Chapter 8
Stochastic Modelling of Geophysical Signal Constituents Within a Kalman Filter Framework

Olga Engels

Abstract Reliable trend estimation is of great importance while analyzing data. This importance is even enhanced when using the estimated trends for forecasting reasons in the context of climate change. While a constant trend might be a valid assumption for describing some geophysical processes, such as the tectonic motion or the evolution of Glacial Isostatic Adjustment (GIA) over very short geologic time frames, it is often too strong of an assumption to describe climatological data that might contain large inter-annual, multi-year variations or even large episodic events. It is therefore suggested to consider signal as a stochastic process. The main objective of the work described in this chapter is to provide a detailed mathematical description of geodetic time series analysis which allows for physically natural variations of the various signal constituents in time. For this purpose, state-space models are defined and solved through the use of a Kalman Filter (KF). Special attention is paid towards carefully estimating the noise parameters, which is an essential step in the KF. It is demonstrated how the time-correlated observational noise can be classified and handled within the state-space framework. The suggested methodology is applied to the analysis of real Gravity Recovery And Climate Experiment (GRACE), Global Positioning System (GPS), Surface Mass Balance (SMB) and global mean sea level time series. The latter is derived based on different satellite altimetry missions. The examples are illustrative in showing how the outlined technique can be used for estimating time-variable rates from different geodetic time-series with different stochastic properties.

Keywords State space model · Integrated random work · Time-varying signal constituents · Colored noise · Shaping filter · Autoregressive (AR) process · Prediction error decomposition

O. Engels (✉)
Institute of Geodesy and Geoinformation, University of Bonn, Nussallee 17,
53115 Bonn, Germany
e-mail: engels@geod.uni-bonn.de

8.1 Introduction

Geodetic observations such as from GPS, GRACE or altimetry is indispensable tool for variety of applications, in particular for those related to climate change. When analyzing geodetic data and making projections into the future, we usually rely on a rate which describes with which speed a process is changing. This rate is usually seen as a constant value and is estimated using a classical Least-Squares Adjustment (LSA). This interpretation of changes might be misleading if we are dealing with climate-related measurements that might include deviations from the deterministic linear trend assumption as well as from the constant seasonal amplitudes and phases. One example is Antarctica with its high inter-annual variations and very high episodic accumulation anomalies which are also called *climate noise* (Wouters et al. 2013). The question is whether these variations should be better modeled in the functional or in stochastic model. If we for instance use GPS to constrain Antarctic GIA, which is any viscoelastic response of the solid earth to changing ice loads and the most uncertain signal in Antarctica, we should correct GPS for elastic uplift. Elastic uplift is an immediate reaction of the solid earth to the contemporaneous mass changes. The contemporaneous mass changes contain interannual variations, multi-year variations or even large episodic events. The assumption of the deterministic trend might not capture all the variability and yield erroneous correction for elastic uplift that, in turn, yields erroneous constraint on GIA which is required for most techniques when estimating ice mass balance. Reliable estimation of ice mass balance is required, among others, for estimating sea level rise. The goal is therefore to estimate the changes as accurate as possible. For this, we model signal constituents stochastically using a state space model. The state space model includes an observation and a state process and can be written as

$$y_t = Z_t \alpha_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, H), \tag{8.1}$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \qquad \eta_t \sim N(0, Q), \quad t = 1, \ldots, n, \tag{8.2}$$

$$\alpha_1 \sim N(a_1, P_1), \tag{8.3}$$

The Eq. (8.1) is called *observation* equation with y_t being an observation vector at time t, α_t being an unknown state vector at time t and ε_t the irregular term with $H = I\sigma_\varepsilon^2$. The design matrix Z_t links y_t to α_t. The observation equation has the structure of a linear regression model where the unknown state vector α_t varies over time. The Eq. (8.2) represents a first order vector autoregressive model and consists of a transition matrix T_t, which describes how the state changes from one time step to the next, and the process noise η_t with $Q = I\sigma_\eta^2$. Process noise variance Q is assumed to be independent from H. The matrix R_t determines which components of the state vector α_t have the non-zero process noise. The initial state α_1 is $N(a_1, P_1)$ with a_1 and P_1 assumed to be known. Since we will restrict ourselves to data that are evenly spaced in time, the index t for the system matrices in Eqs. (8.1), (8.2) will be skipped hereafter.

Modeling signal constituents stochastically while representing them in state space form and using a KF framework to estimate the state parameters is a well-established methodology for treating different problems in econometrics as described in Durbin and Koopman (2012) and Harvey (1989). Durbin and Koopman (2012, Chap. 4.3) formulated the KF recursion to sequentially solve the linear state space model defined in Eqs. (8.1)–(8.3) using following equations:

$$v_t = y_t - Za_t, \qquad F_t = ZP_tZ^T + H,$$

$$a_{t|t} = a_t + P_tZ^TF_t^{-1}v_t, \qquad P_{t|t} = P_t - P_tZ^TF_t^{-1}ZP_t, \qquad (8.4)$$

$$a_{t+1} = Ta_t + K_tv_t, \qquad P_{t+1} = TP_t(T - K_tZ)^T + RQR^T.$$

The $K_t = TP_tZ^TF_t^{-1}$ is the so-called Kalman gain and v_t is the innovation with variance F_t. After computing $a_{t|t}$ and $P_{t|t}$, the state vector and its variance matrix can be predicted using

$$a_{t+1} = Ta_{t|t}, \quad P_{t+1} = TP_{t|t}T^T + RQR^T. \qquad (8.5)$$

By taking the entire time series $y_1 \ldots, y_n$ for $t = 1, \ldots, n$ into account, the state smoothing $\hat{\alpha}_t$ and its error variance V_t can be computed in a backward loop for $t = n, \ldots, 1$ initialized with $r_n = 0$ and $N_n = 0$ according to Durbin and Koopman (2012, Chap. 4.4):

$$r_{t-1} = Z^TF_t^{-1}v_t + L_t^Tr_t, \quad N_{t-1} = Z^TF_t^{-1}Z + L_t^TN_tL_t,$$

$$\hat{\alpha}_t = a_t + P_tr_{t-1}, \qquad V_t = P_t - P_tN_{t-1}P_t. \qquad (8.6)$$

The matrix L_t is given by $L_t = T - K_tZ$. The smoothing yields in general a smaller mean squared error than filtering, since the smoothed state is based on more information compared to the filtered state.

The covariance matrix for the smoothed state $\hat{\alpha}_t$ can be computed according to Durbin and Koopman (2012, Chap. 4.7):

$$\text{Cov}(\alpha_t - \hat{\alpha}_t, \alpha_j - \hat{\alpha}_j) = P_tL_t^TL_{t+1}^T \cdots L_{j-1}^T(I - N_{j-1}P_j) \qquad (8.7)$$

with $j = t + 1, \ldots, n$. If $j = t + 1$, $L_{t+1}^T \ldots L_t^T$ is replaced by the identity matrix I, which has a dimension of the estimated state vector.

In the next section, different time series models applicable to the analysis of geodetic data are summarized and put into the state space form defined in Eqs. (8.1)–(8.3).

8.2 Time Series Models

Different time series models exist as can be found in e.g., Harvey (1989), Durbin and Koopman (2012), Peng and Aston (2011). Here, we provide a detailed description of those models that are usually used to parameterize geodetic time series: trend, harmonic terms, step-like offsets, and coloured noise.

8.2.1 Trend Modelling

To fit a trend to time series, usually a deterministic function is used

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t, \quad t = 1, \ldots, n, \\
\varepsilon_t &\sim N(0, \sigma_\varepsilon^2)
\end{aligned}
\tag{8.8}
$$

with observation vector y_t at time $t = 1, \ldots, n$. The linear trend is $\mu_t = \alpha + \beta \cdot t$ with an intercept α and a slope β. The unmodeled signal and measurement noise in the time series is stored in the error term ε_t and is often assumed to be an independent and identically distributed (iid) random variable with zero mean and variance σ_ε^2.

By obtaining μ_t recursively from

$$
\mu_{t+1} = \mu_t + \beta, \quad \text{with} \quad \mu_0 = \alpha
\tag{8.9}
$$

and generating β_t by random walk process, yields

$$
\begin{aligned}
\mu_{t+1} &= \mu_t + \beta_t + \xi_t, \qquad \xi_t \sim N(0, \sigma_\xi^2), \\
\beta_{t+1} &= \beta_t + \zeta_t, \qquad\qquad \zeta_t \sim N(0, \sigma_\zeta^2).
\end{aligned}
\tag{8.10}
$$

This can be regarded as a local approximation to a linear trend. The trend is linear if $\sigma_\xi^2 = \sigma_\zeta^2 = 0$. If $\sigma_\zeta^2 > 0$, the slope β_t, is allowed to change in time. The larger the variance σ_ζ^2, the greater the stochastic movements in the trend, the more the slope is allowed to change from one time step to the next. Please note that any changes in slope is acceleration. Since there is no physical reason for the intercept to change over time, we model it deterministically by setting $\sigma_\xi^2 = 0$; this leads to a stochastic trend model called an *integrated random walk* (Harvey 1989; Durbin and Koopman 2012; Didova et al. 2016).

Representing the state vector in the state space form yields

$$
\alpha_t = \begin{bmatrix} \mu_t \ \beta_t \end{bmatrix}^T .
\tag{8.11}
$$

The observation equation reads

$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \alpha_t + \varepsilon_t \tag{8.12}$$

with

$$Z = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{8.13}$$

and remaining state space matrices being

$$T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad R = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad Q = \sigma_\eta^2, \quad H = \sigma_\varepsilon^2. \tag{8.14}$$

8.2.2 Modelling Harmonic Terms

Harmonic terms are important signal constituents in geodetic time series that are usually co-estimated with the trend. For this, the Eq. (8.8) is extended with a deterministic harmonic term

$$c_t = c \cdot \cos \omega t + s \cdot \sin \omega t, \tag{8.15}$$

yielding

$$y_t = \mu_t + \sum_{i=1}^{2} (c_i \cdot \cos \omega_i t + s_i \cdot \sin \omega_i t) + \varepsilon_t, \quad t = 1, \dots, n, \tag{8.16}$$

with angular frequency

$$\omega_i = \frac{2\pi}{T_i} T_s, \tag{8.17}$$

where $T_1 = 1$ for an annual signal, and $T_2 = 0.5$ for a semi-annual signal; T_s is the averaged sampling period

$$T_s = \frac{t_n - t_1}{n - 1}. \tag{8.18}$$

To allow harmonic terms to evolve in time, they can be built up recursively similar to the linear trend in the previous section, leading to the stochastic model

$$\begin{aligned} c_t &= c_{t-1} \cdot \cos \omega + s_{t-1} \cdot \sin \omega + \varsigma_t, \\ s_t &= -c_{t-1} \cdot \sin \omega + s_{t-1} \cdot \cos \omega + \varsigma_t^*, \end{aligned} \tag{8.19}$$

where ς_t and ς_t^* are white-noise disturbances that are assumed to have the same variance (i.e., $\varsigma_t \sim N(0, \sigma_\varsigma^2)$) and to be uncorrelated. These stochastic components allow the parameters c and s and in turn the corresponding amplitude A_t and phase ϕ_t to evolve over time

$$A_t = \sqrt{c_t^2 + s_t^2}$$

$$\phi_t = -\tan^{-1}(s_t/c_t) - \tau\omega) \bmod 2\pi, \text{ with } \tau = \frac{t - t_1}{T_s} \tag{8.20}$$

Inserting the stochastic trend and stochastic harmonic models into Eq. (8.8) yields

$$y_t = \mu_t + c_{1,t} + c_{2,t} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{8.21}$$

with $c_{1,t}$ and $c_{2,t}$ being annual and semi-annual terms, respectively. Please note that Eq. (8.21) can be easily extended by additional harmonic terms using the stochastic model of Eq. (8.19) with the corresponding angular frequencies (Harvey 1989; Durbin and Koopman 2012; Didova et al. 2016).

The state vector becomes

$$\alpha_t^{[b]} = \begin{bmatrix} \mu_t & \beta_t & c_{1,t} & s_{1,t} & c_{2,t} & s_{2,t} \end{bmatrix}^T \tag{8.22}$$

with index b emphasizing that the integrated random walk along with the annual and semiannual components represent a *basic* model for geodetic time series. The observation equation gets the form

$$y_t = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \alpha_t + \varepsilon_t \tag{8.23}$$

with

$$Z^{[b]} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}. \tag{8.24}$$

The remaining state space matrices can be written as

$$T^{[b]} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos\omega_1 & \sin\omega_1 & 0 & 0 \\ 0 & 0 & -\sin\omega_1 & \cos\omega_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos\omega_2 & \sin\omega_2 \\ 0 & 0 & 0 & 0 & -\sin\omega_2 & \cos\omega_2 \end{bmatrix}, \tag{8.25}$$

$$R^{[b]} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad Q^{[b]} = \begin{bmatrix} \sigma_\zeta^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{\varsigma_1}^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{\varsigma_1}^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{\varsigma_2}^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{\varsigma_2}^2 \end{bmatrix}, \quad H = \sigma_\varepsilon^2.$$

8.2.3 Modelling Coloured Noise

If the observations are close together, they may contain temporally correlated, so-called *coloured* noise. Here, we aim at co-estimating the coloured noise within the described state space model solved within the KF framework as described in Sect. 8.1. When not modeling the coloured noise in observations such as from GPS, the solutions for the noise parameters might be outside a reasonable range (e.g., zero noise variance or noise variance exceeding a reasonable limit). For this, a so-called *shaping filter* developed by Bryson and Johansen (1965) is used. Since the KF requires a time-independent noise input, the observational noise ϵ_t is parameterized in such a way that the process noise matrix consists of a time-independent noise while the output, the state vector forming ϵ_t, is time-dependent. This is done by extending the state vector α_t in Eq. (8.22) with the noise. For purposes of modeling temporally correlated noise in the geodetic time series within the state space framework, an Autoregressive Moving Average (ARMA) model that subsumes Autoregressive (AR) and Moving Average (MA) models can be utilized (Didova et al. 2016).

An ARMA model of order (p, q) is defined as

$$\varepsilon_t = \sum_{j=1}^{l} \phi_j \varepsilon_{t-j} + \varkappa_t + \sum_{j=1}^{l-1} \theta_j \varkappa_{t-j}, \quad t = 1, \ldots, n, \tag{8.26}$$

with $l = \max(p, q + 1)$, autoregressive parameters $\phi_1, \ldots, \phi_p$ and moving average parameters $\theta_1, \ldots, \theta_q$. $\varkappa_t$ is a serially independent series of $N(0, \sigma_\varkappa^2)$ disturbances. Some parameters of an ARMA model can be zero, which provides two special cases: (i) if $q = 0$, it is an autoregressive process AR(p) of order p and (ii) if $p = 0$, it is a moving-average process MA(q) of order q.

Coloured noise ε_t can be put into state space form as:

$$\alpha_t^{[\varepsilon]} = \begin{bmatrix} \varepsilon_t \\ \phi_2 \varepsilon_{t-1} + \cdots + \phi_l \varepsilon_{t-l+1} + \theta_1 \varkappa_t + \cdots + \theta_{l-1} \varkappa_{t-l+2} \\ \phi_3 \varepsilon_{t-1} + \cdots + \phi_l \varepsilon_{t-l+2} + \theta_2 \varkappa_t + \cdots + \theta_{l-1} \varkappa_{t-l+3} \\ \vdots \\ \phi_l \varepsilon_{t-1} + \theta_{l-1} \varkappa_t \end{bmatrix} \tag{8.27}$$

with $\eta^{[\varepsilon]} = \varkappa_{t+1}$. The index ε emphasizes that the system matrices are attributed to the coloured noise that is modeled using an ARMA-process:

$$T^{[\varepsilon]} = \begin{bmatrix} \phi_1 & 1 & & 0 \\ \vdots & & \ddots & \\ \phi_{l-1} & 0 & & 1 \\ \phi_l & 0 & \cdots & 0 \end{bmatrix}, \quad R^{[\varepsilon]} = \begin{bmatrix} 1 \ \theta_1 \ \cdots \ \theta_{l-1} \end{bmatrix}^T, \quad Z^{[\varepsilon]} = \begin{bmatrix} 1 \ 0 \ 0 \cdots 0 \end{bmatrix}. \tag{8.28}$$

Combining the *basic* time series model with ARMA-model yields

$$\alpha_t = (\alpha_t^{[\varepsilon]}, \alpha_t^{[b]}) \tag{8.29}$$

with the system matrices

$$
\begin{aligned}
Z_t &= (Z^{[\varepsilon]}, Z^{[b]}), \quad T = \mathrm{diag}(T^{[\varepsilon]}, T^{[b]}), \\
R &= \mathrm{diag}(R^{[\varepsilon]}, R^{[b]}), \\
Q &= \mathrm{diag}\big(\big[\sigma_{\varkappa_{t+1}}^2 \ \sigma_\zeta^2 \ \sigma_{\varsigma_1}^2 \ \sigma_{\varsigma_1}^2 \ \sigma_{\varsigma_2}^2 \ \sigma_{\varsigma_2}^2\big]\big).
\end{aligned} \tag{8.30}
$$

Detecting p and q for ARMA(p, q)

The (p, q) of the ARMA model define the amount of ϕ and θ coefficients necessary to parameterize coloured noise ε_t in Eq. (8.27). That means that we first need to know how large p and q have to be chosen. To get an idea about the appropriate (p, q) we can (i) follow Didova et al. (2016) and perform a power density function (PSD) analysis or (ii) we can analyze usually used criteria to identify which model provides the 'best' fit to the given time series.

PSD Analysis

When using a PSD analysis, the idea is that the residuals, obtained after fitting a deterministic function to the given time series, represent an appropriate approximation of the noise contained in the time series. For this, we first set the process noise variance σ_η^2 to zero and σ_ε^2 to one, which is equivalent to the commonly used LSA. We then estimate the state vector using filtering and smoothing recursions described in Sect. 8.1. The state vector can for instance consist of the components contained in the *basic* model described in Eq. (8.22). We estimate the state vector by KF considering quantities introduced in Sect. 8.1

$$\hat{\varepsilon}_t = H(F_t^{-1} v_t - K_t^T r_t). \tag{8.31}$$

The KF is used instead of LSA, because KF allows the residuals to be computed at each time step $t = n, \ldots, 1$ regardless possibly existing data gaps in the time series. The postfit residuals obtained after fitting a deterministic model to the observations represent an approximation of the observational noise. In the next step, we compute the PSD function of the approximate coloured noise. Then, using this PSD function we estimate the parameters of the pure recursive (MA) and non-recursive (AR) part of the ARMA filter by applying the standard Levinson–Durbin algorithm (Farhang-Boroujeny 1998) to $p, q \in \{0, \ldots, 5\}$. We limit the order to 5 to keep the dimension of the state vector α_t relatively short. The estimated parameters are then used to compute the PSD function of the combined ARMA(p, q) solution. Finally, we use Generalized Information Criterion (GIC) to select the PSD of the ARMA model that best fits the PSD of the approximate coloured noise. The (p, q) of this ARMA model

define the amount of ϕ and θ coefficients necessary to parameterize coloured noise ε_t in Eq. (8.27).

Criteria for 'best' fit

It is important to understand that the residuals, obtained after fitting a deterministic function to the given time series, may still contain unmodeled time-dependent portion of the signal. That means that these residuals are only an approximation of the observational noise.

To get an idea about which ARMA(p, q) model is the most appropriate to parameterize the observational noise of a particular time series, we can compare the log-likelihood value of a particular fitted model. Since the loglikelihood value is usually larger for larger number of parameters (for larger p and/or q), we also need a criterion that can deal with different amount of parameters. For this, Akaike Information Criterion (AIC) and the Bayessian Information Criterion (BIC) can be used (Harvey 1989).

ARMA and Long-Range Dependency

ARMA, as a high-frequency noise model, is known to describe a short-range dependency (have a short memory). The noise in GPS time series, however, is believed to contain long-range dependency (have a long memory). Therefore, a power law model is usually used to model GPS noise. According to Plaszczynski (2007), power law noise, which has a form $\frac{1}{f^\alpha}$, is a stochastic process with a spectral density having a power exponent $0 < \alpha \leq 2$. For GPS time series analysis, the power law model with $\alpha = 1$ and $\alpha = 2$ is usually used. In case of $\alpha = 2$, we are talking about a random walk noise, which is an analogue of the Gaussian random walk we employed to model time-varying signal constituents. Plaszczynski (2007) has shown that ARMA models can be used to generate random walk noise. This can be immediately seen from the mathematical description of the random walk process

$$\varepsilon_t = \varepsilon_{t-1} + \varkappa_t \tag{8.32}$$

with ε_t being the observation at time t. The Eq. (8.26) is equivalent to Eq. (8.32) in case $q = 0$, $p = 1$ and $\phi_1 = 1$. That means that AR(1), which is a special case of ARMA, can represent random walk noise.

In case of $\alpha = 1$, we are talking about a flicker noise, which is difficult to represent within the state space model and therefore, can be only approximated. On the one hand, flicker noise can be approximated by a linear combination of independent first-order Gauss-Markov processes, as it has been shown by Dmitrieva et al. (2015). On the other hand, Didova et al. (2016) have shown that also ARMA models can approximate flicker noise, when a special ARMA case—AR(p)—is used. In their supplement, Didova et al. (2016) demonstrated that an infinite number of parameters p would be required to exactly describe flicker noise. However, limiting the maximum order p to 5 to control the dimension of the state vector $\alpha_t^{[\varepsilon]}$ would still be sufficient to approximate flicker noise within the state space formalism.

8.2.4 Modelling of Offsets

Some geodetic data, such as GPS observations might include offsets that must be parameterized to avoid additional errors in the estimated trends (Williams 2003). If the offsets are related to hardware changes, they are step-like and easy to include into the state space model. For this, a variable w_t is defined as:

$$w_t = \begin{cases} 0, & t < \tau_e, \\ 1, & t \geq \tau_e. \end{cases} \tag{8.33}$$

Including this in the observation equation Eq. (8.21) gives

$$y_t = \mu_t + c_{1,t} + c_{2,t} + \delta \, w_t + \varepsilon_t, \quad t = 1, \ldots, n, \tag{8.34}$$

with δ measuring the change in the offset at a known epoch τ_e. For k offsets, the state vector can be written as

$$\alpha_t^{[\delta]} = [\delta_1 \ldots \delta_k]^T. \tag{8.35}$$

We can now combine the different models: (i) the basic model defined in Eq. (8.22), (ii) the coloured noise from Eq. (8.27) modeled here using an ARMA-process, and (iii) the model for k offsets from Eq. (8.35)

$$\alpha_t = (\alpha_t^{[\varepsilon]}, \alpha_t^{[b]}, \alpha_t^{[\delta]}), \tag{8.36}$$

with the system matrices

$$\begin{aligned}
Z &= (Z^{[\varepsilon]}, Z^{[b]}, \mathbf{I}_k), \quad T = \mathrm{diag}(T^{[\varepsilon]}, T^{[b]}, \mathbf{I}_k), \\
R &= \mathrm{diag}(R^{[\varepsilon]}, R^{[b]}, \mathbf{0}_k), \\
Q &= \mathrm{diag}\left(\left[\sigma_{\varkappa_{t+1}}^2 \; \sigma_\zeta^2 \; \sigma_{\varsigma_1}^2 \; \sigma_{\varsigma_1}^2 \; \sigma_{\varsigma_2}^2 \; \sigma_{\varsigma_2}^2\right]\right),
\end{aligned} \tag{8.37}$$

where Z, T and R with corresponding indices have been defined in Sects. 8.2.2 and 8.2.3.

8.2.5 Hyperparameters

The parameters that build the system matrices Q and H decide about the variability of the estimated signal constituents (the variability of the parameters stored in the state vector α). For instance, the larger is σ_ζ^2, the more the slope is allowed to change from one time step to the next; the larger is $\sigma_{\varsigma_1}^2$, the more variability is allowed for the corresponding harmonic term. That means that if we chose one of these parameters too large, it will absorb variations possibly originating from other signal components.

There parameters, therefore, govern the estimates of the state vector and are called *hyperparameters*. These parameters are stored in vector ψ

$$\psi = \left[\psi_\varepsilon \; \psi_\eta\right]^T \tag{8.38}$$

and can be either assumed to have a certain value, as it was done by Davis et al. (2012), or they can be estimated based on the Kalman filter. Because we do not have any a priori information regarding the process noise, we estimate the hyperparameters. One way to do so is by maximizing likelihood. If a process is governed by hyperparameters ψ, which generate observations y_t, the likelihood L of producing the y_t for known ψ is according to Harvey (1989)

$$L(Y_n|\psi) = p(y_1, \ldots, y_n) = p(y_1) \prod_{t=2}^{n} p(y_t|Y_{t-1}). \tag{8.39}$$

The $p(y_t|Y_{t-1})$ represents the distribution of the observations y_t conditional on the information set at time $t-1$, that is $Y_{t-1} = \{y_{t-1}, y_{t-2}, \ldots, y_1\}$. In praxis, we usually work with loglikelihood $logL$ instead of the likelihood L

$$\text{logL}(Y_n|\psi) = \sum_{t=1}^{n} p(y_t|Y_{t-1}). \tag{8.40}$$

The hyperparameters ψ are regarded as *optimal* if the $logL$ is maximized or the $-logL$ is minimized. Since the $E(y_t|Y_{t-1}) = Z_t a_t$, the innovation $v_t = y_t - Z_t a_t$ (Sect. 8.1) with the variance $F_t = Var(y_t|Y_{t-1})$, inserting $N(Z_t a_t, F_t)$ into Eq. (8.40) yields

$$\log L(Y_n|\psi) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}(\log|F_t| + v_t^T F_t^{-1} v_t), \tag{8.41}$$

which is computed from the Kalman filter output (Eq. (8.4)) following Durbin and Koopman (2012, Chap. 7). Harvey and Peters (1990) refer to obtaining the $logL$ in such a way as via the *prediction error decomposition*.

   Because the hyperparameters represent standard deviations that cannot be negative, they are defined for our basic state space vector from Eq. (8.22) as

$$\psi = 0.5 \log \left[\sigma_\varepsilon^2 \; \sigma_\eta^2\right]^T = 0.5 \log \left[\sigma_\varepsilon^2 \; \sigma_\zeta^2 \; \sigma_{\varsigma_1}^2 \; \sigma_{\varsigma_2}^2\right]^T. \tag{8.42}$$

We are numerically searching for the optimal hyperparameters ψ that minimize the $-\text{logL}(Y_n|\psi)$ (the negative logL is called objective function). The lower the dimension of the hyperparameters vector, the faster an optimization algorithm might converge. However, this does not guarantee that the optimal solution will be found if the optimization problem is non-convex. An optimization problem is non-convex, if additionally to the global minimum (that we are aiming at to find), several local

minimum points exist. At these local minimum points, the value of the objective function $-\log L$ is different than at the global minimum. That means that if we start searching for the global minimum in the proximity of a local minimum, the optimization algorithm will suggest the local minimum as the optimal solution. It follows from here that the starting point (also called initial guess) is crucial for finding the optimal set of hyperparameters and in turn, reliable parameters stored in the state vector α that are the signal constituents we are interested in to estimate.

In other words, if the problem is non-convex, there is no guarantee of finding a global minimum. Depending on the initial guess, the solution might be a local minimum meaning that there is non unique solution. The preferred solution is significantly depending on the length of the state space vector, on the length of the time series (the longer the better), on the noise content and kind, on the non-convexity of the problem, etc. What exactly causes the non-convexity and to which extent (data, definition of the transition matrix, or of the state vector, or of the hyperparameters vector, or most likely the interaction of all aforementioned components) is a challenging topic that needs to be investigated, but is out of the scope of this study. Therefore, we recommend to always check the spectral representation of the estimated signal constituents (Sect. 8.3.4) and if independent observations are available, to use them for validation (Sect. 8.3.3).

There are, however, tools to increase the chance of finding the optimal solution by limiting the parameter search space and/or by applying explicit constraints on the hyperparameters (Didova et al. 2016). Yet, we first should decide on which optimization algorithm to use. Since the problem we are dealing with is non-convex, we use an Interior-Point (IP) algorithm as described in Byrd et al. (1999) to find hyperparameters that minimize our objective function. This algorithm is a gradient-based solver, which means that the gradient of the objective function is required. According to Durbin and Koopman (2012, Chap. 7), the gradient of the objective function can be analytically computed using the quantities calculated in Sect. 8.1:

$$\frac{\partial \log L(Y_n|\psi)}{\partial \psi} = \frac{1}{2} \sum_{t=1}^{n} \text{tr} \left\{ (u_t u_t^T - D_t) \frac{\partial H_t}{\partial \psi} \right\} + \frac{1}{2} \sum_{t=2}^{n} \text{tr} \left\{ (r_{t-1} r_{t-1}^T - N_{t-1}) \frac{\partial R_t Q_t R_t^T}{\partial \psi} \right\},$$

(8.43)

where $u_t = F_t^{-1} v_t - K_t^T r_t$ and $D_t = F_t^{-1} + K_t^T N_t K_t$.

To increase the likelihood of getting the optimal solution, we start the IP algorithm for different starting points. The larger the amount of starting points the higher the probability of finding the global minimum, the longer the execution time of the algorithm. One should however ensure that after each run numerically the same optimal solution is obtained. From all the different solutions, the solution is used to estimate the state vector α that provides the smallest objective function value (Anderssen and Bloomfield 1975). The uniformly distributed starting points are randomly generated.

To further increase the likelihood of getting the optimal solution, we generate the starting points within a finite search space. For this, we define lower and upper bounds for our hyperparameters. The lower bounds are set equal to zero, as the standard deviations can not be less than zero. To define upper bound, the traditional LSA is utilized. We first fit a basic deterministic model (trend, annual, semiannual terms) to

the analyzed time series. The variance of the postfit residuals is used as an upper bound for the σ_ε^2. The variance of the postfit residuals obtained after fitting the deterministic model is larger than the σ_ε^2, as it contains additionally to the unmodeled signal and measurement noise, possible fluctuations in the modeled trend, annual and semi-annual components. The σ_ε^2 in Eq. (8.23) does not contain possible fluctuations in the modeled terms, since we model them stochastically as described in Sect. 8.2. The upper bounds for harmonic terms are defined in similar way. Deterministic harmonic terms are simultaneously estimated using LSA within a sliding window of minimum two years. The maximum size of the sliding window corresponds to the length of the analyzed time series. In this way, a sufficient amount of for instance annual amplitudes is estimated for different time periods. The variance computed based on the multiple estimates is regarded as the upper bound for $\sigma_{\varsigma_1}^2$. This is an upper bound, since the standard deviations computed for different time intervals indicate possible signal variations within the considered time span and contain possible variations within the trend component. These standard deviations are always larger than the process noise of the corresponding signal, which only represents the variations from one time step to the next. The upper bound for other harmonic terms are defined in the same way. The search space associated with the trend component σ_ζ^2 is only limited through the lower bound.

The importance of limiting the parameter search space within a non-convex optimization problem is demonstrated in Didova et al. (2016). As already mentioned, the reliability of the estimated hyperparameters can be verified by investigating the amplitude spectrum of the estimated signal constituents. As there is no recipe for solving a non-linear problem that has several local minima (or maxima), any prior knowledge which might be available should be used. This can be easily done by setting explicit constrains for instance on the noise parameter σ_ε^2. However, before introducing a constraint, it should be verified that this constraint is indeed supported by the data (for more details the reader is reffed to Didova et al. 2016).

8.3 Application to Real Data

In this section, we show how the time-varying trends can be estimated from different geodetic time series that feature different stochastic properties. For this, we estimate time-variable rates from GPS and GRACE at the GPS stations in Antarctica that are located in regions where (i) a high signal-to-noise ratio is expected and (ii) an apriori information regarding the geophysical processes exists. For monthly available GRACE time series, a white noise assumption is used. In contrast to that, for daily GPS observations we co-estimate coloured noise using the procedure described in Sect. 8.2.3. If time-varying rates derived from GRACE and GPS exhibit the same behavior, we interpret the estimated variations as a signal and not as noise. To strengthen this interpretation, we derive time-varying rates utilizing monthly SMB data from Regional Atmospheric Climate Model (RACMO) at the same GPS stations. The hypothesis is that (i) all three techniques (RACMO, GRACE, and GPS) should

capture small-scale accumulation variability present in SMB and (ii) this variability can be detected using the described state space framework solved by KF.

Moreover, we analyze the Global Mean Sea Level (GMSL) time series which has a temporal resolution of 10 days. The time series is derived using a combination of different altimetry products over 25 years.

8.3.1 Pre-processing of GRACE and SMB

GRACE and SMB time series are monthly available. GRACE time series are obtained using unconstrained DMT2 monthly GRACE solutions completed to degree n and order m 120 (Farahani et al. 2016). Degree-1 coefficients were added using values generated from the approach of Swenson et al. (2008), and the C_{20} harmonics were replaced with those derived from satellite laser ranging (Cheng and Tapley 2004). Since DMT2 solutions are available starting from February 2003 to December 2011, we focus on analyzing this time span.

SMB is the sum of mass gain (precipitation) and mass loss (e.g., surface runoff) provided at the spatial resolution of 27 km. SMB reflects mass changes within the firn layer only. GRACE signal over Antarctica also reflects mass changes within the firn layer, but additionally it contains changes due to GIA and ice dynamics. We remove the GIA-induced mass changes from the total GRACE signal using GIA rates derived in Engels et al. (2018).

To ensure a fair comparison between GRACE and SMB data in terms of spatial resolution, the dynamic patch-approach described in Engels et al. (2018) is applied to retrieve surface densities from both, GRACE and SMB data.

To enable a direct comparison between the GRACE, SMB and GPS data, we convert GRACE and SMB derived monthly surface densities into vertical deformation as observed by GPS. For this, derived surface densities are first converted into spherical harmonic representation of the surface mass C_{nm}^q, S_{nm}^q according to Sneeuw (1994). In the next step, these spherical harmonics are converted into spherical harmonics in terms of vertical deformation C_{nm}^h, S_{nm}^h following Kusche and Schrama (2005) as

$$\left.\begin{array}{c} C_{nm}^h \\ S_{nm}^h \end{array}\right\} = \frac{3\rho_w}{\rho_e}\frac{h_n'}{2n+1}\left\{\begin{array}{c} C_{nm}^q \\ S_{nm}^q \end{array}\right\} \tag{8.44}$$

using the density of water ρ_w, the density of Earth ρ_e, and Load Love numbers h_n'. Finally, monthly spherical harmonics in terms of vertical deformation are synthesized at the locations of GPS stations resulting in a time series of vertical deformation.

The resulting monthly time series derived based on GRACE and SMB data are used to estimate time-varying rates along with stochastically modeled known harmonics (annual and semiannual components for GRACE and SMB data, and additionally tidal S2 periodic term for GRACE). For both datasets, a constant intercept is co-estimated. The state vector has the form as described in Eq. (8.22) with an additional tidal S2 harmonic term (161 days) to parameterize GRACE time series.

8.3.2 Pre-processing of GPS

We use daily GPS-derived vertical displacements at two permanent GPS stations in Antarctica: (i) at VESL station that is located in Queen Maud Land of East Antarctica and (ii) at CAS1 GPS station that is located in Wilkes Land. The processing of the GPS displacements followed that of Thomas et al. (2011), although GPS observations were intentionally not corrected for non-tidal atmospheric loading. To be more consistent with GRACE-derived data, we corrected the GPS data using the Atmospheric and Oceanic De-aliasing (AOD) product (Flechtner 2007).

The GPS observations contain step-like offsets within the analyzed time period: at the CAS1 station two offsets occurred (in Oct. 2004 and Dec. 2008) and at the VESL station one (in Jan. 2008). Moreover, GPS time series might contain outliers that should be removed from the data prior applying KF to it. This is because KF is not robust against outliers. We used Hampel filter to detect the outliers (Pearson 2011) and removed the observations from the time series even if the outliers were detected only in the horizontal or vertical component.

Another issue when dealing with GPS data is that the observations might be not evenly spaced in time, partially yielding relatively large data gaps. In general, KF can easily deal with irregularly distributed observations. However, we need equally spaced data to be able to model temporally correlated noise of higher orders (Sect. 8.2.3) within the state space framework. For this, we fill short gaps with inter-polated values. Long gaps are filled with NAN values. For the daily GPS data, we defined a gap to be long if more than one week of data is missing (seven consecutive measurements).

To estimate time-varying rates from GPS time series, slope, annual and semi-annual signal constituents are allowed to change in time. The state vector has the form as described in Eq. (8.36) containing step-like offsets and an ARMA-process to parameterize the coloured noise. The order p and q of the ARMA-process was detected by performing the PSD analysis as described in Sect. 8.2.3. Figures 8.1a and b demonstrate the estimated time-varying slope along with the time-varying annual signal for both analyzed GPS sites.

8.3.3 GRACE-SMB-GPS

When comparing the time-varying rates of vertical deformations obtained from the three independent techniques, three important aspects should be considered. First, GPS observations are discrete point measurements that are sensitive to local effects and GRACE and SMB results are spatially smoothed over the patches defined by Engels et al. (2018). Second, the GPS observations used here are global. They refer to a reference frame with origin in the Center-of-Mass (CM) of the total Earth system while the vertical deformations we obtained from GRACE and SMB are regional. To enable a fair comparison of GRACE and SMB time series with those of GPS,
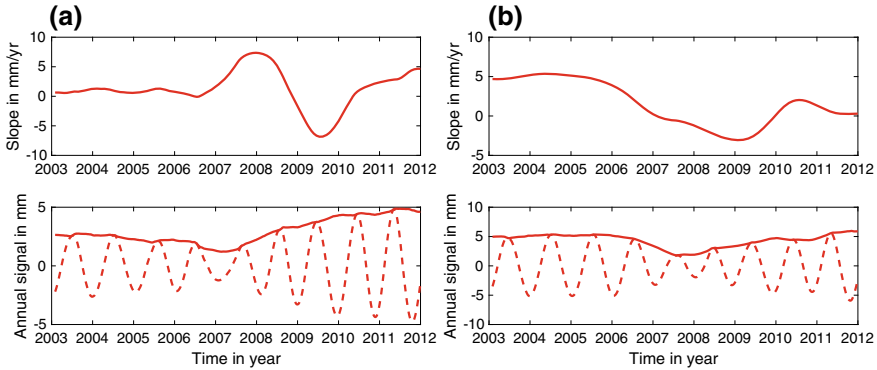
Fig. 8.1 Time-varying slope (top) and annual signal (bottom, dashed line) along with the time-varying annual amplitude (bottom, solid line) for GPS vertical site displacements at the **a** CAS1 and **b** VESL station (without any corrections applied)

we should 'regionalize' GPS observation to Antarctica. For this, we should reduce the signal originating from non-Antarctic sources from the GPS signal. Third, GPS observations contain global GIA whereas GRACE and SMB are GIA-free assuming that a correct GIA signal is subtracted from GRACE data. GIA contaminates the GPS secular trend at very low degrees, mostly driven by GIA in the Northern Hemisphere (Klemann and Martinec 2011) and the leakage from non-Antarctic sources is also mostly originating from changes in the spherical harmonic coefficients of degree-one and C_{20}. We therefore remove the time-varying slope obtained from degree-one and C_{20} time series from the time-varying slope obtained from GPS observations. The assumption is here that these low-degree coefficients are a sufficient first-order approximation of the non-Antarctic leakage.

Figures 8.2 and 8.3 show three time-varying rates estimated using GRACE, SMB, and GPS time series for the VESL and CAS1 station, respectively. In these figures, GPS-derived time-varying rates are corrected for degree-one, C_{20}, and atmospheric non-tidal variations. There is a high correlation of 0.9 and 0.7 between the SMB- and GRACE-derived time-varying rates for the CAS1 and VESL station, respectively. The correlation between GPS- and GRACE-derived time-varying rates is slightly lower: 0.6 for the CAS1, and 0.8 for the VESL station. Although the correlation is generally high, a systematic bias between the three estimates might exist. This bias can be explained by geophysical processes. The bias between the SMB- and GRACE-derived time-varying rates would most likely be due to the fact that SMB data contain variations within the firn layer and GRACE-derived rates represent variations within the firn and ice layer after being corrected for GIA. That means that after subtracting SMB from GRACE rates, the remainder should represent variations mostly associated with ice dynamics. We therefore subtract the mean slope of SMB from the mean slope of GRACE, assuming the difference is representative for ice dynamics. In this way computed *bias* between the SMB- and GRACE-derived time-varying rates is added to the time-varying SMB rates resulting in the shift of the

Fig. 8.2 Time-varying slope for GRACE (blue), GPS (green), and SMB (red) time series at the geolocation of the VESL site in Queen Maud Land, East Antarctica. **a** Original time-varying rates and **b** shifted time-varying rates (blue: GRACE+GIA, red: SMB+ice dynamics+GIA). Time-varying error bars are 1σ
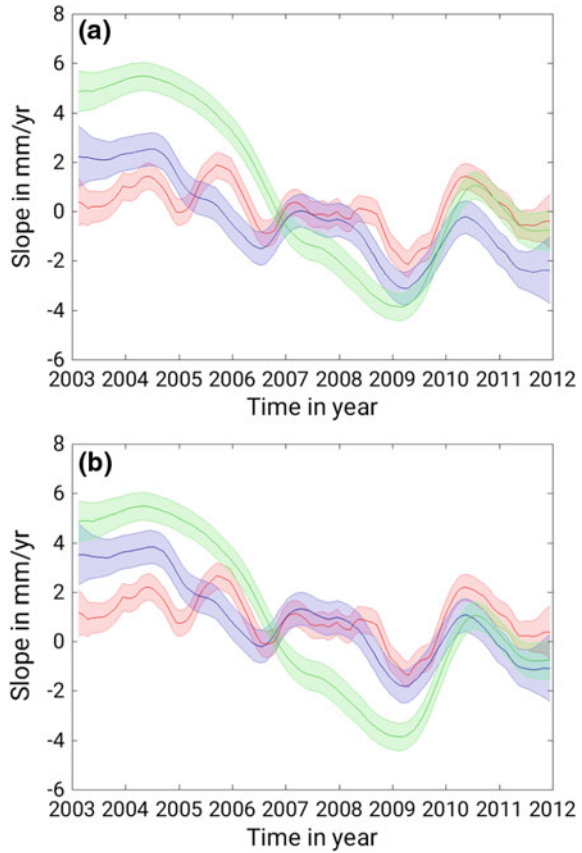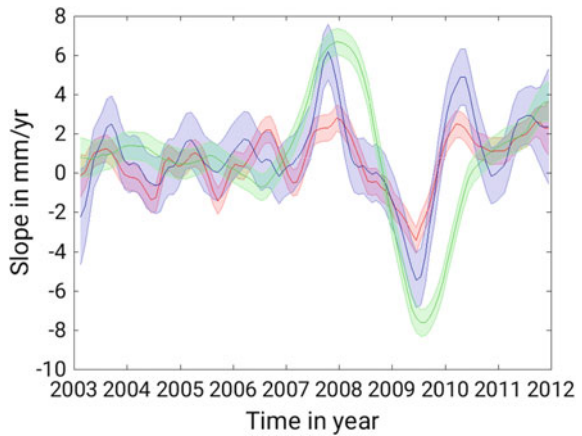
Fig. 8.3 Time-varying slope for GRACE (+GIA) (blue), GPS (green), and SMB (+ice dynamics+GIA) (red) time series at the geolocation of the CAS1 site in Wilkes Land, East Antarctica. Time-varying error bars are 1σ

SMB-derived time-varying rates towards the GRACE-derived time-varying rates (Fig. 8.2b). The mean rate for ice dynamics is estimated to be 0.3 ± 0.09 and -0.5 ± 0.08 mm for CAS1 and VESL station, respectively. Please note that we do not show the original plot of the three time series for the CAS1 station, as the difference between the 'shifted' and 'unshifted' version is small and cannot be detected by visual inspection.

The bias between the GPS- and GRACE-derived time-varying rates would most likely be due to the fact that GPS data contain variations due to both surface processes (firn, ice) and GIA whereas GRACE-derived rates are GIA-free, since we removed the GIA rates from them in the pre-processing step as described in Sect. 8.3.1. It follows that the difference between the mean slope of GPS and the mean slope from GRACE should mainly represent the solid-earth deformation associated with GIA. In this way computed *bias* between the GRACE- and GPS-derived time-varying rates is added to the time-varying GRACE rates resulting in the shift of the GRACE-derived time-varying rates towards the GPS-derived time-varying rates (Figs. 8.2b, 8.3). Please note that the bias attributed to GIA is also added to the time-varying rates derived from SMB allowing a direct comparison between the three independent techniques. The mean rate for GIA is estimated to be -0.2 ± 0.8 and 1.3 ± 0.4 mm for CAS1 and VESL station, respectively.

After correcting the SMB- and GRACE-derived time-varying rates for ice dynamics and GIA, respectively we can compute the agreement between SMB/GRACE and GRACE/GPS time-varying rates in terms of Weighted Root Mean square Residual (WRMS) reduction in percent following Tesmer et al. (2011). This quantity takes into account the magnitude and behavior of the time-varying rates estimated from two different time series as well their uncertainties. For ice dynamics corrected SMB time-varying rates explain 49 and 27% of the GRACE slope WRMS for CAS1 and VESL GPS stations, respectively. For GIA corrected GRACE time-varying rates explain 21 and 40% of the GPS slope WRMS for CAS1 and VESL GPS stations, respectively. Please note the improved agreement between the magnitude of the peaks derived from GRACE and GPS rates at the CAS1 station compared to the results shown in Didova et al. (2016) (their Fig. 9). The better agreement is mainly caused by the dynamic patch approach applied to the GRACE data, which localizes the signal and thus, improves its recovery (Engels et al. 2018).

Despite the visual inspection of Figs. 8.2 and 8.3 the WRMS reduction in percent confirms a good agreement between the temporal variations derived from three independent techniques. If we only compared the deterministic trends from GRACE, SMB and GPS, we would not be able to get any insights into the geophysical processes. Analyzing the time-varying rates allows us to state that all three techniques capture small-scale accumulation variability modeled by SMB at the two GPS location. In particular, both GRACE and GPS seem to observe the same geophysical processes with similar magnitude. We interpret these geophysical processes as signal and not as noise. Under some assumptions as described above, we are even able to separate different signals. We could go further an compare the GIA from this analysis with for instance the GIA used to correct GRACE data, however this is out of the scope of this chapter.

As stated at the beginning of this section, we have chosen the CAS1 and the VESL GPS stations because of existing prior knowledge about the geophysical process that took place there. Lenaerts et al. (2013) reported strong accumulation events in 2009 and 2011 in Dronning Maud Land, EA where the VESL GPS stations is located. As we performed the comparison in terms of vertical deformation, the time-varying rates in Fig. 8.2 contain a clear subsidence of the solid Earth as an immediate response to the high accumulation anomaly in both years. This subsidence is detected by all three independent techniques as well as the subsidence at the CAS1 GPS station in 2009 reported by Luthcke et al. (2013).

8.3.4 Global Mean Sea Level Time Series

We analyze GMSL time series[1] over the last 25 years that are derived using a combination of different altimetry products. GMSL time series has a repeat cycle of 10 days, which is a different sampling characteristic compared to daily GPS- or monthly GRACE-observations. Since the time series might contain irregularly spaced data, we fill short gaps with interpolated values. Long gaps are filled with NAN values as for GPS time series. Here, we define a gap to be long if three consecutive measurements are missing (i.e., one month of altimetry observations).

While analyzing the LSA residuals of GMSL time series, a temporally correlated noise is detected. We model this coloured noise as AR-process within the Kalman Filter (Sect. 8.2.3). To get an idea about which AR(p) model is the most appropriate to parameterize the observational noise of the GMSL time series, we compared the loglikelihood values, AIC, and BIC for AR(p) with $p = 1 \ldots 9$. That means that the time series is parameterized using different AR(p), bias along with slope, annual, and semiannual components that are allowed to change in time. The corresponding state space model is solved by Kalman Filter. The AR(5) is determined to be a preferred parameterization for the temporally correlated noise in the GMSL time series, as for this model we get the minimum AIC and BIC, and the maximum logL from all the nine different solutions. Figure 8.4 shows the deterministic slope estimated by commonly used LSA with its formal errors rescaled by the a posteriori variance. Figure 8.4 also contains the time-varying slope. From the time-varying slope we compute mean slope to allow both estimates (from LSA and KF) to be directly compared. As can be seen in Fig. 8.4, the results of two techniques agree very well. The advantage of having derived the time-varying trend for the GMSL is that we can immediately see that the acceleration is not constant over the analyzed time span, since any change in slope term in Fig. 8.4 reflects acceleration. When computing the acceleration between the 2007 and the begin of the time series, we get an insignificant number of 0.04 ± 0.08 mm/y^2, between 2007 and 2015 there is a significant average acceleration of 0.27 ± 0.09 mm/y^2 and over the entire analyzed time period the average acceleration

---

[1]http://sealevel.colorado.edu/content/2018rel1-global-mean-sea-level-time-series-seasonal-signals-retained, last access on 09.07.2018.

Fig. 8.4 Slope estimates in mm/year: the time-varying slope derived using the Kalman Filter (KF) framework (black); the mean slope derived from the KF time-varying slope (red); the slope estimated using the least-squares adjustment with formal LSA errors rescaled by the a posteriori variance (blue). Error bars are 1σ
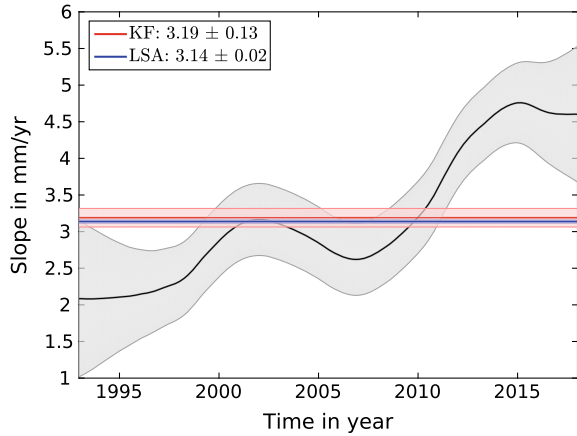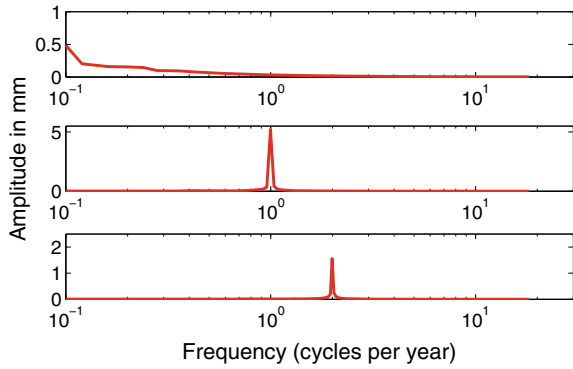

Fig. 8.5 Amplitude spectrums of the estimated slope (top), annual (middle) and semi-annual (bottom) components for the GMSL time series in mm

is estimated to be 0.1 ± 0.06 mm/y^2 (not significant at the 95% confidence level). It should be noted, however, that we utilized the GMSL time series as it is, without removing signals, such as eruption or El Niño Southern Oscillation (ENSO) effects (Nerem et al. 2018), from the time series prior to estimating time-varying rates.

The reliability of the estimated hyperparameters and, in turn, of the different signal constituents is verified using spectral analysis. Figure 8.5 demonstrates that the amplitude spectrums of the estimated slope, annual and semiannual components show significant peaks over the expected frequencies without existing significant peaks elsewhere.

8.4 Conclusions

We estimated time-varying rates from four different time series: GPS, GRACE, SMB, amd GMSL. For each time series, different parameters are estimated. Common to all of them is that along with time-varying rates we also allowed the harmonic

signals to change in time. In this way we avoid the contamination of the time-varying rates by the variability in harmonic terms. The variability of the derived rates, which is governed by hyperparameters, is validated using the inter-comparison of time-varying rates derived from GPS, GRACE and SMB data at the locations of two permanent GPS stations. All three independent techniques capture small-scale accumulation variability present in SMB at these two locations. Such an inter-comparison of time-varying rates that are derived using the described state space framework solved by KF can help decide whether the observed power in the GPS time series at the low frequencies is caused by inaccurately modeled colored noise or is due to geophysical variations. Moreover, any change in the derived time-varying rates reflects an acceleration. The analysis of the GMSL time series over the 25 years suggests the absence of a significant constant acceleration for this time period.

References

Anderssen R, Bloomfield P (1975) Properties of the random search in global optimization. Journal of Optimization Theory and Applications 16(5-6):383–398

Bryson A, Johansen D (1965) Linear filtering for time-varying systems using measurements containing colored noise. IEEE Transactions on Automatic Control 10(1):4–10

Byrd RH, Hribar ME, Jorge N (1999) An interior point algorithm for large-scale nonlinear programming. SIAM J Optim 9(4):877–900

Cheng M, Tapley BD (2004) Variations in the earth's oblateness during the past 28 years. Journal of Geophysical Research: Solid Earth 109(B9)

Davis JL, Wernicke BP, Tamisiea ME (2012) On seasonal signals in geodetic time series. Journal of Geophysical Research 117(B01403)

Didova O, Gunter B, Riva R, Klees R, Roese-Koerner L (2016) An approach for estimating time-variable rates from geodetic time series. Journal of Geodesy 90(11):1207–1221

Dmitrieva K, Segall P, DeMets C (2015) Network-based estimation of time-dependent noise in GPS position time series. Journal of Geodesy 89(6):591–606

Durbin J, Koopman SJ (2012) Time Series Analysis by State Space Methods. Oxford: Oxford University Press

Engels O, Gunter B, Riva R, Klees R (2018) Separating geophysical signals using grace and high-resolution data: a case study in antarctica. Geophysical Research Letters

Farahani HH, Ditmar P, Inácio P, Didova O, Gunter B, Klees R, Guo X, Guo J, Sun Y, Liu X, et al (2016) A high resolution model of linear trend in mass variations from dmt-2: added value of accounting for coloured noise in grace data. Journal of Geodynamics

Farhang-Boroujeny B (1998) Adaptive filters: Theory and applications.

Flechtner F (2007) GRACE 327-750: AOD1B product description document for product release 01 to 04. Tech. rep., GeoFor-schungsZentrum Potsdam, Germany

Harvey AC (1989) Forecasting, structural time series models and the Kalman filter. Cambridge: Cambridge University Press

Harvey AC, Peters S (1990) Estimation procedures for structural time series models. Journal of
    Forecasting 9(2):89–108
Horst R, Pardalos PM, Thoai NV (2000) Introduction to global optimization, nonconvex optimiza-
    tion and its applications, vol. 48
Klemann V, Martinec Z (2011) Contribution of glacial-isostatic adjustment to the geocenter motion.
    Tectonophysics 511(3):99–108
Kusche J, Schrama EJO (2005) Surface mass redistribution inversion from global GPS deformation
    and gravity recovery and climate experiment (GRACE) gravity data. J Geophys Res 110(B09409)
Lenaerts J, Meijgaard E, Broeke MR, Ligtenberg SR, Horwath M, Isaksson E (2013) Recent snowfall
    anomalies in dronning maud land, east antarctica, in a historical and future climate perspective.
    Geophysical Research Letters 40(11):2684–2688
Luthcke SB, Sabaka T, Loomis B, Arendt A, McCarthy J, Camp J (2013) Antarctica, Greenland
    and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution. Journal
    of Glaciology 59(216):613–631
Nerem R, Beckley B, Fasullo J, Hamlington B, Masters D, Mitchum G (2018) Climate-change–
    driven accelerated sea-level rise detected in the altimeter era. Proceedings of the National
    Academy of Sciences p 201717312
Pearson R (2011) Exploring Data in Engineering, the Sciences, and Medicine. Oxford: Oxford
    University Press
Peng JY, Aston JA (2011) The state space models toolbox for MATLAB. Journal of Statistical
    Software 41(6):1–26
Plaszczynski S (2007) Generating long streams of 1/fα noise. Fluctuation and Noise Letters
    7(01):R1–R13
Sneeuw N (1994) Global spherical harmonic analysis by least-squares and numerical quadrature
    methods in historical perspective. Geophysical Journal International 118(3):707–716
Swenson S, Chambers D, Wahr J (2008) Estimating geocenter variations from a combination of
    grace and ocean model output. Journal of Geophysical Research: Solid Earth 113(B8)
Tesmer V, Steigenberger P, van Dam T, Mayer-Gürr T (2011) Vertical deformations from homoge-
    neously processed grace and global gps long-term series. joge 85:291–310
Thomas ID, King MA, Bentley MJ, Whitehouse PL, Penna NT, Williams SDP, Riva REM, Lavallee
    DA, Clarke PJ, King EC, Hindmarsh RCA, Koivula H (2011) Widespread low rates of Antarctic
    glacial isostatic adjustment revealed by GPS observations. Geophysical Research Letters 38(22),
    10.1029/2011GL049277
Williams SD (2003) Offsets in global positioning system time series. Journal of Geophysical
    Research: Solid Earth (1978–2012) 108(B6)
Wouters B, Bamber Já, Van den Broeke M, Lenaerts J, Sasgen I (2013) Limits in detecting accel-
    eration of ice sheet mass loss due to climate variability. Nature Geoscience 6(8):613–616

Olga Engels studied Geodesy and Geoinformation at the University of Bonn, Germany. In 2017,
she defended her PhD thesis "Separating GIA and ice mass change signals in Antarctica using
satellite data" at Delft University of Technology in the Department of Geoscience and Remote
Sensing. She received the Young Author Award 2016 from the International Association of
Geodesy for her paper "An approach for estimating time-variable rates from geodetic time series".
Since March 2017, she is a researcher at the Institute of Geodesy and Geoinformation at the Uni-
versity of Bonn and is working on the assimilation of remote sensing data in a global hydrological
model.