Bias-correction of Kalman filter estimators associated to a linear state space model with estimated parameters

Marco Costa[a,b], Magda Monteiro[a,b]

[a] *School of Technology and Management of Águeda, University of Aveiro, Apartado 473, 3754 – 909 Águeda, Portugal*

[b] *Center for Research & Development in Mathematics and Applications, Campus Universitário de Santiago 3810-193 Aveiro, Portugal*

Abstract

This paper aims to discuss some practical problems on linear state space models with estimated parameters. While the existing research focuses on the prediction mean square error of the Kalman filter estimators, this work presents some results on bias propagation into both one-step ahead and update estimators, namely, non recursive analytical expressions for them. In particular, it is discussed the impact of the bias in the invariant state space models. The theoretical results presented in this work provide an adaptive correction procedure based on any parameters estimation method (for instance, maximum likelihood or distribution-free estimators). This procedure is applied to two data set: in the calibration of radar precipitation estimates and in the global mean land-ocean temperature index modeling.

Keywords: State space model, Kalman filter, Estimated parameters, Bias-correction, Predictions bias, Environmental data

1. Introduction

State space models have been largely applied in several areas of applied statistics. In particular, the linear state space models have desirable properties and they have a huge potential in time series modeling that incorporates latent processes.

Once a model is placed in the linear state space form, the most usual algorithm to predict the latent process, the state, is the Kalman filter algorithm. This algorithm is a procedure for computing, at each time t ($t = 1, 2, \ldots$), the optimal estimator of the state vector based on the available information until t and its success lies on the fact that is an online estimation procedure. The main goal of the Kalman filter algorithm is to find predictions for the unobservable variables based on observable variables related to each other through a set of equations forming the state space model. Indeed, in the context of linear state space models, the Kalman filter produces the best linear unbiased estimators. When the errors and the initial state are Gaussian, the Kalman filter estimators are the best unbiased estimators in the sense of the minimum mean square error. However, the optimal properties only can be guaranteed when all model's parameters are known (Harvey, 1996). If the model is nonlinear, it must be considered the equation of optimal filtering (Stratonovich, 1960; Dobrovidov et al., 2012). However, as it was proved in Markovich (2015), when the unobservable Markov sequence is defined by a linear equation with a Gaussian noise, the equation of optimal filtering coincides with the classical Kalman filter.

In practice, some or even all model's parameters are unknown and have to be estimated. When the true parameters Θ of the linear state space model are, for instance, substituted by their maximum likelihood ML (or other) estimates, $\widehat{\Theta}$, the theoretical properties of Kalman filter estimators are no longer valid. The usual approach in the analysis of the effects (implications) of applying estimates rather than using true values is to recalculate the mean square errors of both one-step-ahead estimator and update estimator of the unknown state β_t, $P_{t|t-1}$ and $P_{t|t}$, respectively. This approach is discussed in the literature, for instance in Ansley and Kohn (1986) and Hamilton (1986) or more recently in Pfeffermann and Tiller (2005) and it relies on the fact that substituting the model parameters by their estimates in the theoretical mean square error (MSE) expression, that assumes known parameters values, results in underestimation of the true MSE.

Indeed, denoting by $\widehat{\beta}_{t|t}(\widehat{\Theta})$ the optimal filter estimator of β_t based on the observations up to time t substituting Θ by $\widehat{\Theta}$, the MSE of the estimation error is

$$\begin{aligned} \text{MSE}_{t|t} &= \text{E}\left\{\left[\widehat{\beta}_{t|t}(\widehat{\Theta}) - \beta_t\right]\left[\widehat{\beta}_{t|t}(\widehat{\Theta}) - \beta_t\right]'\right\} \\ &= P_{t|t} + \text{E}\left\{\left[\widehat{\beta}_{t|t} - \widehat{\beta}_{t|t}(\widehat{\Theta})\right]\left[\widehat{\beta}_{t|t} - \widehat{\beta}_{t|t}(\widehat{\Theta})\right]'\right\}. \end{aligned}$$

The first term of the sum is the uncertainty contribution of the Kalman filter resulting from the estimation of state when the model parameters are known. The second term reflects the uncertainty due to the estimation of parameters.

Usually, the existent literature investigates methodologies to the second parcel, that is, the contribution to the $\text{MSE}_{t|t}$ resulting from 'parameters uncertainty'. In Hamilton (1986) it is suggested the application of Monte Carlo techniques combining with the ML estimation. From another perspective, Ansley and Kohn (1986) proposed to approximate $P_{t|t}$ by $P_{t|t}(\widehat{\Theta})$ and to expand $\widehat{\beta}_{t|t}(\widehat{\Theta})$ around $\widehat{\beta}_{t|t}$ until the second term. These works were extended in a Bayesian approach in Quenneville and Singh (2000). Wall and Stoffer (2002) proposed a bootstrap procedure for evaluating conditional forecast errors that requires the backward representation of the model. Tsimikas and Ledolter (1994) presented an alternative way to build the restricted likelihood function, also using mixed effects models.

Pfeffermann and Tiller (2005) studied non-parametric and parametric bootstrap methods. Also, a bootstrap approach was adopted in the estimation of the mean squared prediction error of the best linear estimator of nonlinear functions of finitely many future observations in a stationary time series in Bandyopadhyay and Lahiri (2010). Rodríguez and Ruiz (2012) proposed two new bootstrap procedures to obtain MSE of the unobserved states which have better finite sample properties than both bootstraps alternatives and procedures based on the asymptotic approximation of the parameter distribution.

In this work it is investigated the parameters bias propagation into Kalman filter estimators, which results allow proposing an adaptive correction algorithm of Kalman filter estimators bias based on an initial parameters estimates. This procedure allows an improvement in modeling of two relevant applications: the calibration of radar precipitation estimates and in the modeling of the global mean land-ocean temperature index between 1880 and 2013.

2. The state space model

Consider the linear state space model represented by the equations

$$Y_t = H_t\beta_t + e_t \tag{1}$$

$$\beta_t = \mu + \Phi(\beta_{t-1} - \mu) + \varepsilon_t, \tag{2}$$

where Y_t is a $k \times 1$ vector time series of observable variables at time t, which are related with the $m \times 1$ vector of unobservable state variables, β_t, known as the state vector, μ is a $m \times 1$ vector of parameters, Φ is a $m \times m$ transition matrix and the disturbances e_t and ϵ_t are $k \times 1$ and $m \times 1$ vectors, respectively, of serially uncorrelated white noise processes with zero mean and covariance matrices $\Sigma_e = E(e_t e_t')$, $\Sigma_\varepsilon = E(\varepsilon_t \varepsilon_t')$ and $E(e_t \epsilon_s') = \mathbf{0}$ for all t and s. Although the state process $\{\beta_t\}$ is not observable, it is generated by a first-order autoregressive process according to (2), the *transition equation.* All the $k \times m$ matrices H_t are assumed to be known at time $t-1$.

An important class of state space models is given by Gaussian linear state space models when the disturbances e_t and ε_t and the initial state are

Gaussian. The state space model (1)-(2) does not impose any restriction on the stationarity of the state process $\{\beta_t\}$. However, in many applications there is no reason to assume that the state process is not stationary.

When the state process's stationarity is suitable it can be assumed that the state vector β_t is a stationary VAR(1) process with mean $E(\beta_t) = \mu$ and transition matrix Φ with all eigenvalues inside the unit circle, i.e.,

$$|\lambda_i(\Phi)| < 1 \text{ for all } \lambda_i \text{ such that } |\Phi - \lambda_i I| = 0, \tag{3}$$

and with covariance matrix Σ, which is the solution of the equation $\Sigma = \Phi\Sigma\Phi' + \Sigma_\varepsilon$.

Usually, the linear state space models are represented considering a state equation as

$$\beta_t = \Phi\beta_{t-1} + \varepsilon_t$$

or in a simply way taking $\Phi = I$, i.e., considering that the state process $\{\beta_t\}$ is a random walk. However, the state space formulation (1)-(2) is more general since this formulation additionally allows the state to be a nonzero mean stationary process. When the state process $\{\beta_t\}$ is non-stationary the transition equation can be rewritten as $\beta_t = C + \Phi\beta_{t-1} + \varepsilon_t$, where $C = (I - \Phi)\mu$ and the state may be non- stationary VAR(1) process.

2.1. The Kalman filter

The Kalman filter provides optimal unbiased linear one-step-ahead and update estimators of the unobservable state β_t. Briefly, the Kalman filter is an iterative algorithm that produces, at each time t, an estimator of the state vector β_t which is given by the orthogonal projection of the state vector onto the observed variables up to that time.

Let $\widehat{\beta}_{t|t-1}$ denote the estimator of β_t based on the observations $Y_1, Y_2, ..., Y_{t-1}$ and let $P_{t|t-1}$ be its covariance matrix, i.e. $E[(\widehat{\beta}_{t|t-1} - \beta_t)(\widehat{\beta}_{t|t-1} - \beta_t)']$, the MSE matrix. Since the orthogonal projection is a linear estimator, the forecast of the observable vector Y_t is given by $\widehat{Y}_{t|t-1} = H_t\widehat{\beta}_{t|t-1}$.

When, at time t, Y_t is available, the prediction error or *innovation*, $\eta_t = Y_t - \widehat{Y}_{t|t-1}$, is used to update the estimate of β_t (*filtering*) through the equation

$$\widehat{\beta}_{t|t} = \widehat{\beta}_{t|t-1} + K_t\eta_t,$$

where K_t is called the Kalman gain matrix and is given by

$$K_t = P_{t|t-1}H_t'(H_tP_{t|t-1}H_t' + \Sigma_e)^{-1}.$$

Furthermore, the MSE of the updated estimator $\widehat{\beta}_{t|t}$, represented by $P_{t|t}$, verifies the relationship $P_{t|t} = P_{t|t-1} - K_tH_tP_{t|t-1}$. On the other hand, at time t, the forecast for the state vector β_{t+1} is given by the equation

$$\widehat{\beta}_{t+1|t} = \mu + \Phi(\widehat{\beta}_{t|t} - \mu)$$

and its MSE matrix is $P_{t+1|t} = \Phi P_{t|t}\Phi' + \Sigma_\varepsilon$. The Kalman filter algorithm is initialized with $\widehat{\beta}_{1|0}$ and $P_{1|0}$. For more details on Kalman filter algorithm see Harvey (1996) and Shumway and Stoffer (2006).

When the state process is stationary, the Kalman filter algorithm can be initialized considering that initial state vector β_0 has $\widehat{\beta}_{1|0} = \mu$ and a covariance matrix $vec(P_{1|0}) = [I_{m^2} - (\Phi \otimes \Phi)]^{-1} vec(\Sigma_\varepsilon)$, where vec and $\otimes$ are the vec operator and the Kronecker product, respectively. In the non-stationarity case, the initialization of the Kalman filter can be incorporated in the estimation procedure or can be specified in terms of a diffuse or non-informative prior (Harvey, 1996).

2.2. Estimation of the parameters

In practice, the parameters $\Theta = (\mu, \Phi, \Sigma_e, \Sigma_\varepsilon)$ are unknown and they must be estimated. When the disturbances e_t and ε_t are normally distributed the Kalman filter estimators minimizes the MSE when the expectation is taken over all the variables since, in this case, the orthogonal projection coincides with the conditional expectation,

$$\widehat{\beta}_{t|t} = E(\beta_t|Y_t, ...) \text{ and } \widehat{\beta}_{t|t-1} = E(\beta_t|Y_{t-1}, ...). \quad (4)$$

Thus, the conditional mean estimator is the minimum mean square estimator of β_t and it is unbiased in the sense that the expectation of the estimation error is zero (Harvey, 1996). So, it is usually assumed the errors normality in several applications, nevertheless, some authors studied other appropriated methodologies for non-Gaussian errors.

The parameters estimation problem in state space models with non-Gaussian errors was treated in more detail in Carlin et al. (1992) and Shephard and Pitt (1997), which focus on Markov chain Monte Carlo to carry out simulation smoothing and Bayesian posterior analysis of parameters. Furthermore, the works of Alpuim (1999) and Costa and Alpuim (2010) were based on distribution-free estimators. Ng et al. (2013) proposed non-parametric ML estimators of forecast distributions in a general non-Gaussian, non-linear state space setting.

The theoretical properties of the Gaussian ML estimates are very desirable since the distribution assumption is not being significantly violated. Under the assumption of normality, the log-likelihood of a sample $(Y_1, Y_2, ..., Y_n)$ can be written through conditional distributions, yielding

$$\log L(\Theta; Y_1, Y_2, ..., Y_n) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\log(|\Omega_t|) - \frac{1}{2}\sum_{t=1}^{n}\eta_t'\Omega_t^{-1}\eta_t,$$

where

$$\Omega_t = H_t P_{t|t-1} H_t' + \Sigma_e. \tag{5}$$

It is possible to obtain the ML estimates maximizing the log-likelihood function in order to the unknown parameters using numerical algorithms, namely, the EM algorithm (Dempster et al., 1977) or the Newton-Raphson algorithm (Harvey, 1996). An alternative is the optimization algorithm BFGS used in Franco et al. (2008).

3. Bias of the Kalman filter estimators

This section analyzes the bias propagation of the estimates of the model's parameters into the state estimators extending the preliminary work of Monteiro and Costa (2012). The usual approaches focus in the correction of the estimated mean square errors of the Kalman filter estimators, while this work focuses on the Kalman filter estimators bias, i.e., on the point estimation of the Kalman filter estimates.

The state process structure of a VAR(1) associated to the Kalman filter estimators implies that the bias propagation is additive in the μ estimation. This fact allows investigating the propagation of this bias into Kalman filter estimators.

The approach presented in the following sections does not assume any distribution or estimation method to the parameters. These results are based on the linearity of the model and unbiased properties of the Kalman filter estimators.

3.1. Linear propagation bias

Consider a linear state space model (1)-(2) where it is admitted that all parameters are known except the vector μ that is estimated with an error, i.e.,

$$\widehat{\mu} = \mu + \lambda,$$

where λ is the estimation error.

Let $\widehat{Y}_{t|t-1}(\widehat{\Theta})$ be the one-step-ahead forecast of Y_t obtained with $\widehat{\Theta}$ and similarly $\widehat{\beta}_{t|t-1}(\widehat{\Theta})$ and $\widehat{\beta}_{t|t}(\widehat{\Theta})$ for the state estimators.

As the Kalman filter estimators are linear on μ, the estimation error λ of μ will influence them additively, i.e.,

$$\begin{aligned}\widehat{Y}_{t|t-1}(\widehat{\Theta}) &= \widehat{Y}_{t|t-1} + \text{bias}(\widehat{Y}_{t|t-1}(\widehat{\Theta}))\\ \widehat{\beta}_{t|t-1}(\widehat{\Theta}) &= \widehat{\beta}_{t|t-1} + \text{bias}(\widehat{\beta}_{t|t-1}(\widehat{\Theta}))\\ \widehat{\beta}_{t|t}(\widehat{\Theta}) &= \widehat{\beta}_{t|t} + \text{bias}(\widehat{\beta}_{t|t}(\widehat{\Theta})).\end{aligned}$$

If the state process is stationary the starting value $\widehat{\beta}_{1|0}(\widehat{\Theta})$ for the Kalman filter is given by the mean of the unconditional distribution of the state vector. So, in this case we have $\widehat{\beta}_{1|0}(\widehat{\Theta}) = \widehat{\mu} = \mu + \lambda$.

If the state is not stationary we consider $\widehat{\beta}_{1|0}(\widehat{\Theta}) = \widehat{\beta}_{1|0} + \lambda_{\widehat{\beta}_{1|0}(\widehat{\Theta})}$.

The bias induced in forecast of Y_t is given by

$$\begin{aligned}\widehat{Y}_{t|t-1}(\widehat{\Theta}) &= H_t\widehat{\beta}_{t|t-1}(\widehat{\Theta})\\ &= \widehat{Y}_{t|t-1} + H_t\text{bias}(\widehat{\beta}_{t|t-1}(\widehat{\Theta}))\end{aligned}$$

which induces a bias in the filtering stage, namely,

$$\begin{aligned}\widehat{\beta}_{t|t}(\widehat{\Theta}) &= \widehat{\beta}_{t|t-1}(\widehat{\Theta}) + K_t(Y_t - \widehat{Y}_{t|t-1}(\widehat{\Theta}))\\ &= \widehat{\beta}_{t|t} + (I - K_tH_t)\text{bias}(\widehat{\beta}_{t|t-1}(\widehat{\Theta})).\end{aligned}$$

Additionally, the bias of the one-step-ahead forecast has the form

$$\begin{aligned}\widehat{\beta}_{t|t-1}(\widehat{\Theta}) &= \widehat{\mu} + \Phi(\widehat{\beta}_{t|t}(\widehat{\Theta}) - \widehat{\mu}) \\ &= \widehat{\beta}_{t|t-1} + (I-\Phi)\lambda + \Phi \text{bias}(\widehat{\beta}_{t-1|t-1}(\widehat{\Theta})).\end{aligned}$$

In a recursively way, we have,

$$\text{bias}(\widehat{\beta}_{1|0}(\widehat{\Theta})) = \lambda$$

$$\text{bias}(\widehat{\beta}_{t|t}(\widehat{\Theta})) = (I - K_t H_t)\text{bias}(\widehat{\beta}_{t|t-1}(\widehat{\Theta})) \tag{6}$$

$$\text{bias}(\widehat{\beta}_{t|t-1}(\widehat{\Theta})) = (I-\Phi)\lambda + \Phi\text{bias}(\widehat{\beta}_{t-1|t-1}(\widehat{\Theta})) \tag{7}$$

which can be written as

$$\begin{aligned}\text{bias}(\widehat{\beta}_{t|t}(\widehat{\Theta})) &= (I - K_t H_t)(I_m - \Phi)\lambda \\ &\quad + (I - K_t H_t)\Phi\text{bias}(\widehat{\beta}_{t-1|t-1}(\widehat{\Theta}))\end{aligned}$$

$$\begin{aligned}\text{bias}(\widehat{\beta}_{t|t-1}(\widehat{\Theta})) &= (I-\Phi)\lambda + \Phi(I - K_{t-1}H_{t-1}) \times \\ &\quad \times \text{bias}(\widehat{\beta}_{t-1|t-2}(\widehat{\Theta}))\end{aligned}$$

through the application of (7) and (6), respectively.

These equations allow obtaining non-recursive analytical expressions for forecast and filter bias. These results are presented in Proposition 1 under the convention $\sum_{k=1}^{t} u_k = 0$ for $t < 1$ and all u_k.

Proposition 1. *Consider a linear state space model (1)-(2) with* $bias(\widehat{\beta}_{1|0}(\widehat{\Theta})) = \lambda$ *and assume that the remaining parameters are known.*
Then, for $t \geq 2$,

$$
\begin{aligned}
bias(\widehat{\beta}_{t|t-1}(\widehat{\Theta})) &= \left[\left(I + \sum_{k=1}^{t-2} \prod_{i=1}^{k} \Phi\left(I - K_{t-i}H_{t-i}\right)\right) \times \right. \\
&\quad \left. \times \left(I - \Phi\right) + \prod_{i=1}^{t-1} \Phi\left(I - K_{t-i}H_{t-i}\right)\right] \lambda
\end{aligned}
$$

and

$$
\begin{aligned}
bias(\widehat{\beta}_{t|t}(\widehat{\Theta})) &= (I - K_t H_t) \\
&\quad \times \left\{\left[I + \sum_{k=1}^{t-2} \prod_{i=1}^{k} \Phi\left(I - K_{t-i}H_{t-i}\right)\right]\right. \\
&\quad \left. \times \left(I - \Phi\right) + \prod_{i=1}^{t-1} \Phi\left(I - K_{t-i}H_{t-i}\right)\right\} \lambda.
\end{aligned}
$$

All technical details and proofs are given in the Appendix.

This proposition shows that, under the considered conditions, the induced forecast and filter bias are proportional to the vector bias whose proportionality constant is given by the expressions above. However, these expressions can be simplified in the invariant models, i.e., when matrices $H_t = H$ do not depend on time, as follows in the next subsection.

3.2. Invariant linear state space models with a stationary state

Consider an invariant linear state space model with equations (1)-(2), i.e., $H_t = H$ for all t, and that the stationarity condition (3) holds. In this case, the Kalman filter converges to the steady-state Kalman filter rapidly.

Briefly, it means that the sequence $\{P_{t|t-1}\}$ converges to a steady matrix $\overline{P}$ which verifies the Riccati equation, and the sequence $\{K_t\}$ converges to a steady matrix $\overline{K}$, (Harvey, 1996), that verifies the equation

$$\overline{K} = \overline{P}H(H'\overline{P}H + \Sigma_e)^{-1}.$$

The next corollary expresses the Proposition 1 for the steady state of the univariate state space model ($m = 1$). To differentiate clearly the results obtained for the univariate case, the following results are presented using lowercase letters (for example, $H \equiv h$, $\Phi \equiv \phi$, $\Sigma_e \equiv \sigma_e^2$, etc.).

Corollary 1. *The limit of equation of the Proposition 1, when t goes to infinity, is given by*

$$\lim_{t \to +\infty} bias(\widehat{\beta}_{t|t-1}(\widehat{\Theta})) = \frac{(1-\phi)}{1-\phi(1-\overline{k}h)}\lambda$$

and

$$\lim_{t \to +\infty} bias(\widehat{\beta}_{t|t}(\widehat{\Theta})) = \frac{(1-\phi)}{1-\phi(1-\overline{k}h)}(1-\overline{k}h)\lambda.$$

Since in the steady-state

$$\overline{k}h = \frac{\overline{p}h^2}{\overline{p}h^2 + \sigma_e^2},$$

we have $0 < \overline{k}h < 1$. So, it can be concluded that the bias of Kalman filter update estimator are smaller than the one-step ahead bias. When h is large, $\overline{k}h$ is approximately equal to 1, thus, in this case, the update and forecast bias are approximately zero and $\lambda(1-\phi)$, respectively. If h is small, then

$\overline{k}h$ is approximately zero and, in this case, both update and forecast bias are equal to λ. Since bias of the one-step ahead and update estimators are related with bias λ, it is important to find an estimator for it.

4. The bias-correction procedure

In this section it is proposed a procedure which combines the estimation of the bias λ through the Kalman filter recursions with the bias propagation equations obtained in the Proposition 1.

The Kalman filter estimators bias obtained in Proposition 1 can be written as

$$\text{bias}(\widehat{\beta}_{t|t-1}(\widehat{\Theta})) = A_{t-1}(\widehat{\Theta})\lambda$$

and

$$\text{bias}(\widehat{\beta}_{t|t}(\widehat{\Theta})) = B_{t}(\widehat{\Theta})\lambda,$$

where $A_{t-1}(\widehat{\Theta})$ and $B_t(\widehat{\Theta})$ are functions of $\widehat{\Theta}$ at time $t-1$ and t, respectively.

Thus,

$$\begin{aligned}\widehat{\beta}_{t|t-1}(\widehat{\Theta}) - \widehat{\beta}_{t|t}(\widehat{\Theta}) &= \widehat{\beta}_{t|t-1} - \widehat{\beta}_{t|t} + \text{bias}(\widehat{\beta}_{t|t-1}(\widehat{\Theta})) - \\ &\quad -\text{bias}(\widehat{\beta}_{t|t}(\widehat{\Theta}))\end{aligned}$$

by that,

$$\begin{aligned} E[\widehat{\beta}_{t|t-1}(\widehat{\Theta}) - \widehat{\beta}_{t|t}(\widehat{\Theta})] &= E[\widehat{\beta}_{t|t-1} - \widehat{\beta}_{t|t}] \\ &\quad + [A_{t-1}(\widehat{\Theta}) - B_t(\widehat{\Theta})]\lambda. \end{aligned}$$

As the Kalman filter estimators are unbiased in the sense that the expectation of the estimation error is zero, follows that

$$E(\widehat{\beta}_{t|t-1} - \beta_t) = E(\widehat{\beta}_{t|t} - \beta_t) = 0,$$

so,

$$E[\widehat{\beta}_{t|t-1}(\widehat{\Theta}) - \widehat{\beta}_{t|t}(\widehat{\Theta})] = [A_{t-1}(\widehat{\Theta}) - B_t(\widehat{\Theta})]\lambda.$$

On the one hand, the factor $[A_{t-1}(\widehat{\Theta}) - B_t(\widehat{\Theta})]$ depends solely on the vector of parameters estimates. On the other hand, we can drop the expectation operator in $E[\widehat{\beta}_{t|t-1} - \widehat{\beta}_{t|t}]$ which is asymptotically equivalent (Harvey 1996, pp 142), i.e.,

$$E[\widehat{\beta}_{t|t-1}(\widehat{\Theta}) - \widehat{\beta}_{t|t}(\widehat{\Theta})] \approx \widehat{\beta}_{t|t-1}(\widehat{\Theta}) - \widehat{\beta}_{t|t}(\widehat{\Theta}).$$

An estimator $\widehat{\lambda}$ can be obtained through the least squares method, i.e.,

$$\widehat{\lambda} = \left\{ \sum_{t=1}^{n} [A_{t-1}(\widehat{\Theta}) - B_t(\widehat{\Theta})]'[A_{t-1}(\widehat{\Theta})) - B_t(\widehat{\Theta})] \right\}^{-1} \times$$

$$\sum_{t=1}^{n}[A_{t-1}(\widehat{\Theta}) - B_t(\widehat{\Theta})]'(\widehat{\beta}_{t|t-1}(\widehat{\Theta}) - \widehat{\beta}_{t|t}(\widehat{\Theta})). \tag{8}$$

On the one hand, the one-step-ahead forecast and the update estimate of the state have different uncertainties as estimators of β_t. If the state process variability is prevalent over the observation equation variance there are a significant disparity between $\widehat{\beta}_{t|t-1}(\widehat{\Theta})$ and $\widehat{\beta}_{t|t}(\widehat{\Theta})$. On the other hand, if the sample size is not significantly large, the approximation of the expectation to the difference on the state estimates is not a good option. In both cases it is suggested to take the median as a robust measure, i.e.,

$$\widehat{\lambda}_i = \frac{\text{median}\{\widehat{\beta}_{t|t-1}(\widehat{\Theta}) - \widehat{\beta}_{t|t}(\widehat{\Theta})\}_i}{\text{median}\{A_{t-1}(\widehat{\Theta}) - B_t(\widehat{\Theta})\}_i}, \tag{9}$$

where the quotient is defined as element by element of vectors when the state process $\{\beta_t\}$ is multivariate. This approach is recommended having into account its robustness to outliers existence.

When the state process $\{\beta_t\}$ is stationary it can be performed a recursive procedure combining the parameter estimation method and state bias correction until a convergence criteria be satisfied. This procedure allows to correct the remaining parameters simultaneously with the mean bias. However, when the state is a non-stationary process the parameters estimation method indicates $\widehat{\mu} = \mathbf{0}$ since the global mean of $\{\beta_t\}$ does not exist. In this case, the recursive scheme does not make sense and the procedure for correcting the bias is performed a single time.

The proposed procedure of bias correction is implemented by the next algorithm.

Algorithm. Let $(y_1, y_2, ..., y_n)$ be a time series generated by the model (1)-(2) and a small positive value δ.

1. Estimate the parameters by an estimation method and take these estimates as
   $\widehat{\Theta}^{(1)} = (\widehat{\mu}^{(1)}, \widehat{\Phi}^{(1)}, \widehat{\Sigma}_e^{(1)}, \widehat{\Sigma}_\varepsilon^{(1)})$;
2. Let $\widehat{\Theta}^{(i)}$ be the vector of parameters in the iteration i:
   (a) Compute the Kalman filter estimates, $\widehat{\beta}_{t|t-1}$ and $\widehat{\beta}_{t|t}$, by the Kalman filter algorithm with $\widehat{\Theta}^{(i)}$;
   (b) Compute the functions $A(\widehat{\Theta}^{(i)})$ and $B(\widehat{\Theta}^{(i)})$ according to (6) and (7);
   (c) Estimate the bias λ according to the estimator (8) or the estimator (9);
   (d) Re-estimate the vector
   $$\widehat{\mu}^{(i+1)} = \widehat{\mu}^{(i)} + \widehat{\lambda};$$
   (e) Obtain the new estimates $\widehat{\Phi}^{(i+1)}$, $\widehat{\Sigma}_e^{(i+1)}$ and $\widehat{\Sigma}_\varepsilon^{(i+1)}$ using the adopted estimation method;
   (f) Take $\widehat{\Theta}^{(i+1)} = \{\widehat{\mu}^{(i+1)}, \widehat{\Phi}^{(i+1)}, \widehat{\Sigma}_e^{(i+1)}, \widehat{\Sigma}_\varepsilon^{(i+1)}\}$;
   (g) If $\widehat{\Theta}^{(i+1)}$ verifies a convergence condition, for instance
   $$||\widehat{\Theta}^{(i+1)} - \widehat{\Theta}^{(i)}|| < \delta,$$

then $\widehat{\Theta}^* = \widehat{\Theta}^{(i+1)}$, else return to 2. a).

3. Run the Kalman filter algorithm and obtain the corrected Kalman filter estimates $\widehat{\beta}^*_{t|t-1}$ and $\widehat{\beta}^*_{t|t}$ taking into account the parameters $\widehat{\Theta}^*$.

5. Applications

The aim of this section is to present and discuss two applications of the proposed methodology in order to show practical improvements in state space modeling through the bias-correction procedure. The first discusses the case of a state space model with a stationary state and the second explores the non-stationary case.

5.1. Calibration of radar measurements via rain gauge data

Rainfall is a difficult phenomenon to model and predict due to strong spatial and temporal heterogeneity (Bruno et al., 2014). Hourly rainfall data may be provided by both weather radar and rain gauges. However, rain gauges are sparsely distributed on the ground and they provide local measurements whereas radar data are available on a fine grid of pixels (for instance cells with size 2Km×2Km) allowing a spatial estimation of the rainfall. Nevertheless, radar measurements are less accurate then rain gauges estimates. Thus, it is very usual to combined both measurements in order to obtain accurate mean area estimates of the rainfall. One of the most popular approach to combined both estimates is to relate them using state space models. There are many state space formulations used in the literature (Chumchean et al., 2006; Costa and Alpuim, 2011; Leö et al., 2013).

The main idea is to consider that radar measurements (or their transformation) can be calibrated through a state space model based on the rain gauges measurements by a stochastic relation.

Consider G_t and R_t the rain gauges and the radar estimates, respectively, with $t = 1, 2, ..., n$. The radar estimate R_t is the mean area rainfall of the cell where the rain gauge is located. These estimates are related through the state space model

$$\begin{aligned} G_t &= R_t\beta_t + e_t \\ \beta_t &= \mu + \Phi(\beta_{t-1} - \mu) + \varepsilon_t, \end{aligned}$$

where the radar estimate R_t is known and the state β_t, at time t, is a stochastic calibration factor in the sense that it corrects the estimate R_t given the rain gauge's estimate. On the one hand, the observation equation's error e_t can be seen as an error associated to both the rain gauge device and the measurement reading process. On the other hand, the state equation error ε_t is associated to the calibration process variability.

The data analyzed correspond to 24 hours of a storm occurred at April 28, 2000 in the Alenquer River basin in Portugal located around 40Km north of Lisbon. This area has several rain gauges and is under the radar umbrella installed in Cruz do Leão. It is considered the rainfall estimates of both the rain gauge located in Olhalvo location and the respectively radar estimates associated to the cell 2Km×2Km where this rain gauge is situated.

Table 1: Parameters estimates in the iterative procedure.

iteration	$\widehat{\mu}$	$\widehat{\phi}$	$\widehat{\sigma}_{\varepsilon}^2$	$\widehat{\sigma}_e^2 \times 10^{-4}$
1	1.62448	0.24429	0.69714	3.3881
2	1.19880	0.39074	0.79139	2.4127
3	1.21813	0.38034	0.78542	2.4234
4	1.21437	0.38235	0.78658	2.4211
5	1.21508	0.38197	0.78636	2.4215
6	1.21494	0.38204	0.78640	2.4214
7	1.21497	0.38204	0.78640	2.4214
8	1.21497	0.38204	0.78640	2.4214

Table 2: Estimate of the mean bias and the convergence criterion in the iterative procedure.

iteration	$\widehat{\lambda}$	$\|\|\widehat{\Theta}^{(i)} - \widehat{\Theta}^{(i-1)}\|\|$
2	$4.26{\times}10^{-1}$	$4.60{\times}10^{-1}$
3	$-1.93{\times}10^{-2}$	$2.27{\times}10^{-2}$
4	$3.76{\times}10^{-3}$	$4.42{\times}10^{-3}$
5	$-7.15{\times}10^{-4}$	$8.41{\times}10^{-4}$
6	$1.37{\times}10^{-4}$	$1.61{\times}10^{-4}$
7	$-2.63{\times}10^{-5}$	$2.63{\times}10^{-5}$
8	$1.80{\times}10^{-8}$	$1.80{\times}10^{-8}$

Due to small sample dimension, Table 1 presents the parameters estimates obtained in the iteration procedure considered the estimator (9). The adopted parameter estimation method was the ML considering Gaussian disturbances. The estimation method fitted a stationary AR(1) to the calibration factor, as in other works in this scope (Brown et al., 2001). However, the first bias estimate was 0.426, approximately 26% of the initial estimate of μ. After eight iterations the norm $||\widehat{\Theta}^{(i)} - \widehat{\Theta}^{(i-1)}||$ is less then 10^{-7} and the bias estimate is close to 10^{-8} (see Table 2).

Table 3: Mean square errors of radar calibrate estimates using both forecast and filtered calibration factors.

	$\text{MSE}_{t\|t-1}$	$\text{MSE}_{t\|t}$
ML	0.6394	2.012×10^{-5}
with correction	0.5550	8.997×10^{-6}
variation	-13.21%	-55.28%

The assessment of the methodology's performance can be done, in each context, through various appropriate indicators. In this case, model's adjustment is assessed by the ability to calibrate the radar observations by the one-step-ahead forecasts $\widehat{\beta}_{t|t-1}(\widehat{\Theta})$ or by the update estimates $\widehat{\beta}_{t|t}(\widehat{\Theta})$. Thus, we considered the following measures

$$\text{MSE}_{t|t-1} = \frac{1}{n}\sum_{t=1}^{n}(G_t - R_t\widehat{\beta}_{t|t-1})^2 \tag{10}$$

and

$$\text{MSE}_{t|t} = \frac{1}{n}\sum_{t=1}^{n}(G_t - R_t\widehat{\beta}_{t|t})^2. \tag{11}$$

Table 3 shows the model's performance measures with the ML estimates and after the bias-correction procedure. The proposed approach allows a reduction of the 13.21% and 55.28% of the $\text{MSE}_{t|t-1}$ and $\text{MSE}_{t|t}$, respectively. The correction procedure had more impact proportionally in the reduction of the mean square errors associated with de radar calibration when the update estimates are used. Figure 1 presents the accumulated precipitation

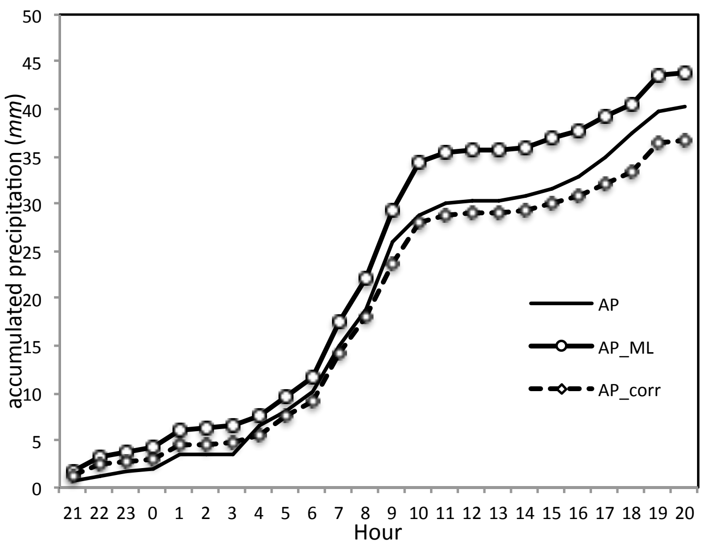

Figure 1: Accumulated precipitation during the storm (AP – based on rain gauge data; AP_ML – estimated with the Gaussian ML; AP_corr– estimated with the corrected parameters).

during the storm considering both non-corrected and corrected parameters estimates. The results show that the corrected parameters produce an accumulated precipitation up to each hour closest to the rain gauge data, which are assumed more accurate. However, as indicated by Corollary 1, in absolute value the correction is greater in the one-step-ahead forecasts.

5.2. Modeling the global mean land-ocean temperature index

The proposed methodology was applied to the global mean land-ocean temperature index, 1880 to 2013, with the base period 1951-1980. Data set is available in the Surface Temperature Analysis (GISTEMP) in the site of the NASA Goddard Institute for Space Studies (GISS). The available data on the global surface temperature are the combination of various data sources

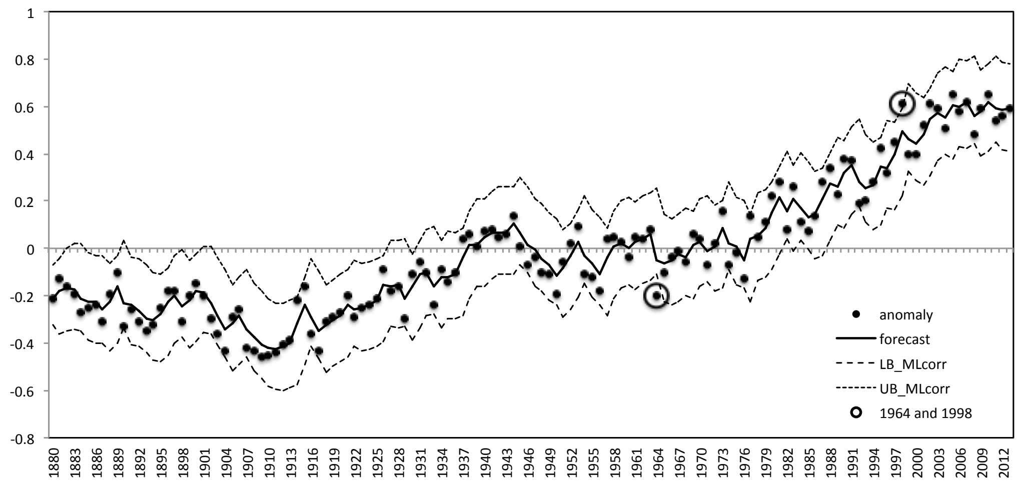

Figure 2: Anomalies, one-step-ahead forecasts and the respective empirical confidence levels at 95% for the bias-corrected case.

(data over land, satellite measurements of sea surface temperature (SST) since 1982, and a ship-based analysis for earlier years). Error sources include incomplete station coverage, quantified by sampling a model-generated data set with realistic variability at actual station locations, and partly subjective estimates of data quality problems (Hansen et al., 2006). The temporal correlation is a relevant feature of the environmental data and it has real impact on data modeling (Alpuim and El-Shaarawi, 2009).

Let Y_t be the global mean temperature anomaly (°C) in the year $t = 1880, ..., 2013$ which is modeled by the equations:

$$\begin{aligned} Y_t &= \beta_t + e_t \\ \beta_t &= \mu + \phi(\beta_{t-1} - \mu) + \varepsilon_t. \end{aligned}$$

Table 4: Parameters estimation.

	$\widehat{\mu}$	$\widehat{\phi}$	$\widehat{\sigma}_{\varepsilon}^2 \times 10^{-3}$	$\widehat{\sigma}_{e}^2 \times 10^{-3}$
ML	0	1.00296	1.763	4.878
corrected	-3.390	—	—	—

The model's adjustment confirms the empirical analysis from Figure 2 that the anomaly is a non-stationary process. In fact, the Gaussian ML estimation produces an estimate for ϕ greater then one. Thus, the correction procedure was applied only one time because, in this case, the method will not converge considering a stationary representation. In practice, it implies that the correction process focuses only in the bias of the state estimates keeping the other parameters unchanged.

Table 4 presents the ML estimates of the parameters and the bias-correction according to the algorithm and the application of the equation (9) once data are non-stationary. The ML estimates induce the state equation $\beta_t = \phi\beta_{t-1} + \varepsilon_t$ and the bias-correction procedure indicates the introduction of the constant $\widehat{\mu}(1-\widehat{\phi}) = 0.010028$°C in the model. So, the correction procedure suggests that the state equation error has a non-zero mean of 0.010028.

Although the state process is not stationary, the Kalman filter enters in a steady state very quickly since in the Kalman filter $p_{t|t-1} \to \bar{p}$ and $k_t \to \bar{k}$. Therefore, the limits of Corollary 1 were achieved. The limit forecast bias is 0.0224°C and the limit filtered bias is 0.01239°C in each year.

Thus, this procedure allows estimating these three types of bias: the bias $\widehat{\mu}(1-\widehat{\phi})$ suggests that this constant can be viewed as the mean of

Table 5: Mean square errors of both one-step-ahead estimates and update estimates of the anomalies considering the Gaussian ML estimators and with the bias correction.

	$\mathrm{MSE}_{t\|t-1}$	$\mathrm{MSE}_{t\|t}$
ML	8.757×10^{-3}	2.679×10^{-3}
ML corrected	8.660×10^{-3}	2.646×10^{-3}
variation (%)	-1.114%	-1.208%

the state equation error and it is induced directly by the uncertainty of the parameter estimation; the value in the middle is the update estimate bias which accommodates both parameters uncertainty and the state uncertainty when Y_t is known; the greatest bias, as expected, is the forecast prediction bias since it is based on the observation Y_{t-1} and incorporates the observation equation uncertainty.

The model's adjustment performance was assessed by both measures (10) and (11), which results are presented in Table 5. On the one hand, performance measures show that the bias correction procedure allows a reduction of the MSE in both performance measures. On the other hand, the models performance can be assessed by empirical confidence intervals of the one-step-ahead forecasts at 95%, i.e.,

$$\widehat{Y}_{t|t-1} \pm 1.96\sqrt{\widehat{\Omega}_t},$$

where $\widehat{\Omega}_t$ is the MSE of $\widehat{Y}_{t|t-1}$ obtained in the Kalman filter recursions (5). Considering the Gaussian ML estimates with no correction four observations are outside the respective empirical confidence interval (in the years 1914, 1964, 1977 and 1998). With the bias correction only two observations are outside of the respective empirical confidence interval (in the years of 1964 e 1998). The most relevant in this comparison is that the performance's improvement of empirical confidence intervals is due solely to the bias correction procedure since the amplitude of these intervals remained unchanged. Figure 2 shows the anomalies, one-step-ahead forecasts and their respective empirical confidence levels at 95% for the bias corrected case.

6. Discussion

This work proposed a bias-correction procedure of the Kalman filter estimators associated to a state space model with estimated parameters. The analysis of the bias propagation of the constant term of the state equation allows determining analytical expressions to both Kalman filter estimators. These results were obtained for a general state space model and particularly analyzed in the invariant models and in the stationary process case. Theoretical results allowed to design a procedure that corrects the initial parameters estimates in order to improve the Kalman filter estimates accuracy. Applications showed that this approach can improve the adjustment of state space models and to enhance analyses of interest in data application's context.

Acknowledgements

Authors were partially supported by Portuguese funds through the CIDMA - Center for Research and Development in Mathematics and Applications, and the Portuguese Foundation for Science and Technology ("FCT– Fundação para a Ciência e a Tecnologia"), within project UID/MAT/04106/2013.

Appendix

Proof of the Proposition 1

The proof is given by the mathematical induction method. It can be seen that bias($\hat{\beta}_{2|1}$) verifies the expression of Proposition 1 through the application of (8) and the convention adopted, i.e.

$$
\begin{aligned}
\text{bias}(\widehat{\beta}_{2|1}(\widehat{\Theta})) &= (I-\Phi)\lambda + \Phi(I-K_1H_1)\text{bias}(\widehat{\beta}_{1|0}(\widehat{\Theta})) \\
&= (I-\Phi)\lambda + \Phi(I-K_1H_1)\lambda \\
&= \left[\left(I + \sum_{k=1}^{0}\prod_{i=1}^{k}\Phi\left(I-K_{t-i}H_{t-i}\right)\right)(I-\Phi)\right. \\
&\quad \left. + \prod_{i=1}^{1}\Phi\left(I-K_{t-i}H_{t-i}\right)\right]\lambda.
\end{aligned}
$$

Consider now that the expression is valid for all instants up to time t. Therefore, at time $t+1$, applying (8) we have

$$
\text{bias}(\widehat{\beta}_{t+1|t}(\widehat{\Theta})) = (I-\Phi)\lambda + \Phi(I-K_tH_t)\text{bias}(\hat{\beta}_{t|t-1}(\widehat{\Theta})).
$$

Under the induction hypothesis it becomes

$$
\text{bias}(\widehat{\beta}_{t+1|t}(\widehat{\Theta})) = (I-\Phi)\lambda + \Phi(I-K_tH_t)\times
$$

$$
\times \left[\left(I + \sum_{k=1}^{t-2} \prod_{i=1}^{k} \Phi \left(I - K_{t-i} H_{t-i} \right) \right) (I - \Phi) + \prod_{i=1}^{t-1} \Phi \left(I - K_{t-i} H_{t-i} \right) \right] \lambda
$$

$$
\begin{aligned}
\text{bias}(\widehat{\beta}_{t+1|t}(\widehat{\Theta})) &= \\
&= (I - \Phi)\lambda + \left[\left(\Phi(I - K_t H_t) + \sum_{k=1}^{t-2} \prod_{i=0}^{k} \Phi \left(I - K_{t-i} H_{t-i} \right) \right) \times \right. \\
&\quad \left. \times (I - \Phi) + \prod_{i=0}^{t-1} \Phi \left(I - K_{t-i} H_{t-i} \right) \right] \lambda \\
&= \left[I + \left(\sum_{k=0}^{t-2} \prod_{i=0}^{k} \Phi \left(I - K_{t-i} H_{t-i} \right) \right) (I - \Phi) \right. \\
&\quad \left. + \prod_{i=0}^{t-1} \Phi \left(I - K_{t-i} H_{t-i} \right) \right] \lambda.
\end{aligned}
$$

The final result is obtained through a change of variable in the summation and product operators.

The proof of the result to $\text{bias}(\widehat{\beta}_{t|t}(\widehat{\Theta}))$ follows applying (6) and the result to $\text{bias}(\widehat{\beta}_{t|t-1}(\widehat{\Theta}))$.

References

Alpuim, T.,1999. Noise variance estimators in state space models based on the method of moments. Annales de l'I.S.U.P. XXXXIII, 3–23.

Alpuim, T., El-Shaarawi, A., 2009. Modeling monthly temperature data in Lisbon and Prague. Environmetrics 20, 835–852.

Ansley, C.F., Kohn, R., 1986. Prediction mean squared error for state state space models with estimated parameters. Biometrika 73, 467–473.

Bandyopadhyay, S., Lahiri, S.N., 2010. Resampling-based bias-corrected time series prediction. J. Stat. Plan. Inference 140, 3775–3788.

Brown, M., Diggle, P., Lord, M., Young, P., 2001. Space-Time Calibration of Radar Rainfall Data. J. Roy. Stat. Soc. C-App. 50, 221–241.

Bruno, F., Cocchi, D., Greco, F., Scardovi, E., 2014. Spatial reconstruction of rainfall fields from rain gauge and radar data. Stoch. Env. Res. Risk. A. 28, 1235–1245.

Carlin, B.P., Polson, N.G., Stoffer, D., 1992. A Monte Carlo approach to nonnormal and nonlinear state-space modelling. J. Am. Statist. Assoc. 87, 493–500.

Chumchean, S., Seed, A., Sharma, A., 2006. Correcting of real-time radar rainfall bias using a Kalman filtering approach. J. Hydrol. 317, 123–137.

Costa, M., Alpuim, T., 2010. Parameter estimation of state space models for univariate observations. J. Stat. Plan. Inference 140, 1889–1902.

Costa, M., Alpuim, T., 2011. Adjustment of state space models in view of area rainfall estimation. Environmetrics 22, 530–540.

Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. J. Roy. Stat. Soc. B Met. 39, 1–38.

Dobrovidov, A.V., Koshkin, G.M., Vasiliev, V.A. 2012. *Non-Parametric state space models*, Kendrick Press, Heber City, UT.

Franco, G. C., Santos, T. R., Ribeiro, J. A., Cruz, F. R. B., 2008. Confidence intervals for hyperparameters in structural models. Commun. Stat. Simulat. 37, 486–497.

Hamilton, J. D., 1986. A standard error for the estimated state vector of a state-space model. J. Economet. 33, 387–397.

Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D.W., Medina-Elizade, M., 2006. Global temperature change. In *Proceedings of the national Academy of Sciences of United States of American* 103 (39), 14288–14293.

Harvey, A.C., 1996. *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Leö, R., Mikkelsen, P. S., Rasmussenand, R. M., Madsen, H., 2013. State-space adjustment of radar rainfall and skill score evaluation of stochastic volume forecasts in urban drainage systems. Water Sci. Technol. 68, 584–590.

Markovich, L. A., 2015. Inferences from optimal filtering equation. Lith. Math. J. 55(3), 413–432.

Monteiro, M., Costa, M., 2012. A note on prediction bias for state space models with estimated parameters. In *Proceeding of the ICNAAM2012, AIP Conference Proceedings* 1479, 2090–2093.

Ng, J., Forbes, C.S., Martin, G.M., McCabe, B., 2013. Non-parametric estimation of forecast distributions in non-Gaussian, non-linear state space models. Int. J. Forecasting 29, 411–430.

Pfeffermann, D., Tiller, R., 2005. Bootstrap approximation to prediction MSE for state space-models with estimated parameters. J. Time Ser. Anal. 21, 219–236.

Quenneville, B., Singh, A.C., 2000. Bayesian prediction mean squared error for state space models with estimated parameters. J. Time Ser. Anal. 26, 893–916.

Rodríguez, A., Ruiz, E., 2012. Bootstrap prediction mean squared errors of unobserved states based on the Kalman filter with estimated parameters. Comput. Stat. Data Anal. 56, 62–74.

Shephard, N., Pitt, M.K., 1997. Likelihood Analysis of Non-Gaussian Measurement Time Series. Biometrika 84, 653–667.

Shumway, R.H., Stoffer, D.S., 2006. *Time series analysis and its applications: with R examples*, Springer, Cambridge.

Stratonovich, R. L., 1960. Conditional Markov processes. Theory Probab. Appl. 5, 156–178.

Tsimikas, J., Ledolter, J., 1994. REML and best linear unbiased prediction in state space models. Commun. Stat. Theory 23, 2253–2268.

Wall, K. D., Stoffer, D. S., 2002. A state space approach to bootstrapping conditional forecasts in ARMA models. J. Time Ser. Anal. 23, 733–751.