), dynamic linear models offer intuitive means of providing qualitative prior information in the form of the model equations and quantitative information by prior distributions on variance parameters. By the estimation procedure we aim at finding variance parameters that are consistent with the given observation uncertainty, i.e. the model can predict the observations within their accuracy. This means that the scaled prediction residuals that are used defining the likelihood function should behave like an independent Gaussian random variable. We can assess these assumptions by different residual analysis diagnostics.

As we are effectively looking for slowly varying trends in the data, we will set prior constraints to variance parameters to reflect this. For example, we might assume that the change within a month in the background level is on the average some percentage of the overall time series mean. The estimation procedure will then divide the observed variability into model components (level, trend, seasonality) in proportions that reflect the prior choices. The standard model diagnostic tools, such as autocorrelation analysis and normal probability plots, can be used to reveal possible discrepancies in the model assumptions that have to be considered. As the model residuals are calculated from one-step predictions, the diagnostics will reveal both over-fit and a lack of fit.

The model error covariance \mathbf{W} (Eq. 13) is time invariant, with nonzero diagonal terms for the trend parameter, a common value for the four parameters defining the variability in the seasonal components, and one value for the variance defining the autoregressive component. For MCMC estimation of the variance parameters, we use log-normal prior distributions for the corresponding standard deviations. The motivation for this is that the standard deviations are positive by definition so logarithmic scale is natural and commonly used and it allows prior specification in meaningful units. If a random variable x follows a log-normal distribution $\log N(\mu, \sigma^2)$, then $\log(x)$ follows the standard Gaussian distribution $N(\mu, \sigma^2)$. Because of this transformation, it is more intuitive to work with a log mean parameter, as in $\log N(\log(\mu), \sigma^2)$, where μ is now the approximate mean of the untransformed x and the scale parameter of the log-normal distribution σ can be interpreted as an approximation of the relative standard deviation of the original variable in question. The prior distribution for the standard deviation of the monthly level change σ_{trend} was set to be log-normal with the log mean equal to 1/12 % of the average level of the ozone observations and having scale parameter 1. This scale corresponds to a relatively wide (130 % SD)² prior uncertainty for σ_{trend} . All the σ parameters describe allowed

²This can be derived from the fact that if $x \sim \log N(\mu, \sigma^2)$ then its mean and variance are $E(x) = \exp(\mu + 1/2\sigma^2)$ and $D(x) = \exp(2\mu)\exp(\sigma^2)(\exp(\sigma^2) - 1)$, see e.g. Wikipedia entry for the log-normal distribution.

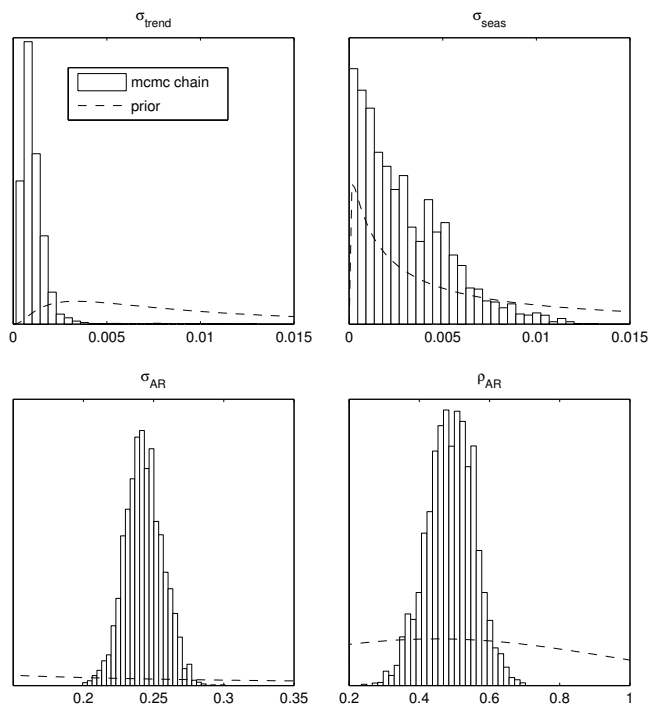


Figure 1. Results of MCMC analysis for the four unknown model parameters: trend change standard deviation σ_{trend} , seasonal change standard deviation σ_{seas} , autoregressive standard deviation σ_{AR} and the autoregressive coefficient ρ . The histograms are estimates of the corresponding marginal posterior probability densities obtained by MCMC analysis with 10 000 simulation steps. The parameters are for 40–50° N, 35–45 km as in Fig. 2. The prior probability densities given by dashed lines are log-normal for the standard deviations and truncated Gaussian for the AR parameter ρ . For the analysis the ozone observations have been scaled by their observed standard deviation. This means that the values given on the x axes are also scaled so that the overall variance of the ozone observations is one.

change in one time interval of the model and data (month), i.e. the unit for σ_{trend} is 1 cm^{-3} . For the seasonal standard deviation, σ_{seas} , we set the log mean parameter to 1 % of the overall variability in the observed ozone values and the scale parameter to 2 (corresponding to 730 % SD). For the standard deviation of the AR(1) component (Eq. 10), we use 30 % of the observed variability for the log mean and similarly 2 for the scale. For the autoregressive coefficient ρ the prior was Gaussian $N(0.45, 0.5)$ truncated to be in the interval $[0, 1]$. Table 1 collects the prior specifications. The results for the trend analyses or the estimated parameter values themselves are not sensitive to the chosen priors. The MCMC for each separate fit was run using 10 000 simulation steps, and the resulted parameter samples (chains in MCMC terminology) were checked for convergence of the algorithm. The histograms of the marginal posterior samples of the parameters with the prior probability density functions overlaid are shown in Fig. 1 for one zonal band and altitude region. The prior for the trend variability would allow the trend to change

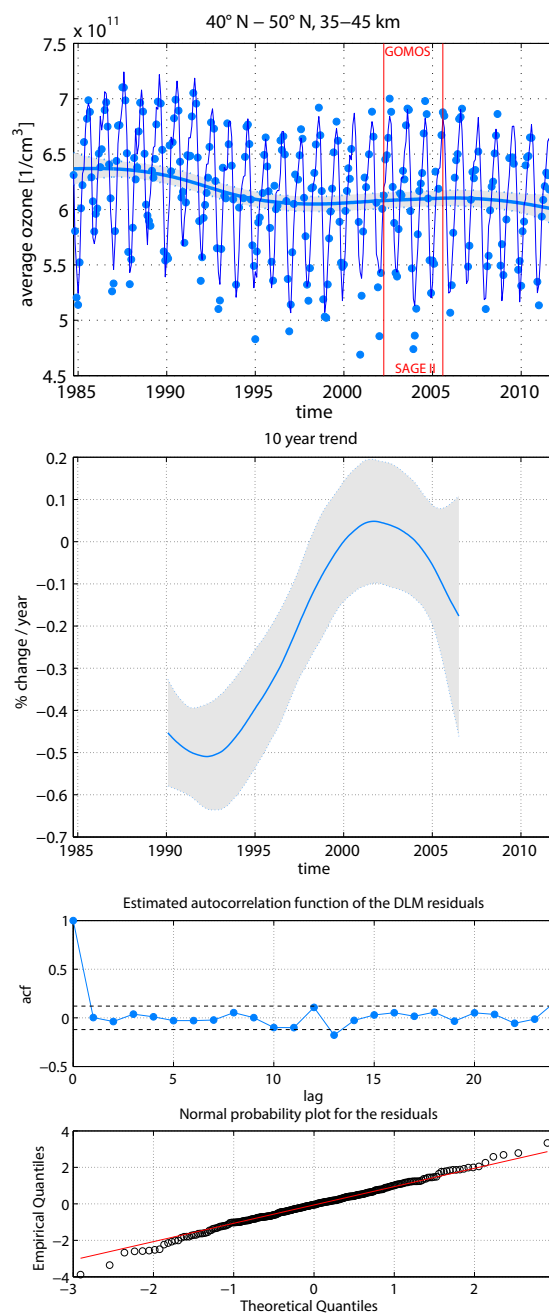


Figure 2. DLM fit for average ozone at 40–50° N, 35–45 km. In the upper panel the dots are the observations used in the analysis, the solid line following the observations is the DLM fit obtained by a Kalman filter. The smooth solid line is the background level component of the model with 95 % probability envelope. The beginning of GOMOS observations as well as the end of SAGE II are shown by red vertical lines. In the lower left panel an analysis of the 10-year trend is performed, showing the mean estimated trend and its 95 % probability envelope. The lower right panel shows model diagnostic analyses on the residuals by an estimated autocorrelation function (ACF) and by a normal probability plot. In the autocorrelation function plot, dashed horizontal lines correspond to the approximate region where the coefficients do not significantly (95 %) deviate from zero. The unit of lag is months.

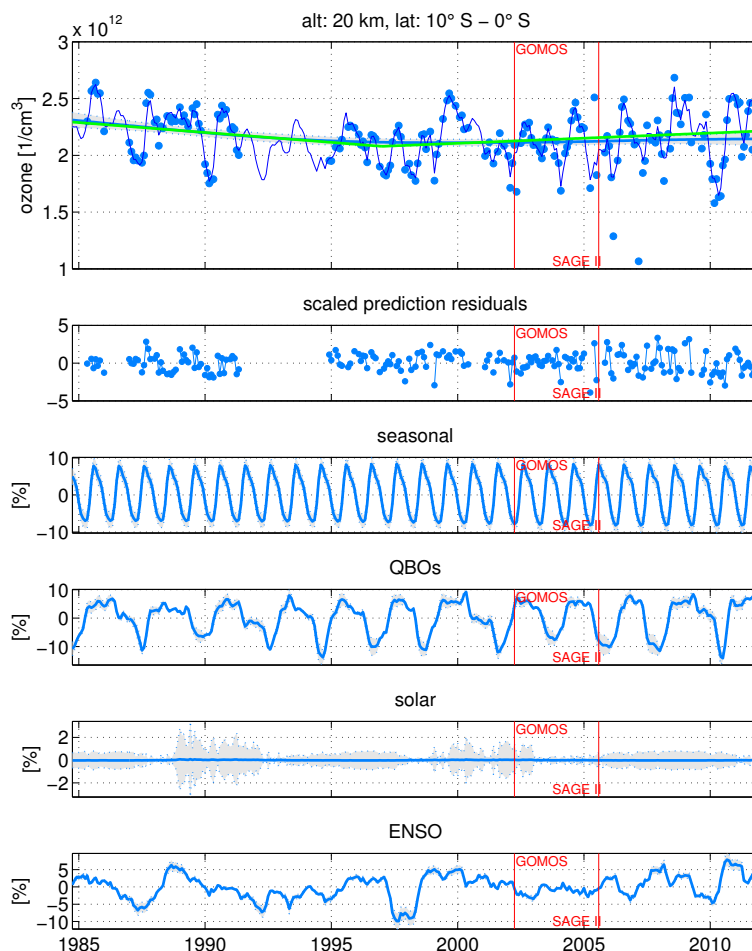


Figure 3. Some of the DLM model components for 10–0° S, 20 km. The top panel has the observations and the level components as in Fig. 2. In addition, the piecewise linear regression model used by Kyrölä et al. (2013) is drawn with a solid green line. The second panel shows model residuals, scaled by the estimated residual standard deviation, i.e. the variability in the y axis direction should be approximately standard Gaussian $N(0, 1)$. The other panels show means and 95 % probability envelopes (grey shading) of the DLM components related to seasonality and to the proxy variables – solar, QBO and ENSO – i.e. the original proxy variables multiplied by the estimated regression coefficients. For the proxy variables, the y axis scaling is the percentage of the total ozone variability, i.e. the ozone observations were scaled by their standard deviation, individually for each data set analysed. At 20 km we see a large number of missing observations, some possible outliers, and a strong effect of the ENSO index. There is only a slight difference between DLM and linear models, however, although the latter does not include ENSO.

by about same amount as it does in linear trend analyses – or let it not change if the data do not support it – but, also, should not capture too fine changes due to natural variability that could be attributed, e.g. to the autoregressive residual error component in the model.

The simple parameterization of the model error term with three unknown parameters was selected by performing initial fits with different parameterizations, and using sensible initial values refined by a maximum likelihood optimization. The models were then diagnosed by studying the residuals by using normal probability plots and plots of the estimated autocorrelation function. When a good candidate model was found, an MCMC analysis was used to study the uncertainty and identifiability of the variance parameters. Lastly, the in-

teresting trend features of the time series were studied by plotting the estimated background level with uncertainty confidence bounds, and drawing realizations from the posterior distribution of the level term and checking the statistical significance of hypothesized features. For example, statistical significance for the trend was assessed by checking whether 95 % posterior probability region included zero value or not.

2.5 Estimating trends

A trend is a change in the statistical properties of background state of the system (Chandler and Scott, 2011), the simplest case being linear trend, where, when applicable, we only need to specify the trend coefficient and its uncertainty. In the

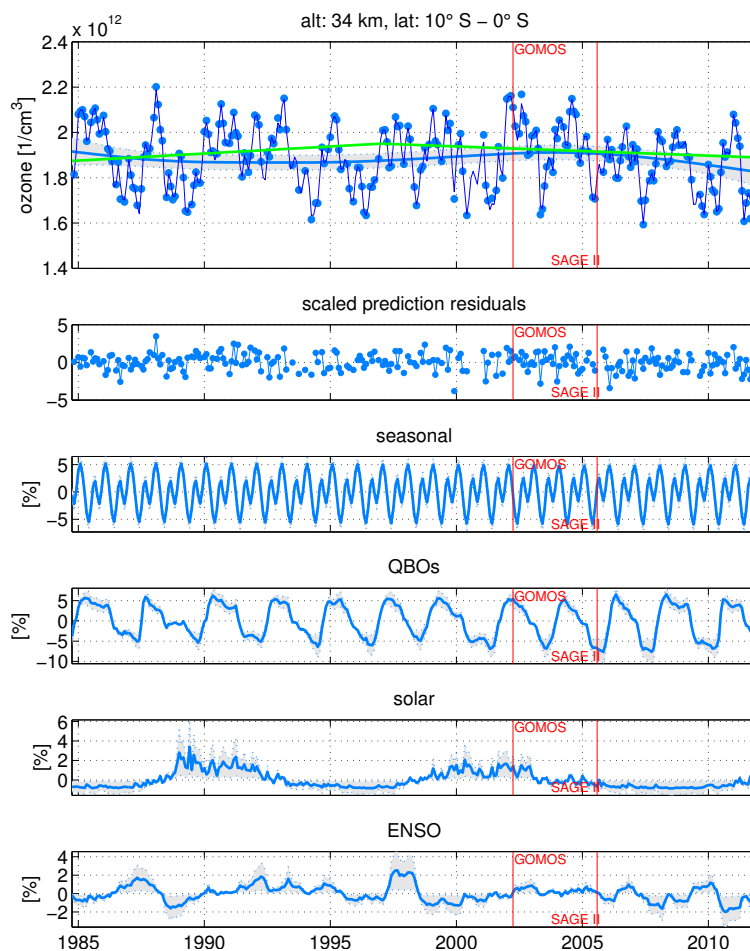


Figure 4. Some of the DLM model components for 10–0° S, 34 km, see Fig. 3 for explanation. We see noticeable difference between DLM and linear regression fits drawn with a solid green line on the top panel.

Table 1. Specification of priors for auxiliary model parameters estimated by MCMC. The prior distributions for the model error standard deviation parameters were log-normal, $\log N(\log(\mu), \sigma^2)$ with values of μ and σ given in the table. For simplicity, the same relative priors are used in all the altitude–latitude regions. In Fig. 1 the prior distributions are plotted together with MCMC chain histograms that estimate the posterior distributions for one altitude–latitude region.

Estimated parameter	Prior mean μ	Prior scale σ
Trend standard deviation, σ_{trend}	1/12 % of O ₃ mean	1
Seasonal standard deviation, σ_{seas}	1 % of O ₃ variability	2
Autoregressive standard deviation, σ_{AR}	30 % of O ₃ variability	2
Autoregressive coefficient, ρ	prior is Gaussian $N(0.45, 0.5^2)$ truncated to [0, 1]	

companion paper (Kyrölä et al., 2013) the trend and a change in the trend is studied by using a piecewise linear model with a predefined change point. Natural systems evolve continuously in time and it is not always appropriate to approximate the background evolution with a constant, or piecewise linear, trend. Within the state space dynamic linear model framework, the trend can be defined as the change in the estimated background level, i.e. the change in μ_t defined in

Eq. (3). Posterior sampling from the background level provides an efficient method for studying uncertainties in different trend estimates.

Temporal changes in the system can be studied by visually inspecting the background level and its estimated uncertainty. We can draw samples from the posterior distribution of the level μ_t to assess hypotheses about the evolution of the process. For example, in Sect. 3 we study the change in

the mean ozone level in 10-year periods. We take into account the uncertainty in the model prediction and in the estimated variance parameters by sampling possible background levels μ_t from its posterior uncertainty distribution. We will do this consecutively and, for each sampled realization, calculate the 10-year change in the mean ozone for each time t as $\text{trend}(t) = \mu_{t+60} - \mu_{t-60}$ (time units in months). This sample of trends provides us direct way to analyse trends by calculating, for example, the mean 10-year trend with 95 % uncertainty limits. The general procedure is the following:

1. Using MCMC with the Kalman filter likelihood, produce a sample from the marginal posterior probability distribution of the auxiliary parameters defining the error covariance matrix \mathbf{W} and model matrix \mathbf{G} .
2. Draw one realization of the matrices \mathbf{G} and \mathbf{W} using the posterior distribution provided by MCMC in the previous step.
3. Simulate one realization of the model state $\mathbf{x}_{1:N}$ using the Kalman simulation smoother assuming fixed \mathbf{G} and \mathbf{W} from the previous step and calculate trend-related statistics of interest from this realization.
4. Repeat from step 2 to calculate summaries from the posterior distribution of the quantity of interest.

3 Results

The model parameters were fitted separately to each data set, i.e. to each height interval and zonal band. We performed the analyses using vertical average profile data with both the original interpolated 1 km altitude grid and by forming averaged ozone densities for three altitude regions: 25–35, 35–45, and 45–55 km. The 10° wide zonal bands start from 60° S and go to 60° N. By considering each zonal band independently and summing several altitudes, we have tried to reduce the model to a minimum one that still shows interesting long-term changes and is consistent with our assumptions. Initially, a multivariate estimation was considered by fitting several altitudes and zonal bands together, but this complicated the analyses considerably and did not gain additional insight. In principle, we could use the observations in more refined resolution and model several time series in one estimation step, and even use the individual satellite retrievals instead of spatio-temporal averages.

Figure 2 shows an example of our modelling results in the combined altitude–latitude region, 40–50° N, 35–45 km. The original data are displayed together with DLM estimates and with the estimated mean level component μ_t that is used to make statistical inference about the trend. The fits are obtained by a Kalman smoother (Eqs. A1–A7 in Appendix A) and the 95 % uncertainty regions by combination of a simulation smoother (Eqs. A8–A12) and MCMC. A separate panel (on the lower left side) displays the decadal trend obtained

from the level component using MCMC analysis to account for the uncertainty in the variance parameters. Looking at the observations in a 10-year perspective, the trend has been statistically significantly negative up to the year 1997, as the grey area stays below zero. After 1997, the 10-year trend does not statistically differ from zero. After the DLM decomposition, the model residual term is assumed to be uncorrelated Gaussian noise. The two lower right panels in Fig. 2 show residual diagnostics. These are used to look for deviations from the modelling assumptions. If the scaled residuals pass the check of being independent and Gaussian with unit variance, then the fit agrees with our assumptions and modelling results are consistent with the data. In this case the residuals are almost perfectly uncorrelated Gaussian with unit variance. The normal probability plot compares empirical quantiles (i.e. order statistics) of the residual to those obtained from the theoretical Gaussian distribution and in the case of normally distributed residuals, the points should stay mostly on a straight line, as they do here.

The results from MCMC analysis for the same data set as in Fig. 2, are shown in Fig. 1. The figure has the prior probability distributions together with marginal posterior distributions as MCMC chain histograms for the estimated auxiliary model parameters (referred as θ in Appendix A). We see that the parameter posterior distributions are mostly determined by the data and not by the prior. The standard deviation of the trend, σ_{trend} , is estimated to be relatively small, which supports the search for smooth background variability. This behaviour is typical for all the fits done. The variability in the seasonal component has more uncertainty, and it also varies more between the fits for different altitude–latitude regions. The autoregressive parameters σ_{AR} and ρ have narrow posterior distributions relative to the prior, i.e. these values can be accurately obtained from the data.

Figures 3 and 4 show the fitted model components for two individual altitudes, 20 and 34 km, both at 10–0° S. For the seasonal cycle, we plot the sum of the observed seasonal components, $u_{t,1} + u_{t,2}$ in Eq. (12); for proxies we plot $\beta_i z_{t,i}$, i.e. the proxy coefficient times the value of the proxy. These fits are provided by the Kalman smoother as in the case of the mean level μ_t . At 20 km, we see a strong effect of ENSO. The total variability explained by ENSO is about 20 % (i.e. its effect is between –10 and 10 %), which is similar to the effects of the seasonal variation and of the QBO. Although ENSO was not used by Kyrölä et al. (2013), the fit by linear regression, shown with a solid green line in the upper panel, does agree reasonably well with the DLM level. At 34 km, there is a clear difference between the DLM fit and the piecewise linear analysis with change point fixed at year 1997. Non-linear changes in the background level are not captured by the two-piece linear model. Among the proxy variables, the two QBOs, combined in the figure, have the most significant effect, explaining approximately 10 % of the total variability in ozone at that region.

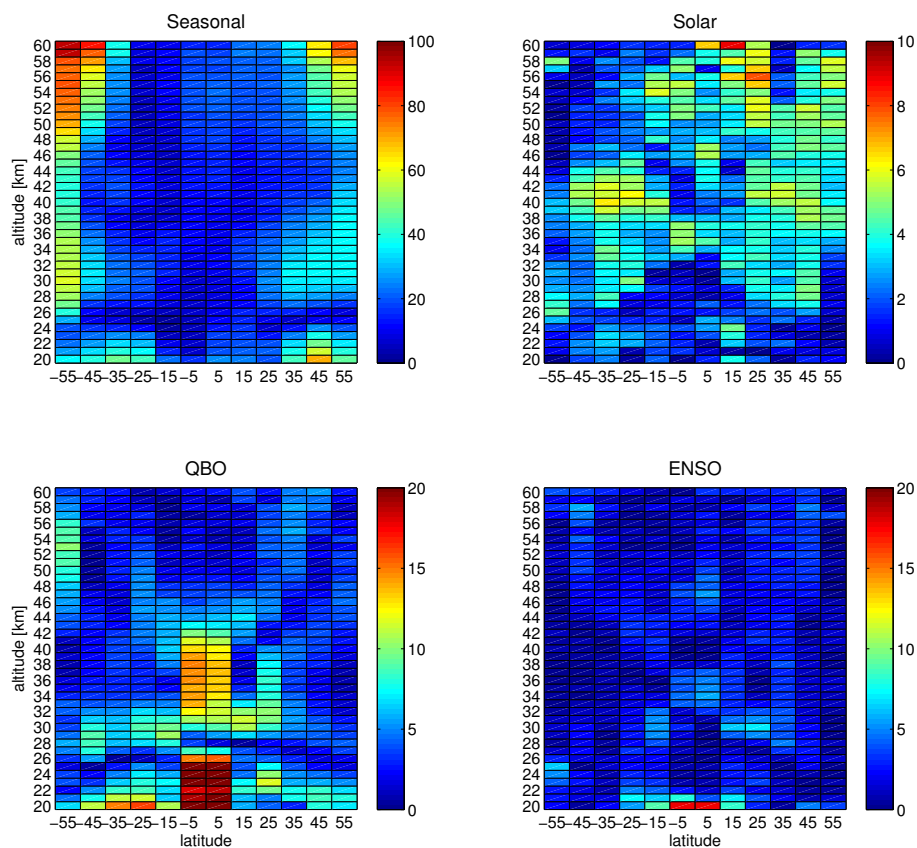


Figure 5. The effects of seasonal variation and of the proxy index variables on the variability of ozone. Here 100% means the observed standard deviation of ozone at each latitude altitude bin. The colouring corresponds to the range of variability during the whole observation period 1984–2011 of each model component given as percentages with respect to the total variability in ozone. We see, e.g. that the ENSO index significantly affects the ozone variability only at 20 km near the equator, where the effect is about 20 %, also shown in the lowest panel of Fig. 3. The maximal solar effect is always less than 10 %. Note that the colour scales differ.

Figure 5 shows the effects of the fitted variable for the seasonal variation and of the proxy index variables on the total observed variability of ozone. We calculated, using the data with 1 km altitude spacing, the range of variability during the whole observation period 1984–2011 of each model component given as percentages with respect to the mean ozone value. The ENSO index has a significant effect (about 20 % of the variability) only at 20 km near the equator. Near 60 km and at higher latitudes, almost all of the ozone variability can be attributed to seasonality. The maximal solar effect is always less than 10 %.

The key results of this paper can be found in Figs. 6 and 7 where we show fits to the ozone time series for 12 latitude belts and for three altitude regions (25–35 km, 35–45 and 45–55 km). We have shown the data points from the merged SAGE II–GOMOS time series, the DLM fit and the slowly varying background level μ_t . Overall, it is easy to see that the fits usually follow very accurately the data points. There are a few scattered outliers, but it is difficult to find any specific pattern for these deviations. As in Fig. 2, the solid trend line is the estimated mean level μ_t and it is shown with 95 %

uncertainty region. The results for the background curve can be roughly divided into three classes. The simplest case, a continuous decay of ozone during the period 1984–2011, is evident in the southernmost latitude band (50–60° S) in all altitude regions. A continuous decay is also present in 50–60° N in the lowest altitude region (25–35 km). The second class, a recovery between an initial decay and a final decay, is seen from the equator to 50° N in the lowest altitude region (25–35 km) and at 40–60° N and 35–45 km. The third class covering the rest of the cases shows decay and recovery. In the Southern Hemisphere all altitudes and latitudes (10–50° S) belong to this class. In the Northern Hemisphere all latitudes in 45–55 km and latitudes from the equator to 40° N belong to this class.

In Fig. 8 the results of decadal trend analyses are collected and plotted separately for the three altitude ranges. The calculations are explained in Sect. 2.5 and as final step 7 in Appendix A. The colouring shows the trend, i.e. the average change in ± 5 years around each time axis point. Blue colour means a negative and red a positive trend. The change is most visible at higher altitudes (45–55 and 35–45 km). At the

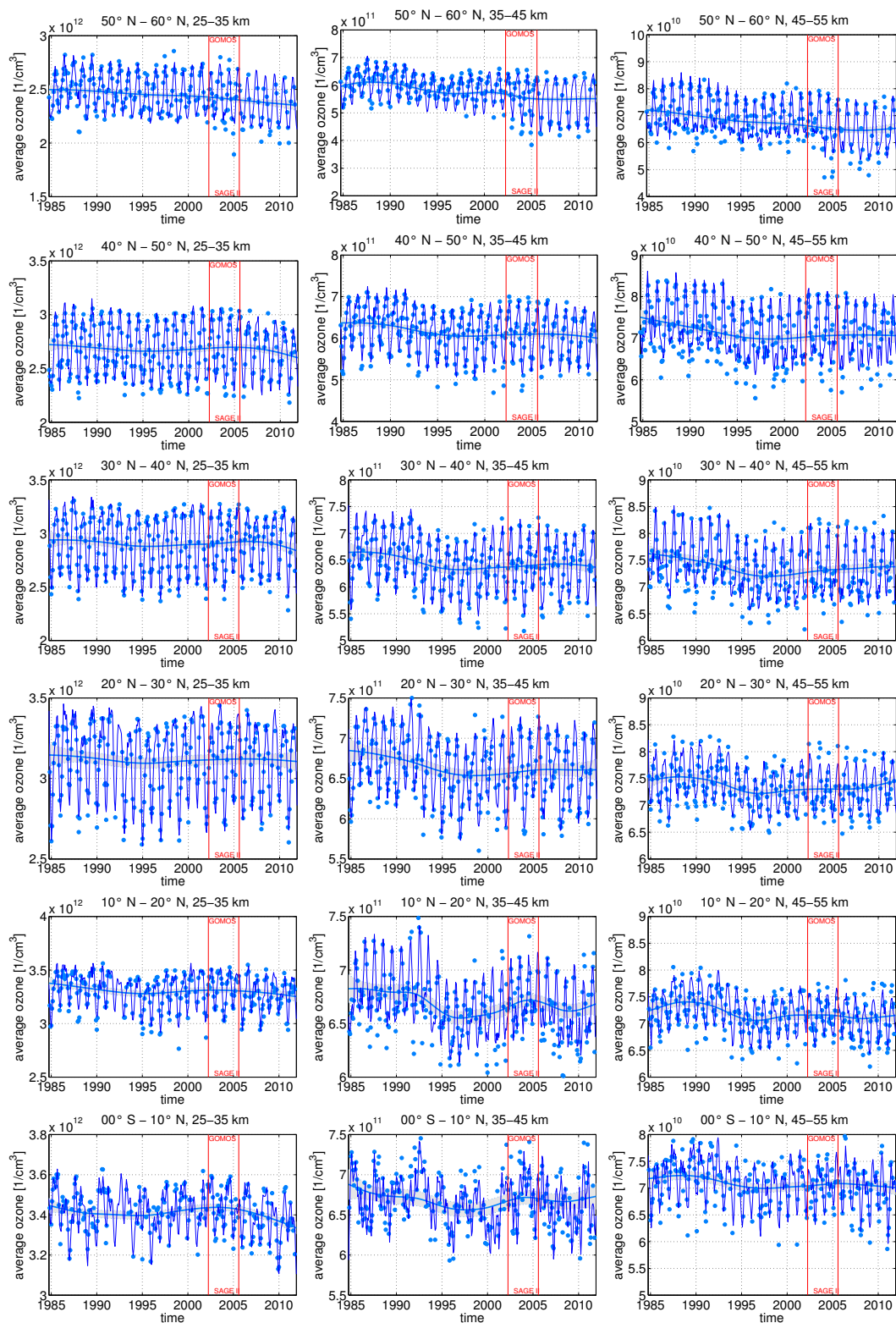


Figure 6. All fits collected for the Northern Hemisphere. The three altitude regions as columns, Northern Hemisphere 10° zonal bands as rows starting from the northernmost zone. The smooth solid curve shows the estimated background level with 95 % probability envelope. The dots are the observations used in the analysis. The solid line following the observations is the DLM fit obtained by Kalman filter formulas.

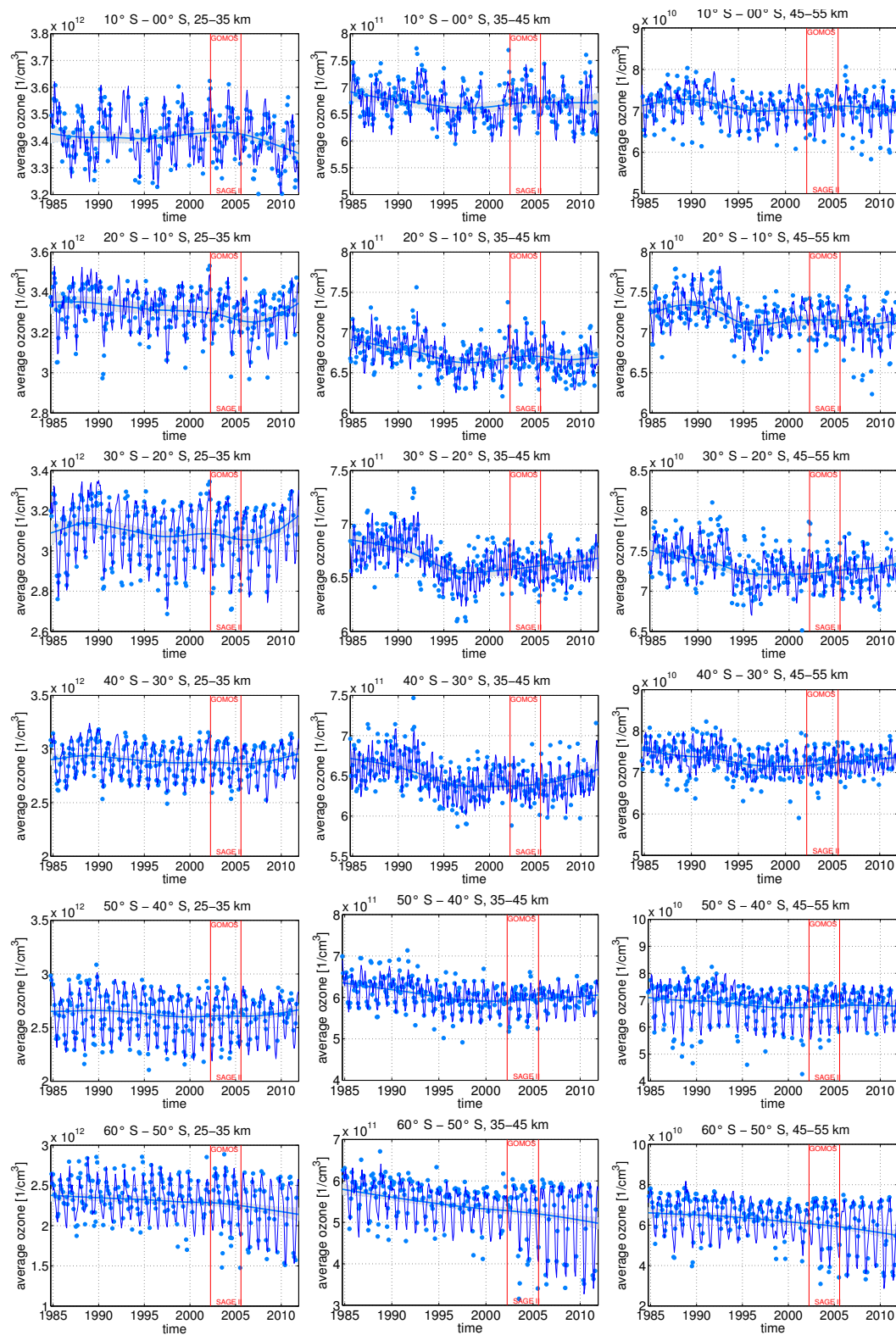


Figure 7. All fits collected for the Southern Hemisphere. The three altitude regions as columns, Southern Hemisphere 10° zonal bands as rows starting from the equatorial zone. See Fig. 6 for explanation.

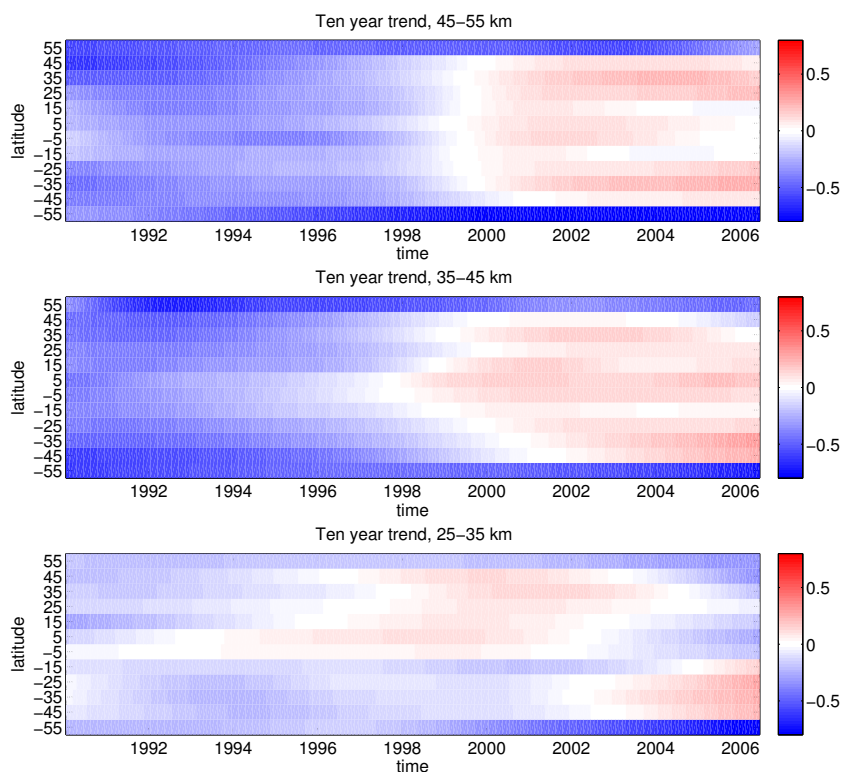


Figure 8. The 10-year trend in percentage change / year from the DLM fits for each altitude region and zonal band. One individual trend analysis is shown in the lower left panel of Fig. 2.

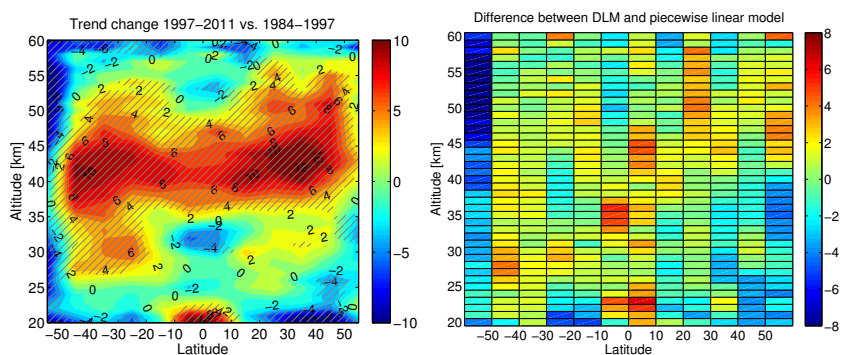


Figure 9. To compare DLM results to linear regression analysis in Kyrölä et al. (2013) we reproduce its Figure 16 by using the DLM approach. The left panel shows the change in ozone trends in % per decade between the periods 1997–2011 and 1984–1997. The shaded area indicate regions where the trend difference is nonzero with over 95 % posterior probability. The right panel shows the differences (in $\% \text{ decade}^{-1}$) between the numbers that produce the contour plot on the left panel and the corresponding one in Kyrölä et al. (2013) (Fig. 16).

lower altitudes (25–35 km) there are some visible changes, but they are mostly masked by a larger variability and the noise level of the satellite observations at these altitudes. The northernmost (50–60° N) and the southernmost (60–50° S) zones differ significantly from the other zones. Looking at the individual plots (Figs. 6 and 7), we see some evidence of instrument-related differences and one possible explanation comes from the different non-uniform sampling character-

istics, both temporal and spatial, of SAGE II and GOMOS (Toohey et al., 2013).

One particular feature of interest in the data is the suggested stratospheric ozone recovery due to the prohibition of CFC compounds by the Montreal treaty in 1987. Several studies indicate a possible turning point around the year 1997 (Newchurch et al., 2003; Jones et al., 2009; Steinbrecht et al., 2009). According to our analyses, significant changes in the average background level of ozone can be seen at

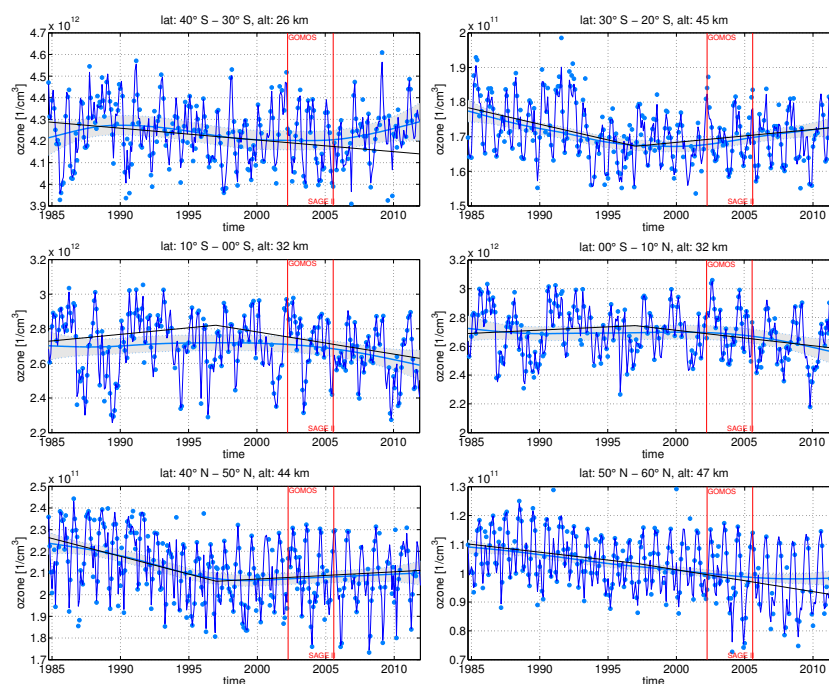


Figure 10. Comparing DLM and the linear model. Six examples of DLM vs. piecewise linear model fits using data with 1 km altitude resolution. The solid black line is the piecewise linear trend from Kyrölä et al. (2013). The smooth solid blue line is the background level component of the DLM model with 95 % probability envelope. At 40–50° N and 44 km the methods agree. In the other panels we see differences of various degrees.

mid-latitude regions (40–20° S and 20–40° N) and in the altitude range 35–55 km. In these regions the trend, measured as average change in 10 years, has mostly turned from negative to positive before the year 2000. Further, in those geographical regions where the “V” type shape of the change is appropriate, the turning point is around the years 1997–2001. The change points and trend features have variations by latitude and altitude. In the Tropics one can see that the change of trend year was much earlier than at mid-latitudes. This suggests that a model with fixed change point in all regions, as used by Kyrölä et al. (2013), does not capture all features in the trends. In some regions there are signs of reduction of ozone after the year 2007. Near the equator the long-term background changes are mostly masked by irregular variations in the ozone concentration.

3.1 Comparison with piecewise linear analysis

Kyrölä et al. (2013) analysed the trend change point by calculating several fixed piecewise linear regression analyses and choosing the one that provided largest difference between before and after the point change in the linear trend. In the DLM approach the trend can change its value continuously and we can analyse directly where the most likely change points are. One use of a DLM model would be the identification of possible trend change points, and provide hypotheses on which a traditional static regression analysis would be

performed. The results of the piecewise linear trend analysis in Kyrölä et al. (2013) are in agreement to some extent with the results here, but the DLM analysis suggests that the trend change point depends on latitude and altitude, as shown in Fig. 8. Also, there is some evidence against using a simple two-piece linear model to describe the background changes. To further compare the methods, we studied the difference in trend between the years from 1984 to 1997 and from 1997 to 2011 using the DLM method. Figure 16 of Kyrölä et al. (2013) corresponds to our Fig. 9 (left panel). Overall, the DLM method produces somewhat larger differences (i.e. faster ozone recovery) in before- and after-1997 trends. This difference varies and is at most 3 % / 10 years. An opposite effect is seen in the southern most latitude band (50–60° S), where the DLM method suggests stronger decline of ozone than the linear method. The uncertainty (as a standard deviation of the estimated change) in DLM trend difference is about 1.5 % / 10 year on the average. For the piecewise linear model it is approximately 0.5 % larger. Because of this relatively large uncertainty, the differences between DLM and piecewise linear trend analysis are mostly not statistically significant. The right panel in Fig. 9 shows the difference in the results between the linear and the DLM approach. The analysis for the before-and-after-1997 trend differs mostly at the lowest and highest latitudes and at the equator. At the southernmost zone the difference is the largest and can be attributed to the ability of the DLM method to dynamically

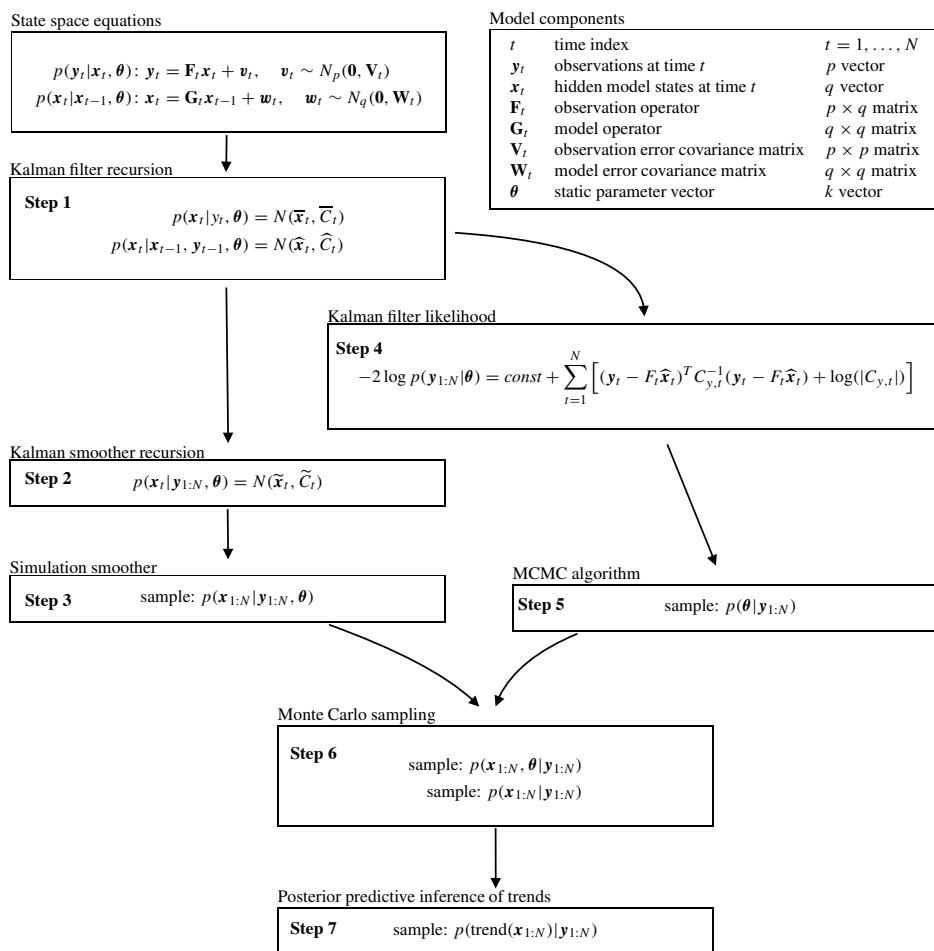


Figure 11. A flowchart showing the dependencies of the DLM calculations explained in Appendix A.

adapt to changing seasonal patterns in the observations, see Fig. 7. However, this might be caused by the different spatio-temporal sampling of the two instruments. Further analysis is beyond the scope of this paper – see the discussion in the companion paper (Kyrölä et al., 2013).

Figure 10 shows six examples of the DLM approach vs. the piecewise linear model fits from Kyrölä et al. (2013). In five of the cases the conclusions disagree in some respect. In these cases the DLM trend has more variability than the fixed-point model trend – the linear model may miss some important features in the data, and the rigidness of the piecewise linear model may cause spurious results. In some cases the change point is later than at the beginning of the year 1997. However, at those regions where the one-change-point model is valid, the conclusions are the same with both modelling approaches.

4 Conclusions

We have shown that a dynamic linear model (DLM) is well suited for modelling ozone time series. In contrast to some classical time series analyses, DLM does not require stationarity, it allows for missing observations and takes uncertainties in the observations into account. By using Markov chain Monte Carlo (MCMC) simulation analysis, the uncertainty in the structural variance parameters can be accounted for. The state space method directly includes a model error term, which makes the analysis more robust to mis-specification of the model. The analysis allows full statistical uncertainty quantification, and it is extendible to more refined analyses, if those seem necessary. Here, a relatively straightforward and conceptually simple analysis reveals interesting features and, also, validates the more ad hoc choices used in piecewise linear global analyses, such as in the companion paper by Kyrölä et al. (2013).

Trend analysis can be a delicate matter and it is always challenging to give causal explanations. With a properly set up DLM model we can detect smooth changes in the

background state. By using proxy variables we can filter out the effect of known external forcing, such as the solar effect. The DLM analysis provides a method to detect and quantify trends, but the statistical model itself does not provide explanations. It can verify that the observations are consistent with the selected model. Model diagnostics will eventually falsify wrong models and other badly selected prior specifications.

We used a local linear trend model with two harmonic functions for the seasonal effect and four proxy time series for the solar flux, Quasi-Biennial Oscillations and ENSO. An autoregressive model component was used to account for possible residual autocorrelation. The results show a statistically significant change point in the combined SAGE II–GOMOS time series approximately after year 1997 at altitudes 35–55 km and mid-latitudes, between 50–20° S and 20–50° N. This change point in time varies with latitude and altitude. There are locations where the estimated changes are opposite to those expected and the length of the data set is still short relative to some cycles of natural variability in the atmospheric processes. At the southernmost zones analysed here (60–50° S) the DLM method shows larger decline of ozone. This might be attributed to the sampling differences of the two instruments, which is better handled by the DLM method. In many regions the behaviour of the mean ozone concentration is too complicated to allow for a simple piecewise linear description. Compared to the two-piece fixed change point linear model used in the companion paper, we see stronger recovery of ozone in many regions – the difference is up to 3 % / decade.

Appendix A: Statistical computations for a DLM model

As DLM calculations are based on standard Kalman filter algorithms, they can be programmed by most numerical analysis software such as Matlab or the statistical language R. Some additional effort is needed for the parameter estimation by MCMC. We have used Matlab and checked the results with the R package DLM (Petris et al., 2009). The Matlab code we used for the DLM analyses, including the MCMC part, is available from <http://helios.fmi.fi/~lainema/dlm/>. We will briefly describe the computations here. A flowchart of the steps necessary in the computations is given in Fig. 11. A concise account of the state space approach and Kalman filter based estimation in time series trend analysis can be found in Sect. 5.5 of Chandler and Scott (2011) – for even more details we refer to already cited literature (Harvey, 1990; Petris et al., 2009; Durbin and Koopman, 2012).

We start with the state space equations (1–2), system matrices \mathbf{F}_t , \mathbf{G}_t , \mathbf{V}_t , \mathbf{W}_t , and the model state \mathbf{x}_t according to definitions in Sect. 2.2. We follow the Bayesian statistical approach, where we are interested in the joint posterior uncertainty distribution of the model states $\mathbf{x}_{1:N}$ and the unknown parameters defining the system matrices, given the observations $\mathbf{y}_{1:N}$ for the whole time series. We collect the extra static parameters (i.e. unknowns in addition to the state $\mathbf{x}_{1:N}$) into vector $\boldsymbol{\theta}$ and write this probability distribution as $p(\mathbf{x}_{1:N}, \boldsymbol{\theta} | \mathbf{y}_{1:N})$. In the present case, this parameter vector is four-dimensional: $\boldsymbol{\theta} = [\sigma_{\text{trend}}, \sigma_{\text{seas}}, \sigma_{\text{AR}}, \rho]^T$.

Step 1

(The steps refer to the flowchart in Fig. 11.) If we assume that the initial distributions at time $t = 0$ are available, then the Kalman filter forward recursion (see e.g. Rodgers, 2000) can be used to calculate the distribution of the state vector \mathbf{x}_t given the observations up to time t , $p(\mathbf{x}_t | \mathbf{y}_{1:t}, \boldsymbol{\theta}) = N(\bar{\mathbf{x}}_t, \bar{\mathbf{C}}_t)$, which is Gaussian for each $t = 1, 2, \dots, N$. This consists of first calculating, as prior, the mean and covariance matrix of one-step-ahead predicted states $p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}, \boldsymbol{\theta}) = N(\hat{\mathbf{x}}_t, \hat{\mathbf{C}}_t)$ and the covariance matrix of the predicted observations $\mathbf{C}_{y,t}$ as

$$\hat{\mathbf{x}}_t = \mathbf{G}_t \bar{\mathbf{x}}_{t-1} \quad \text{prior mean for } \mathbf{x}_t, \quad (\text{A1})$$

$$\hat{\mathbf{C}}_t = \mathbf{G}_t \bar{\mathbf{C}}_{t-1} \mathbf{G}_t^T + \mathbf{W}_t \quad \text{prior covariance for } \mathbf{x}_t, \quad (\text{A2})$$

$$\mathbf{C}_{y,t} = \mathbf{F}_t \hat{\mathbf{C}}_t \mathbf{F}_t^T + \mathbf{V}_t \quad \text{covariance for predicting } y_t \quad (\text{A3})$$

and then the posterior state and its covariance using the Kalman gain matrix \mathbf{K}_t as

$$\mathbf{K}_t = \hat{\mathbf{C}}_t \mathbf{F}_t^T \mathbf{C}_{y,t}^{-1} \quad \text{Kalman gain}, \quad (\text{A4})$$

$$\mathbf{v}_t = \mathbf{y}_t - \mathbf{F}_t \hat{\mathbf{x}}_t \quad \text{prediction residual}, \quad (\text{A5})$$

$$\bar{\mathbf{x}}_t = \hat{\mathbf{x}}_t + \mathbf{K}_t \mathbf{v}_t \quad \text{posterior mean for } \mathbf{x}_t, \quad (\text{A6})$$

$$\bar{\mathbf{C}}_t = \hat{\mathbf{C}}_t - \mathbf{K}_t \mathbf{F}_t \hat{\mathbf{C}}_t \quad \text{posterior covariance for } \mathbf{x}_t. \quad (\text{A7})$$

These equations are iterated for $t = 1, \dots, N$. As initial values, we can use $\bar{\mathbf{x}}_0 = \mathbf{0}$ and $\bar{\mathbf{C}}_0 = \kappa \mathbf{I}$, i.e. a vector of zeros

and a diagonal matrix with some large value κ in the diagonal. Note that the only matrix inversion required in the above formulas is the one related to the observation prediction covariance matrix $\mathbf{C}_{y,t}$, which is of size 1×1 when we analyse univariate time series.

Step 2

The Kalman filter provides distributions of the states at each time t given the observations up to the current time. As we want to do analysis that accounts for all of the observations, we need to have the distributions of the states for each time, given all the observations $\mathbf{y}_{1:N}$. By the linearity of the model, these distributions are again Gaussian, $p(\mathbf{x}_t | \mathbf{y}_{1:N}, \boldsymbol{\theta}) = N(\tilde{\mathbf{x}}_t, \tilde{\mathbf{C}}_t)$. Using the matrices generated by the Kalman forward recursion, the Kalman smoother backward recursion will give us these so-called smoothed states for $t = N, N-1, \dots, 1$ by using the following equations and setting r_{N+1} and N_{N+1} to zero:

$$\mathbf{L}_t = \mathbf{G}_t - \mathbf{G}_t \mathbf{K}_t \mathbf{F}_t \quad \text{auxiliary variable} \quad (\text{A8})$$

$$\mathbf{r}_t = \mathbf{F}_t^T \mathbf{C}_{y,t}^{-1} \mathbf{v}_t + \mathbf{L}_t^T \mathbf{r}_{t+1} \quad \text{--}, \quad (\text{A9})$$

$$\mathbf{N}_t = \mathbf{F}_t^T \mathbf{C}_{y,t}^{-1} \mathbf{F}_t + \mathbf{L}_t^T \mathbf{N}_{t+1} \mathbf{L}_t \quad \text{--}, \quad (\text{A10})$$

$$\tilde{\mathbf{x}}_t = \hat{\mathbf{x}}_t + \hat{\mathbf{C}}_t \mathbf{r}_t \quad \text{smoothed state mean} \quad (\text{A11})$$

$$\tilde{\mathbf{C}}_t = \hat{\mathbf{C}}_t - \hat{\mathbf{C}}_t \mathbf{N}_t \hat{\mathbf{C}}_t \quad \text{smoothed state covariance.} \quad (\text{A12})$$

Step 3

In order to study trends, we still need the full joint distribution of all the states given all the observations and the parameters, i.e. $p(\mathbf{x}_{1:N} | \mathbf{y}_{1:N}, \boldsymbol{\theta})$. This distribution does not have a closed-form solution. Instead, we can simulate realizations from it using a so-called simulation smoother algorithm (Durbin and Koopman, 2012, Sect. 4.9). The state space system equations (1) and (2) provide a direct way to recursively produce realizations of both the states and the observations. However, these will be independent of the original observations. Durbin and Koopman (2012) provide a simple way to produce samples from the distribution $\mathbf{x}_{1:N}^* \sim p(\mathbf{x}_{1:N} | \mathbf{y}_{1:N}, \boldsymbol{\theta})$. Briefly, the simulation smoother algorithm is the following. First, sample from the state space equations to get $\check{\mathbf{x}}_{1:N}$ and $\check{\mathbf{y}}_{1:N}$. Then, use the Kalman smoother with the new observations $\check{\mathbf{y}}_{1:N}$ to get smoothed states $\tilde{\check{\mathbf{x}}}_{1:N}$. Lastly, add the state residual process to the original smoothed states to obtain $\mathbf{x}_{1:N}^* = \check{\mathbf{x}}_{1:N} - \tilde{\check{\mathbf{x}}}_{1:N} + \tilde{\mathbf{x}}_{1:N}$. This can be repeated and used to study possible realizations of the model states, for example possible mean level dynamics as done in our ozone trend studies.

Step 4

So far, the parameter vector $\boldsymbol{\theta}$ has been assumed fixed. To estimate it, we need the marginal likelihood function $p(\mathbf{y}_{1:N} | \boldsymbol{\theta})$, i.e. the distribution of the observations, given the

parameter, with the state $\mathbf{x}_{1:N}$ uncertainty accounted for, i.e. integrated out. Fortunately, this likelihood is available as a by-product of the Kalman filter calculations for each fixed θ . By the assumed Markov property of the state space equations, this marginal likelihood function can be evaluated sequentially as a product of the individual time wise marginal likelihoods as

$$p(\mathbf{y}_{1:N}|\theta) = p(\mathbf{y}_1|\theta) \prod_{t=2}^N p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \theta). \quad (\text{A13})$$

These individual one-step-ahead predictive distributions of the observations are provided by the Kalman filter. In the case of a linear Gaussian model this likelihood is proportional to (ignoring a constant that does not depend on the model parameters)

$$p(\mathbf{y}_{1:N}|\theta) \propto \exp \left\{ -\frac{1}{2} \sum_{t=1}^N \left[(\mathbf{y}_t - \mathbf{F}_t \hat{\mathbf{x}}_t)^T \mathbf{C}_{y,t}^{-1} (\mathbf{y}_t - \mathbf{F}_t \hat{\mathbf{x}}_t) + \log(|\mathbf{C}_{y,t}|) \right] \right\}, \quad (\text{A14})$$

where $\hat{\mathbf{x}}_t$ is the one-step-ahead mean prediction of the state and $\mathbf{C}_{y,t}$ is the covariance matrix of predicted observation, both obtained by the Kalman filter formulas, given above in Eqs. (A1) and (A3).

Step 5

With this likelihood function and with a prior distribution for the parameter vector θ , a MCMC simulation can be performed to produce a sample of parameter values from the marginal posterior distribution $p(\theta|\mathbf{y}_{1:N})$. Thus, we can estimate and study the uncertainty in θ by the MCMC. We use an adaptive version of the random walk Metropolis–Hastings (MH) algorithm by Haario et al. (2006). It is a sequential Monte Carlo algorithm that proposes a value for the parameter vector from a suitably constructed proposal distribution and then either accepts or rejects it depending on the ratio of the posterior distribution at the proposed value to that calculated using the previously accepted value. By the Bayes formula, the posterior distribution is, up to a normalizing constant, equal to the likelihood times the prior. The basic idea behind MCMC is the fact that the acceptance probability depends only on the ratio of the posterior at two parameter values, and this will cancel out the hard-to-calculate proportional constant in the Bayes formula for the posterior distribution. See Gamerman (2006) for more details on the MCMC methodology. A MCMC Matlab toolbox used in the calculations is available from <http://helios.fmi.fi/~lainema/mcmc/>.

Step 6

As final steps in our DLM calculations, we can combine the simulation smoother and the MCMC outputs to produce samples from the joint posterior distribution $p(\mathbf{x}_{1:N}, \theta|\mathbf{y}_{1:N})$, as well as from the marginal posterior distribution of the state, $p(\mathbf{x}_{1:N}|\mathbf{y}_{1:N})$, where the uncertainty in θ has been accounted for.

Step 7

Lastly, a sample from this last distribution is used for the trend analyses by calculating trend statistics that depend on the estimated model states, as explained in Sect. 2.5.

Acknowledgements. This research has been supported by the Academy of Finland MIDAT and INQUIRE projects. We thank the valuable comments given by the editor and the anonymous referees.

Edited by: J. Urban

References

- Bates, B. C., Chandler, R. E., and Bowman, A. W.: Trend estimation and change point detection in individual climatic series using flexible regression methods, *J. Geophys. Res.*, 117, 1–9, doi:10.1029/2011JD017077, 2012.
- Bertaux, J. L., Kyrölä, E., Fussen, D., Hauchecorne, A., Dalaudier, F., Sofieva, V., Tamminen, J., Vanhellefont, F., Fanton D'Andon, O., Barrot, G., Mangin, A., Blanot, L., Lebrun, J. C., Pérot, K., Fehr, T., Saavedra, L., Leppelmeier, G. W., and Fraisse, R.: Global ozone monitoring by occultation of stars: an overview of GOMOS measurements on ENVISAT, *Atmos. Chem. Phys.*, 10, 12091–12148, doi:10.5194/acp-10-12091-2010, 2010.
- Chandler, R. E. and Scott, E. M.: *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*, John Wiley & Sons, 388 pp., 2011.
- Chu, W. P., McCormick, M. P., Lenoble, J., Brogniez, C., and Pruvost, P.: SAGE II inversion algorithm, *J. Geophys. Res.*, 94, 8339–8351, 1989.
- Doucet, A., Freitas, N., and Gordon, N.: *Sequential Monte Carlo methods in practice*, Statistics for engineering and information science, Springer, 581 pp., 2001.
- Durbin, T. and Koopman, S.: *Time Series Analysis by State Space Methods*, Oxford Statistical Science Series, Oxford University Press, second edn., 346 pp., 2012.
- Gamerman, D.: *Markov chain Monte Carlo – Stochastic simulation for Bayesian inference*, Chapman & Hall, second edn., 344 pp., 2006.
- Haario, H., Laine, M., Mira, A., and Saksman, E.: DRAM: Efficient adaptive MCMC, *Stat. Comp.*, 16, 339–354, doi:10.1007/s11222-006-9438-0, 2006.
- Hamilton, J. D.: *Time Series Analysis*, Princeton University Press, 799 pp., 1994.
- Harvey, A. C.: *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, 572 pp., 1990.
- Jones, A., Urban, J., Murtagh, D. P., Eriksson, P., Brohede, S., Haley, C., Degenstein, D., Bourassa, A., von Savigny, C., Sonkaew, T., Rozanov, A., Bovensmann, H., and Burrows, J.: Evolution of stratospheric ozone and water vapour time series studied with satellite measurements, *Atmos. Chem. Phys.*, 9, 6055–6075, doi:10.5194/acp-9-6055-2009, 2009.
- Kyrölä, E., Laine, M., Sofieva, V., Tukiainen, S., Tamminen, J., Päivärinta, S., Zawodny, J., and 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, doi:10.5194/acp-13-10645-2013, 2013.
- Lee, J. and Berger, J. O.: Space-time modeling of vertical ozone profiles, *Environmetrics*, 14, 617–639, doi:10.1002/env.608, 2003.
- Meiring, W.: Oscillations and Time Trends in Stratospheric Ozone Levels: A Functional Data Analysis Approach, *J. Am. Stat. Assoc.*, 102, 788–802, doi:10.1198/016214506000000825, 2007.
- Migon, H. S., Gamerman, D., Lopes, H. F., and Ferreira, M. A.: *Handbook of Statistics, Vol. 25, Bayesian Thinking: Mod. Comp.*, chap. Dynamic Models, 553–588, Elsevier, doi:10.1016/S0169-7161(05)25019-8, 2005.
- Newchurch, M. J., Yang, E.-S., Cunnold, D. M., Reinsel, G. C., Zawodny, J. M., and Russell, J. M.: Evidence for slowdown in stratospheric ozone loss: First stage of ozone recovery, *J. Geophys. Res. Atmos.*, 108, 4507, doi:10.1029/2003JD003471, 2003.
- Petris, G., Petrone, S., and Campagnoli, P.: *Dynamic Linear Models with R, Use R!*, Springer, 265 pp., 2009.
- Rodgers, C. D.: *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Scientific, 238 pp., 2000.
- Särkkä, S.: *Bayesian Filtering and Smoothing*, Institute of Mathematical Statistics Textbooks, Cambridge University Press, 254 pp., 2013.
- SPARC: *Assessment of Trends in the Vertical Distribution of Ozone*, Tech. Rep. 43, World Meteorological Organization Global Ozone Research and Monitoring Project, WCRP-SPARC Report 1, 1998.
- Steinbrecht, W., Claude, H., Schonenborn, F., McDermid, I. S., Leblanc, T., Godin-Beekmann, S., Keckhut, P., Hauchecorne, A., van Gijssels, J. A. E., Swart, D. P. J., Bodeker, G. E., Parrish, A., Boyd, I. S., Kampf, N., Hocke, K., Stolarski, R. S., Frith, S. M., Thomason, L. W., Remsberg, E. E., von Savigny, C., Rozanov, A., and Burrows, J. P.: Ozone and temperature trends in the upper stratosphere at five stations of the Network for the Detection of Atmospheric Composition Change, *Int. J. Remote Sens.*, 30, 3875–3886, doi:10.1080/01431160902821841, 2009.
- Toohey, M., Hegglin, M. I., Tegtmeier, S., Anderson, J., Añel, J. A., Bourassa, A., Brohede, S., Degenstein, D., Froidevaux, L., Fuller, R., Funke, B., Gille, J., Jones, A., Kasai, Y., Krüger, K., Kyrölä, E., Neu, J. L., Rozanov, A., Smith, L., Urban, J., von Clarmann, T., Walker, K. A., and Wang, R. H. J.: Characterizing sampling biases in the trace gas climatologies of the SPARC Data Initiative, *J. Geophys. Res.-Atmos.*, 118, 1–16, doi:10.1002/jgrd.50874, 2013.