Prediction of ESTSP Competition Time Series by Unscented Kalman Filter and RTS Smoother

Simo Särkkä, Aki Vehtari and Jouko Lampinen

Helsinki University of Technology
Department of Electrical and Communications Engineering
Laboratory of Computational Engineering
Box 9203, FIN-02015 HUT, Finland

Abstract. This article presents a solution to the time series prediction competition of the ESTSP 2007 conference. The solution is based on *optimal filtering*, which is a methodology for computing recursive solutions to statistical inverse problems, where a time varying stochastic state space model is measured through a sequence of noisy measurements. In the solution, the overall behavior of the time series is first modeled by constructing a linear state space model, which captures most of the visible features of the time series. Residual analysis techniques are then used for correcting the yet unmodeled features of the time series. These corrections result in a non-linear state space model, which is solved using a combination of linear Kalman filter, non-linear unscented Kalman filter and Rauch-Tung-Striebel smoother. The unknown parameters of the state space model are optimized to give the best possible prediction over 50 steps.

1 Introduction

In this article, a solution to the ESTSP 2007 time series prediction competition is presented. The solution is based on modeling the time series with a non-linear state space model, which is then estimated with unscented Kalman filter and Rauch-Tung-Striebel smoother. The solution is similar to the time series prediction method used in [1], but the underlying linear stochastic model is different and an additional non-linear correction term is included in the model.

1.1 Optimal Filtering and Smoothing

The celebrated *Kalman filter* [2, 3, 4] considers *optimal filtering*, that is, *recursive statistical inference* of linear state space models of the form

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}_{k-1}\,\mathbf{x}_{k-1} + \mathbf{q}_{k-1} \\ \mathbf{y}_k &= \mathbf{H}_k\,\mathbf{x}_k + \mathbf{r}_k, \end{aligned} \tag{1}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the state, $\mathbf{y}_k \in \mathbb{R}^m$ is the measurement, $\mathbf{q}_{k-1} \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}_{k-1})$ is the process noise, and $\mathbf{r}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{R}_k)$ is the measurement noise. The matrix $\mathbf{A}_{k-1}$ is the transition matrix of the dynamic model and $\mathbf{H}_k$ is the measurement model matrix. In addition to producing the optimal estimates, the Kalman filter can be used for computing optimal n-step prediction of the state space model given a sequence measurements from the model.

The *Extended Kalman filter (EKF)* [3, 5, 6] and *Unscented Kalman filter (UKF)* [7, 8, 5] are extensions of the Kalman filter to estimation and prediction of non-linear state space models of the form

$$\begin{aligned} \mathbf{x}_k &= \mathbf{f}(\mathbf{x}_{k-1}, k-1) + \mathbf{q}_{k-1} \\ \mathbf{y}_k &= \mathbf{h}(\mathbf{x}_k, k) + \mathbf{r}_k, \end{aligned} \tag{2}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the state, $\mathbf{y}_k \in \mathbb{R}^m$ is the measurement, $\mathbf{q}_{k-1} \sim \mathrm{N}(0, \mathbf{Q}_{k-1})$ is the Gaussian process noise, $\mathbf{r}_k \sim \mathrm{N}(0, \mathbf{R}_k)$ is the Gaussian measurement noise, $\mathbf{f}(\cdot)$ is the dynamic model function and $\mathbf{h}(\cdot)$ is the measurement model function.

Optimal smoothing methods [9, 10, 11] can be used for computing better estimates of the signal than optimal filters in the cases that the whole history of measurements from the time series can be used for computing the estimates. As optimal filtering methods produce optimal estimates, which are optimal when only causal estimators are considered, *optimal smoothing methods* produce optimal estimates based on the whole history of observations. Obviously, these estimates can be, in principle, computed from the posterior distribution of the states given the measurements, but the idea of optimal smoothing methods is to provide computationally efficient methods for computing these estimates.

1.2 Stochastic Differential Equations

Stochastic differential equation [12, 13] is a white noise driven differential equation of, for example, the form

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}, t) + \mathbf{L}(t)\,\mathbf{w}(t), \tag{3}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ is the (continuous-time) state, $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}_+ \mapsto \mathbb{R}^n$ is the drift function, $\mathbf{L}(t) \in \mathbb{R}^{n \times s}$ is the dispersion matrix, and $\mathbf{w}(t) \in \mathbb{R}^s$ is a white noise process with spectral density matrix $\mathbf{Q}_c(t)$.

The theory of stochastic differential equations is well known, and it is commonly formulated in terms of *Itô calculus*, which is the theory of differential calculus of stochastic processes (see, e.g., [12, 13]). In rigorous mathematical sense the stochastic differential equation (3) should be actually interpreted as a *stochastic integral equation* of the form

$$\mathbf{x}(t) - \mathbf{x}(s) = \int_s^t \mathbf{f}(\mathbf{x}, t)\,\mathrm{d}t + \int_s^t \mathbf{L}(t)\,\mathrm{d}\boldsymbol{\beta}(t), \tag{4}$$

which can be written more compactly as

$$\mathrm{d}\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}(t)\,\mathrm{d}\boldsymbol{\beta}(t), \tag{5}$$

where $\boldsymbol{\beta}(t)$ is a Brownian motion with diffusion matrix $\mathbf{Q}_c(t)$. If we define the *white noise* $\mathbf{w}(t)$ as the formal derivative of Brownian motion $\mathbf{w}(t) = \mathrm{d}\boldsymbol{\beta}(t)/\mathrm{d}t$, the equation (5) can be formally written in form (3). This kind of white noise formulation only makes sense in the case when the dispersion matrix is independent of the state, that is, $\mathbf{L}(\mathbf{x}, t) = \mathbf{L}(t)$, because it is the case when the Itô and Stratonovich interpretations of the SDE are equivalent.

In this article, the dynamic model will be a *linear time invariant* (LTI) stochastic differential equation of the form

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{F}\,\mathbf{x}(t) + \mathbf{L}\,\mathbf{w}(t), \tag{6}$$

where $\mathbf{x}(t)$ is the state, $\mathbf{F}$ and $\mathbf{L}$ are constant matrices, and $\mathbf{w}(t)$ is a white noise process with a constant spectral density matrix $\mathbf{Q}_c$. The theory of LTI equations is much less complicated than the general Itô calculus and the relevant results of it can be found in beginning Chapters many estimation theory oriented books (e.g., [4, 6]).

2 Time Series Prediction

In this section the models and methods for time series prediction, and the prediction results are described. The model is constructed as follows:

1. First a linear state space model is constructed, which captures the dominant features of the time series.
2. A non-linear correction term is added to the model, which models the unmodeled non-linearity in the periodic signal.
3. The remaining auto-regressive component in the residual is compensated by fitting an AR-model to the residual.

The models are estimated using Kalman filter, unscented Kalman filter and Rauch-Tung-Striebel smoother.

2.1 Linear State Space Model

By looking at the time series data, which is shown in Figure 1, it can be seen to be sum of two main components, *a bias component* and *a periodic component*:

- The bias component can be modeled as a *Brownian motion*, which is formally the integral of continuous-time white noise $w_b(t)$, and it can be formulated as a solution of the stochastic differential equation model:

$$\frac{\mathrm{d}x_b}{\mathrm{d}t} = w_b(t). \tag{7}$$

- The periodic component can be modeled as a resonator with a certain angular velocity ω. The variations from perfect sinusoidal can be modeled by including a random white noise forcing term $w_r(t)$ to the resonator equation. This results in the stochastic differential equation model:

$$\frac{\mathrm{d}^2 x_r}{\mathrm{d}t^2} = -\omega^2\, x_r + w_r(t). \tag{8}$$

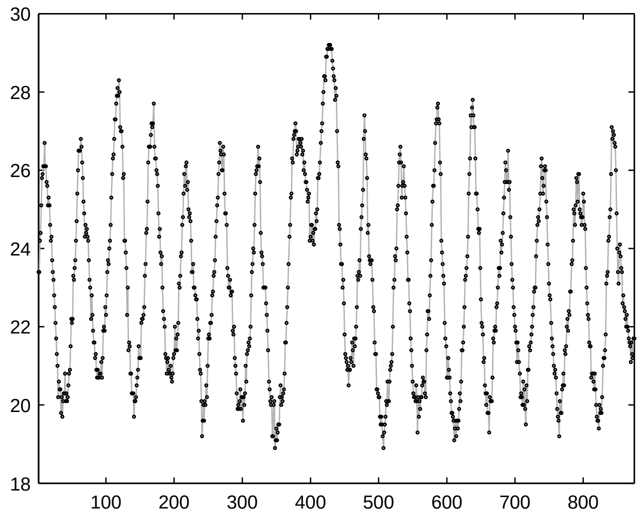

Fig. 1: ESTSP 2007 time series prediction competition data.

The actual time series is the sum of the bias and periodic components. In order to model the deviations of the data from the model a small Gaussian measurement noise r_k is assumed to be present in the measurements:

$$y_k = x_b(k) + x_r(k) + r_k, \tag{9}$$

where $k = 1, 2, \ldots$. The model can be equivalently written in form

$$\underbrace{\begin{pmatrix} \mathrm{d}x_b/\mathrm{d}t \\ \mathrm{d}x_r/\mathrm{d}t \\ \mathrm{d}^2x_r/\mathrm{d}t^2 \end{pmatrix}}_{\mathrm{d}\mathbf{x}/\mathrm{d}t} = \underbrace{\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -\omega^2 & 0 \end{pmatrix}}_{\mathbf{F}} \underbrace{\begin{pmatrix} x_b \\ x_r \\ \mathrm{d}x_r/\mathrm{d}t \end{pmatrix}}_{\mathbf{x}} + \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}}_{\mathbf{L}} \underbrace{\begin{pmatrix} w_b \\ w_r \end{pmatrix}}_{\mathbf{w}} \tag{10}$$

$$y_k = \underbrace{\begin{pmatrix} 1 & 1 & 0 \end{pmatrix}}_{\mathbf{H}} \underbrace{\begin{pmatrix} x_b \\ x_r \\ \mathrm{d}x_r/\mathrm{d}t \end{pmatrix}}_{\mathbf{x}} + r_k. \tag{11}$$

If we define the state as $\mathbf{x} = (x_b\ x_r\ \mathrm{d}x_r/\mathrm{d}t)^T$, the model can be seen to be a linear Gaussian continuous-discrete filtering model:

$$\begin{aligned} \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} &= \mathbf{F}\,\mathbf{x}(t) + \mathbf{L}\,\mathbf{w}(t) \\ y_k &= \mathbf{H}\,\mathbf{x}(k) + r_k. \end{aligned} \tag{12}$$

By the well known formulas for linear systems (see, e.g., [4]) the transition matrix and covariance matrix of the equivalent discrete time model are given as

$$
\begin{aligned}
\mathbf{A} &= \exp(\mathbf{F}) \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & \frac{\sin(\omega)}{\omega} \\ 0 & -\omega\,\sin(\omega) & \cos(\omega) \end{pmatrix} \\
\mathbf{Q} &= \int_0^1 \exp\Big((1-\tau)\,\mathbf{F}\Big)\,\mathbf{L}\,\mathbf{Q}_c\,\mathbf{L}^T\,\exp\Big((1-\tau)\,\mathbf{F}\Big)^T \mathrm{d}\tau \\
&= \begin{pmatrix} q_b & 0 & 0 \\ 0 & \frac{q_r\,\omega - q_r\,\cos(\omega)\,\sin(\omega)}{2\omega^3} & \frac{q_r\,\sin^2(\omega)}{2\omega^2} \\ 0 & \frac{q_r\,\sin^2(\omega)}{2\omega^2} & \frac{q_r\,\omega + q_r\,\cos(\omega)\,\sin(\omega)}{2\omega} \end{pmatrix},
\end{aligned}
\tag{13}
$$

where q_b and q_r are the spectral densities of w_b and w_r, respectively. If we denote the discrete-time state as $\mathbf{x}_k \triangleq \mathbf{x}(k)$, the model can be written as

$$
\mathbf{x}_k = \mathbf{A}\,\mathbf{x}_{k-1} + \mathbf{q}_{k-1} \tag{14}
$$

$$
y_k = \mathbf{H}\,\mathbf{x}_k + r_k, \tag{15}
$$

which is a linear state space model, and suitable for the Kalman filter.

The parameters ω, q_b and q_r can be estimated such that they give the best possible prediction as follows:

1. Find the places in the second half of the time series, which are similar to the place where the time series ends.
2. Select a discrete set of possible parameter values and predict 50 steps a head from the selected time series places.
3. Compute the errors in the predictions and select the parameter values that give the least total error when the predictions at all the selected time series places are summed.

The result of prediction with the estimated parameter values is shown in the Figure 2. The result is the smoothing result, computed by running Kalman filter and Rauch-Tung-Striebel smoother through the data. The estimated bias $x_b(t)$ is also separately shown in the figure. The measurement noise variance is set to an arbitrary value $\sigma_r^2 = 1$, because its value does not affect the prediction.

2.2 Non-Linear Correction Term

Although, the prediction of linear state space model in Figure 2 already captures the essential features of the time series, it is far from perfect. One feature, which can be seen in the time series is that the periodic component does not seem to be perfect sinusoidal, but possibly a non-linear transformation of a sinusoidal. This can be checked by plotting the measurements minus the estimated biases

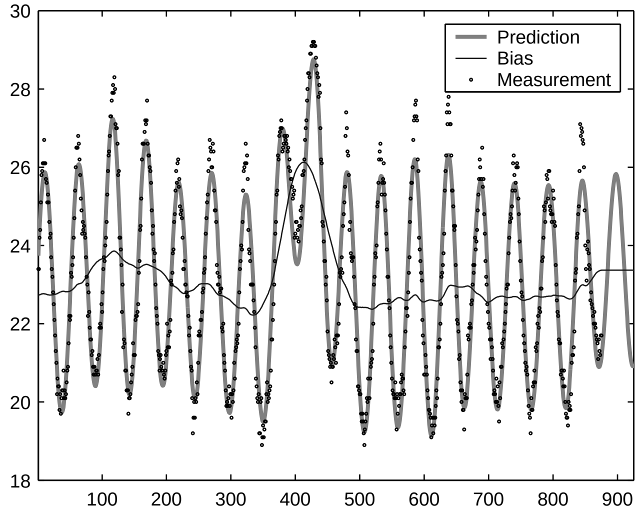

Fig. 2: Linear model smoothing and prediction result.

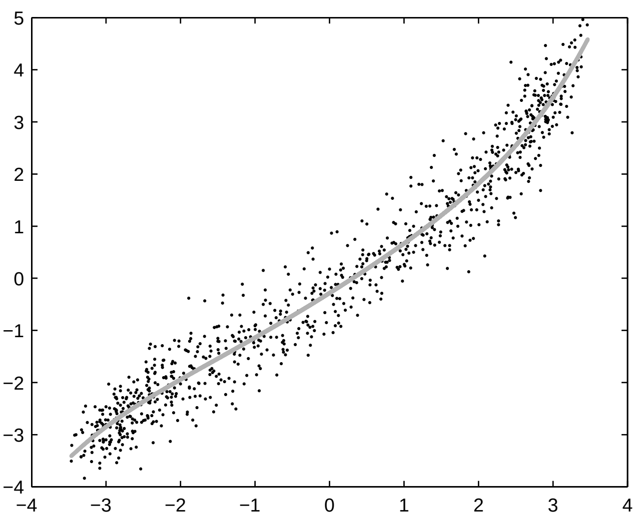

Fig. 3: Measurements minus the biases $y_k - \hat{x}_b$ as function of estimates of the resonator state $\hat{x}_r(k)$ and the fitted polynomial.

$y_k - \hat{x}_b$ as function of estimates of the resonator state $\hat{x}_r(k)$. The result is shown in the Figure 3 together with 5th order polynomial fitted to the function.

The fitted 5th order polynomial can be now included as part of the measurement model as follows:

$$y_k = x_b(r) + \sum_{i=0}^{5} c_i \left(x_r(k) \right)^i + r_k, \tag{16}$$

where c_i are the coefficients of the polynomial. The measurement model is now non-linear, but fortunately of the form (2), which is suitable for the *unscented Kalman filter* (UKF).

In order to get an estimate of the time series based on all the measurements instead of the filtering estimate, which is based on the preceding data, Rauch-Tung-Striebel (RTS) smoother was ran over the UKF filtering result. Note that because the dynamic model is linear, linear RTS smoother is enough and there is no need to use a non-linear smoother.

2.3 Auto-regressive Model for Residual

The non-linear correction term in (16) models the signal-dependence of the residual quite well, but there still exists significant auto-correlation in the residual as can be seem in the Figure 4. There seems to be a small periodic component in the residual, which needs to be compensated.

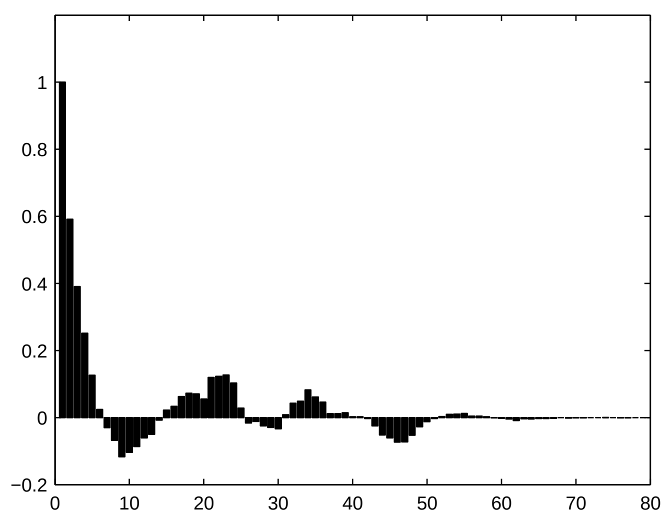

Fig. 4: Auto-correlation of the residual.

The periodicity of the residual time series $\{e_k : k = 1, ..., N\}$ can be modeled with a second order auto-regressive (AR) model [14]

$$e_k = \sum_{i=1}^{2} a_i \, e_{k-i} + r_k^{\mathtt{ar}}. \tag{17}$$

The order of the model was selected to be two, because a second order AR-model is able to capture the single periodic component in the residual and there is no evidence of higher order phenomena. In article [1], an AR-model of the same order was also successfully applied to modeling similar residual periodicity. The variance of the Gaussian noise $r_k^{\mathtt{ar}}$ was set to a suitable value $\sigma_{\mathtt{ar}}^2 = 1$. The coefficients of the AR-model were estimated with the linear least squares method.

After the AR-model has been estimated from the residual time series data, the final estimation solution is obtained by running the Kalman filter and the Rauch-Tung-Striebel smoother to the model

$$\begin{aligned} d_k &= \sum_{i=1}^{2} a_i\, d_{k-i} + v_k^{\mathtt{p}} \\ e_k &= d_k + r_k^p, \end{aligned} \tag{18}$$

where the Gaussian process noise $v_k^{\mathtt{p}}$ has variance $q^{\mathtt{p}} = 10^{-8}$ and the measurement noise r_k^p has the variance $\sigma_p^2 = 10^{-9}$. The residual estimate $\hat{d}_k$ is summed to the prediction computed by the UKF.

2.4 Final Prediction Result

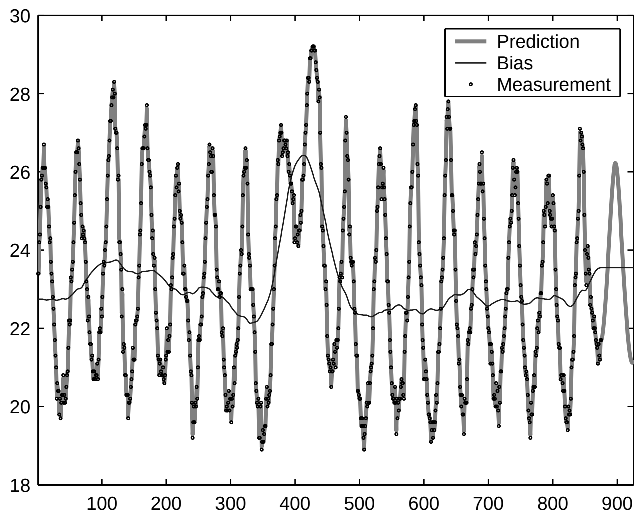

Fig. 5: The final prediction result.

The final prediction result is shown in the Figure 5. The prediction was obtained as follows:

1. A high number of parameter values were selected and the values were searched, which gave the best prediction according to the principle described in Section 2.1.

2. With each parameter values, the non-linear state space model described in Section 2.2 with the linear dynamic model (14) and non-linear measurement model (16) was estimated with *unscented Kalman filter (UKF)* and the prediction was done by iterating the prediction step of the filter 50 times. The non-linear correction in the prediction was applied only to the predicted mean, which corresponds to EKF type of approximation. This was selected, because according to the experiments it gave a lower prediction error than the UKF type of approximation. In the estimation step the UKF worked better than EKF.

3. The Rauch-Tung-Striebel (RTS) smoother was run over the time series to get an estimate, which is conditioned to all the measurements. The auto-regressive model was fitted to the smoother residual and the auto-regressive correction was then added to the prediction results as described in Section 2.3. This estimation was done using linear Kalman filter and RTS smoother.

2.5 Discussion

The approach used in this article is *exploratory* in the sense that a simple linear model is first constructed and it is then improved step by step. The approach is not incremental, because at the final step the whole non-linear model is estimated and used as whole instead of using the results of the models on the previous steps as such.

The Rauch-Tung-Striebel smoother is only used in estimation stage, and actually, it is only needed for computation of the residuals of the solutions. The actual prediction is performed by using the prediction step of the unscented Kalman filter and the non-linear correction is applied to the prediction result. However, the smoother result is needed for estimating the non-linear residual correction term and the AR-model of the residual.

The selection of parameters is based on *predictive criterion*, which tries to mimic the competition error criterion as well as possible. The idea of the approach is to find the parameter values, which minimize the expected value of the prediction error in the competition, when the expected value is computed over the posterior distribution of the other parameters in the model. This can be interpreted as optimal Bayesian decision [15] for selection of point estimates for the parameters. In this case it is not optimal to use, for example, maximum a posterior (MAP) estimate, minimum mean squared error (MMSE) estimate, that is, the posterior mean or any other such point estimates, because they minimize wrong error criteria. Instead, it is best to explicitly minimize the 50 step prediction error criterion of the competition.

3 Conclusion

In this article a solution to the time series prediction competition of the ESTSP 2007 conference has been presented. The solution is based on constructing a non-linear state space model for the time series, which is then estimated using the unscented Kalman filter (UKF) and Rauch-Tung-Striebel (RTS) smoother. The residual auto-correlations were compensated with an AR-model. The parameters were selected by systematically testing various combinations of parameters and by selecting the ones, which gave the least error in 50 step prediction.

References

[1] Simo Särkkä, Aki Vehtari, and Jouko Lampinen. Time series prediction by Kalman smoother with cross-validated noise density. In *Proceedings of the International Joint Conference on Neural Networks*, pages 1653–1658, July 2004.

[2] Rudolph E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME, Journal of Basic Engineering*, 82:35–45, March 1960.

[3] Andrew H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[4] Mohinder S. Grewal and Angus P. Andrews. *Kalman Filtering, Theory and Practice Using MATLAB*. Wiley Interscience, 2001.

[5] Simon Haykin, editor. *Kalman Filtering and Neural Networks*. Wiley, 2001.

[6] Yaakov Bar-Shalom, Xiao-Rong Li, and Thiagalingam Kirubarajan. *Estimation with Applications to Tracking and Navigation*. Wiley Interscience, 2001.

[7] Simon J. Julier, Jeffrey K. Uhlmann, and Hugh F. Durrant-Whyte. A new approach for filtering nonlinear systems. In *Proceedings of the 1995 American Control, Conference, Seattle, Washington*, pages 1628–1632, 1995.

[8] Simon J. Julier and Jeffrey K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, March 2004.

[9] H. E. Rauch, F. Tung, and C. T. Striebel. Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8):1445–1450, 1965.

[10] J.S. Meditch. *Stochastic optimal linear estimation and control*. McGraw-Hill New York, 1969.

[11] Andrew P. Sage and James L. Melsa. *Estimation Theory with Applications to Communications and Control*. McGraw-Hill Book Company, 1971.

[12] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1991.

[13] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, 6 edition, 2003.

[14] Monson H. Hayes. *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, Inc., 1996.

[15] José M. Bernardo and Adrian F. M. Smith. *Bayesian Theory*. John Wiley & Sons, 1994.