Time Series Analysis

Lecture 6: ARIMA models summary
State space models

Tohid Ardeshiri

Linköping University
Division of Statistics and Machine Learning

September 27, 2019

Time domain: The Big Picture

Time Series data

↓

Exploratory data analysis

↓

Make data stationary

↓

Suggest a model M(ϕ, θ)

$\phi = (\phi_1, \ldots, \phi_p)$

$\theta = (\theta_1, \ldots, \theta_q)$

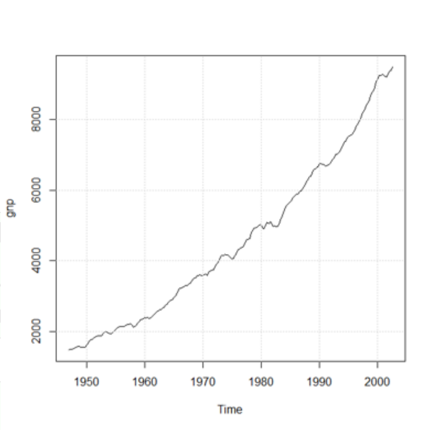
→

- Estimate ϕ's, θ's, ...

Mod

- A
  ar

Prediction



Time domain: The Big Picture

Time Series data

Exploratory data analysis

Make data stationary

Suggest a model $M(\phi, \theta)$

$\phi = (\phi_1, ..., \phi_p)$

$\theta = (\theta_1, ..., \theta_q)$

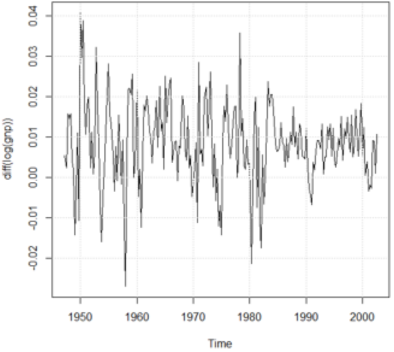
Prediction

Mod...

- A... ar... p

- Estimate ϕ's, θ's, ...

$Y_t = \nabla(\log(X_t))$



Time domain: The Big Picture

Time Series data

Exploratory data analysis

Make data stationary

Suggest a model $M(\phi, \theta)$

$\phi = (\phi_1, ..., \phi_p)$

$\theta = (\theta_1, ..., \theta_q)$

Model Estimation

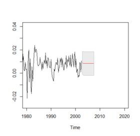
- Estimate $\phi's$, $\theta's$, ...

Model diagnostics and Model selection

- Are there better models with other p and q or even different kind?

no

Prediction



Time domain: The Big Picture

Time Series data

Exploratory data analysis

Make data stationary

Suggest a model M(ϕ, θ)

$\phi = (\phi_1, ..., \phi_p)$

$\theta = (\theta_1, ..., \theta_q)$

Model Estimation

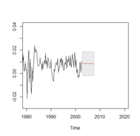
- Estimate $\phi's$, $\theta's$, ...

Model diagnostics and Model selection

- Are there better models with other p and q or even different kind?

no

Prediction



Time domain: The Big Picture

Time Series data

Exploratory data analysis

Make data stationary

Suggest a model M(ϕ, θ)

$\phi = (\phi_1, ..., \phi_p)$

$\theta = (\theta_1, ..., \theta_q)$

Model Estimation

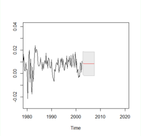
- Estimate $\phi' s$, $\theta' s$, ...

Model diagnostics and Model selection

- Are there better models with other p and q or even different kind?

no

Prediction



Time domain: The Big Picture

Time Series data

Exploratory data analysis

Make data stationary

Suggest a model M(ϕ, θ)

$\phi = (\phi_1, ..., \phi_p)$

$\theta = (\theta_1, ..., \theta_q)$

Model Estimation

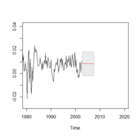
- Estimate $\phi' s$, $\theta' s$, ...

Model diagnostics and Model selection

- Are there better models with other p and q or even different kind?

no

Prediction



Time domain: The Big Picture

Time Series data

Exploratory data analysis

Make data stationary

Suggest a model $M(\phi, \theta)$

$\phi = (\phi_1, ..., \phi_p)$

$\theta = (\theta_1, ..., \theta_q)$

Model Estimation

- Estimate $\phi's$, $\theta's$, ...

Model diagnostics and Model selection

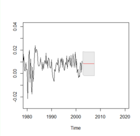
- Are there better models with other p and q or even different kind?

yes

Revise the model

no

Prediction



Model selection

Fit the tentative models, compare them

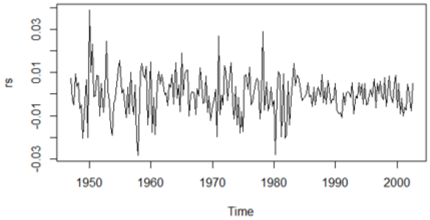
- Analytical measures: AIC, BIC
  - Penalize models with many parameters $\rightarrow$ simpler models
- Residual analysis

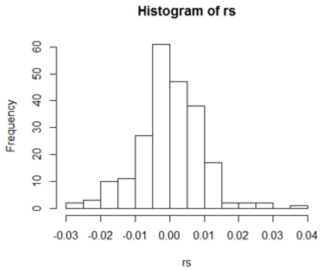
Residual analysis

- Residuals $r_t = x_t - \hat{x}_t^{t-1}$?they are innovations
  - Note: computed from one-step-ahead predictions!
  - Measures predictive quality of the model (compare OLS)
- Residual analysis
  - Visual inspection: stationary? Patterns?
  - Histograms, Q-Q plots
  - ACF, PACF
  - Runs test
  - Box-Ljung test

Residual analysis - Visual inspection

Histogram and visual inspection

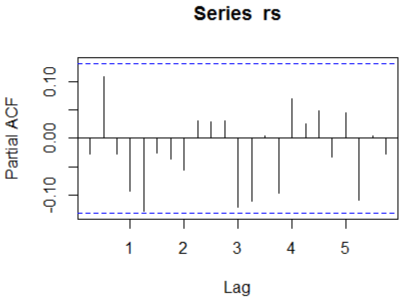


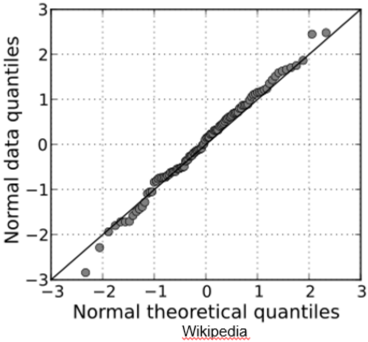


If looks white is good

If looks Normal is good

Residual analysis - ACF /PACF Q-Q plots





If between the blue lines good

If along the diagonal line GOOD

Statistical tests

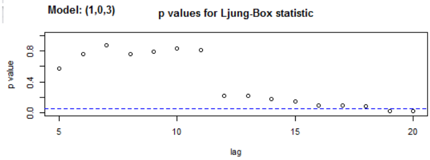
Tests are used to test independence

Runs test

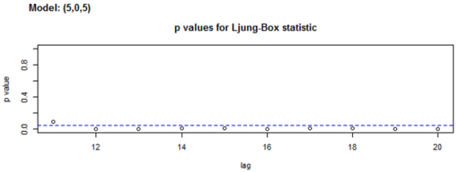
- H_0 : x_t values are i.i.d. **p-value NOT small**
- H_a : x_t values are not i.i.d. **p-value small**

Box-Ljung test

- H_0 : data are independent **p-value NOT small**
- H_a : data are not independent **p-value small**



GOOD



BAD

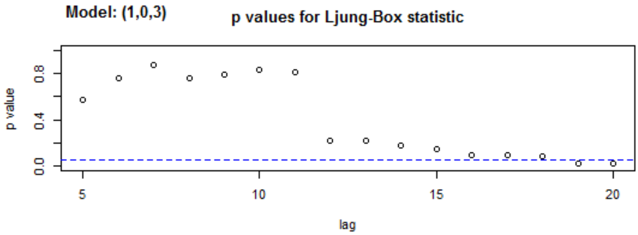
Overfitting

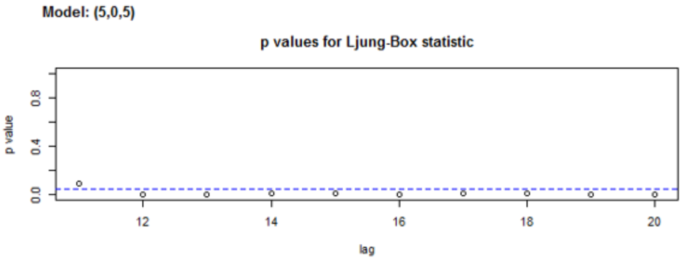
- Occams razor: among equally good models, choose the simplest one
- Overfitting: taking too complex models leads to bad predictions

- If ARIMA(p, d, q) has almost the same predictive quality as ARIMA(p', d', q') , take the one with less parameters

Overfitting

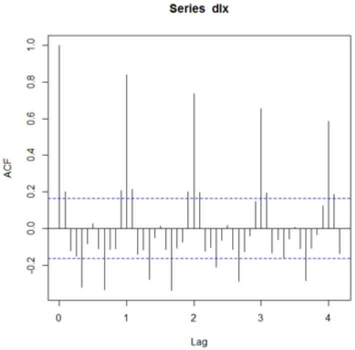
- Example: Recruitment series
  - Fit ARIMA(1,0,3) and ARIMA(5,0,5)

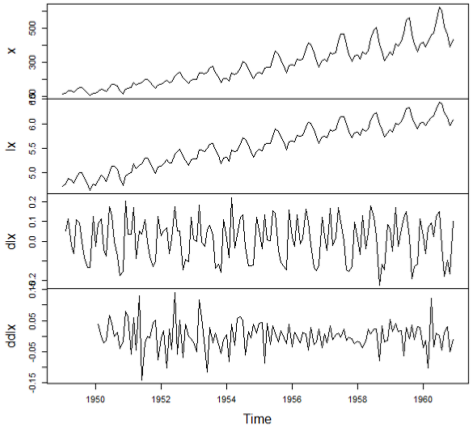




SARIMA - Air passangers

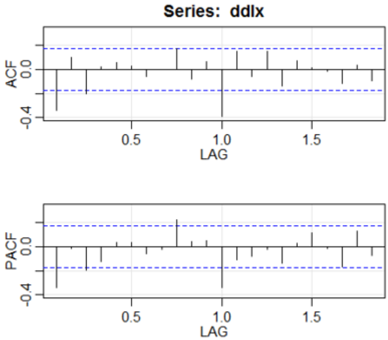
- Example: Air passangers





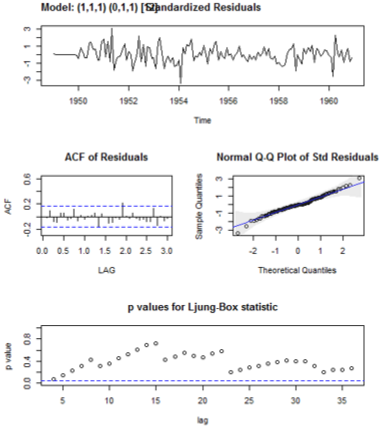
SARIMA - Air passangers

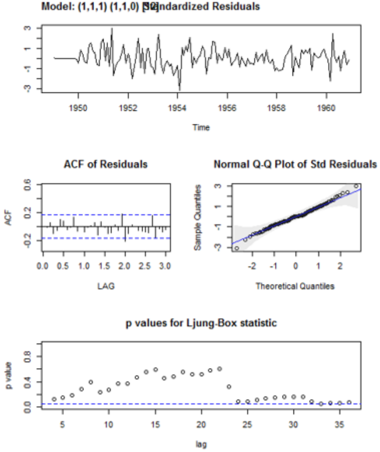
- Example: Air passangers



$\text{ARIMA}(0,1,1)_{12}$ or
$\text{ARIMA}(1,1,0)_{12}$

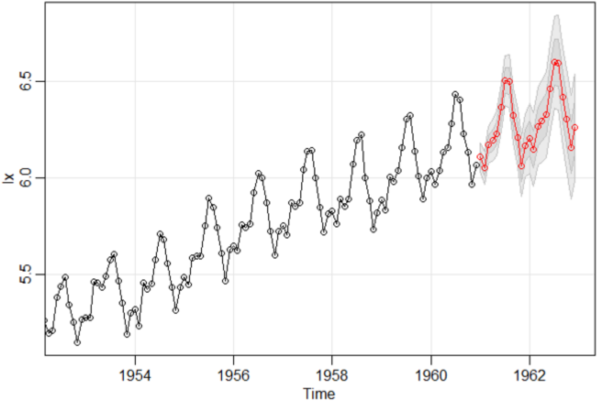
SARIMA - Air passangers





SARIMA

- Forecasting



Read home

- Shumway and Stoffer, Chapter 1, 2 and 3

ARIMA models

Time series models so far

$$\phi^p(B)x_t = \theta^q(B)w_t$$

Model	Concise form
AR(p)	$\phi^p(B)x_t = w_t$
MA(q)	$x_t = \theta^q(B)w_t$
ARMA(p, q)	$\phi^p(B)x_t = \theta^q(B)w_t$
ARIMA(p, d, q)	$\phi^p(B)(1-B)^d x_t = \theta^q(B)w_t$
ARMA$(P, Q)_s$	$\Phi^P(B^s)x_t = \Theta^Q(s)w_t$
ARIMA$(P, D, Q)_s$	$\Phi^P(B^s)(1-B^s)^D x_t = \Theta^Q(B^s)w_t$
ARMA$(p, q) \times (P, Q)_s$	$\Phi^P(B^s)\phi^p(B)x_t = \Theta^Q(B^s)\theta^q(B)w_t$
ARIMA$(p, d, q) \times (P, D, Q)_s$	$\Phi^P(B^s)\phi^p(B)(1-B^s)^D(1-B)^d x_t = \Theta^Q(B^s)\theta^q(B)w_t$

* The notation used in this slide deviates from the notation used in the course literature so far.

Vectorized AR(2) model **Whiteboard**

Consider an AR(2) model

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$$

Let $\mathbf{z}_t = \begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix}$ and $e_t = \begin{bmatrix} w_t \\ 0 \end{bmatrix}$.

Show that we rewrite the AR(2) model in the state space form:

$$\begin{aligned} \mathbf{z}_t &= \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \mathbf{z}_{t-1} + e_t \\ x_t &= \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{z}_t, \end{aligned}$$

ARIMA models in State Space form **Whiteboard**

$$\phi^p(B)x_t = \theta^q(B)w_t$$

Can we rewrite any model of this form as a state space model?

$$\begin{aligned} \mathbf{z}_t &= A\mathbf{z}_{t-1} + e_t, \\ \mathbf{x}_t &= C\mathbf{z}_t + \nu_t, \end{aligned}$$

ARIMA models in State Space form **Whiteboard**

$$\phi^p(B)x_t = \theta^q(B)w_t$$

Outline of the solution:

Let $r = \max(p, q+1)$,
$\phi^r(B) = 1 - \phi_1 B - \cdots - \phi_r B^r$,
$\theta^r(B) = 1 + \theta_1 B - \cdots - \theta_{r-1} B^{r-1}$,
$\phi^r(B)(\theta^r(B))^{-1}x_t = w_t$. Hence, for $z_t = (\theta^r(B))^{-1}x_t$ we can have
$\phi^r(B)z_t = w_t$

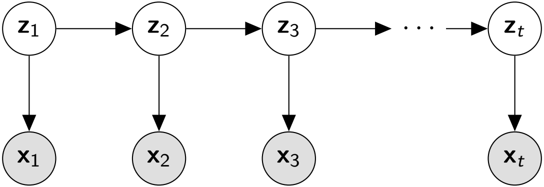
$$\mathbf{z}_t = \begin{bmatrix} z_t \\ z_{t-1} \\ z_{t-2} \\ \vdots \\ z_{t-r+1} \end{bmatrix} \text{ and } \mathbf{z}_t = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_r \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \mathbf{z}_{t-1} + \begin{bmatrix} w_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$x_t = \begin{bmatrix} 1 & \theta_1 & \theta_2 & \cdots & \theta_r \end{bmatrix} \mathbf{z}_t$$

State Space models - graphical models

$$\mathbf{z}_t = A\mathbf{z}_{t-1} + e_t, \qquad e_t \sim f_e(\cdot)$$
$$\mathbf{x}_t = C\mathbf{z}_t + \nu_t, \qquad \nu_t \sim f_\nu(\cdot)$$

A probabilistic graphical model for stochastic dynamical system with latent state $\mathbf{z}_k$ and observations $\mathbf{x}_k$

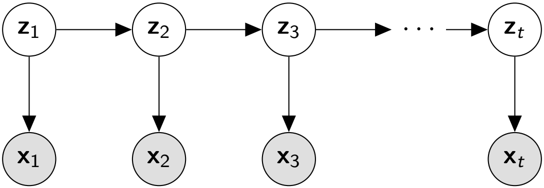


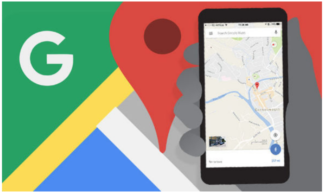
The main tool here is the probability Calculus; Bayes rule and marginalization.

Dynamical systems - more general case

$$\mathbf{z}_t = \mathcal{F}(\mathbf{z}_{t-1}) + e_t, \qquad e_t \sim f_e(\cdot)$$

$$\mathbf{x}_t = \mathcal{C}(\mathbf{z}_t) + \nu_t, \qquad \nu_t \sim f_\nu(\cdot)$$

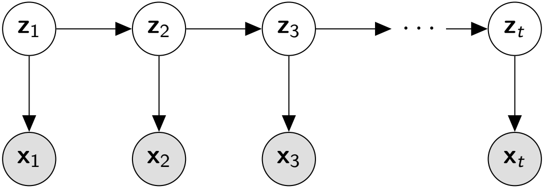




State Space models - Linear and Gaussian

Our main focus will be on linear and Gaussian models:

$$\mathbf{z}_t = A\mathbf{z}_{t-1} + e_t, \qquad e_t \sim N(0, Q)$$
$$\mathbf{x}_t = C\mathbf{z}_t + \nu_t, \qquad \nu_t \sim N(0, R)$$



Bayesian Inference

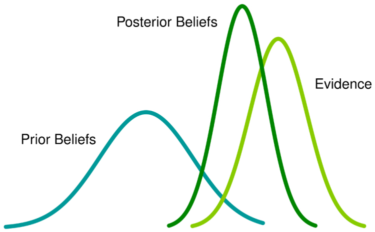
Bayesian inference is a means of combining prior beliefs with the data (evidence) to obtain posterior beliefs.

Example: likelihood update

$$f(\mathbf{z}|\mathbf{x}) \propto f(\mathbf{x}|\mathbf{z})f(\mathbf{z})$$

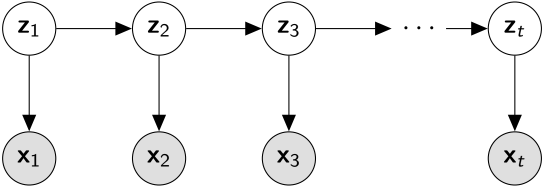
Probability Calculus

$$f(\mathbf{z},\mathbf{x}) = f(\mathbf{z}|\mathbf{x})f(\mathbf{x})$$
$$f(\mathbf{z},\mathbf{x}) = f(\mathbf{x}|\mathbf{z})f(\mathbf{z})$$



Online recursive algorithms

Consider a stochastic dynamical system represented by the following recursion



$$\mathbf{z}_1 \sim f(\mathbf{z}_1), \tag{1a}$$

$$\mathbf{x}_k \sim f(\mathbf{x}_k|\mathbf{z}_k), \tag{1b}$$

$$\mathbf{z}_{k+1} \sim f(\mathbf{z}_{k+1}|\mathbf{z}_k). \tag{1c}$$

The Bayesian filtering recursion corresponds to computing the posterior distributions $f(\mathbf{z}_k|\mathbf{x}_{1:k})$;

$$f(\mathbf{z}_k|\mathbf{x}_{1:k}) = \frac{f(\mathbf{z}_k|\mathbf{x}_{1:k-1})f(\mathbf{x}_k|\mathbf{z}_k)}{\int f(\mathbf{z}_k|\mathbf{x}_{1:k-1})f(\mathbf{x}_k|\mathbf{z}_k)\, \mathrm{d}\mathbf{z}_k}. \tag{2}$$

The density $f(\mathbf{z}_k|\mathbf{x}_{1:k-1})$ in the numerator of (2) which is called the predicted density of $\mathbf{z}_k$ and is obtained by integration as in

$$f(\mathbf{z}_k|\mathbf{x}_{1:k-1}) = \int f(\mathbf{z}_k|\mathbf{z}_{k-1})f(\mathbf{z}_{k-1}|\mathbf{x}_{1:k-1})\, \mathrm{d}\mathbf{z}_{k-1}. \tag{3}$$

Properties of the Normal density function

Property 1: $f(\mathbf{z})f(\mathbf{x}|\mathbf{z}) = f(\mathbf{z}, \mathbf{x})$

$$N(\mathbf{z}; \mu, \Sigma)N(\mathbf{x}; C\mathbf{z}, R) = N\left(\begin{bmatrix}\mathbf{z}\\ \mathbf{x}\end{bmatrix}; \begin{bmatrix}\mu\\ C\mu\end{bmatrix}, \begin{bmatrix}\Sigma & \Sigma C^{\mathrm{T}}\\ C\Sigma & C\Sigma C^{\mathrm{T}} + R\end{bmatrix}\right)$$

Property 2: marginalization and conditioning

If x, y were jointly normal:

$$f(x, y) = N\left(\begin{bmatrix}x\\ y\end{bmatrix}; \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\right)$$

then

$$\begin{aligned}
f(x) &= N(x; \mu_1, \Sigma_{11})\\
f(y) &= N(y; \mu_2, \Sigma_{22})\\
f(x|y) &= N(x; \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\\
f(y|x) &= N(y; \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})
\end{aligned}$$

The Kalman Filter's Foundation

Let $\mathbf{z}$ have a normal prior distribution with mean μ and covariance Σ , i.e., $\mathbf{z} \sim N(\mathbf{z}; \mu, \Sigma)$.

An observation $\mathbf{x}$ with the likelihood function $f(\mathbf{x}|\mathbf{z}) = N(\mathbf{x}; C\mathbf{z}, R)$ is in hand where C is a matrix with proper dimensions and R is a covariance matrix. The posterior distribution of $\mathbf{z}$ can be obtained using the Bayes' rule

$$f(\mathbf{z}|\mathbf{x}) = \frac{f(\mathbf{z})f(\mathbf{x}|\mathbf{z})}{\int f(\mathbf{z})f(\mathbf{x}|\mathbf{z})\,\mathrm{d}\mathbf{z}} \tag{4}$$

$$= \frac{N(\mathbf{z}; \mu, \Sigma)N(\mathbf{x}; C\mathbf{z}, R)}{\int N(\mathbf{z}; \mu, \Sigma)N(\mathbf{x}; C\mathbf{z}, R)\,\mathrm{d}\mathbf{z}}. \tag{5}$$

The posterior distribution $f(\mathbf{z}|\mathbf{x})$ has an analytical solution and turns out to be the normal distribution $N(\mathbf{z}; \mu', \Sigma')$ where

$$\mu' = \mu + K(\mathbf{x} - C\mu), \tag{6a}$$

$$\Sigma' = \Sigma - KC\Sigma, \tag{6b}$$

where

$$K = \Sigma C^{\mathrm{T}}(C\Sigma C^{\mathrm{T}} + R)^{-1}. \tag{7}$$