Article

Adaptive Conditional Bias-Penalized Kalman Filter for Improved Estimation of Extremes and Its Approximation for Reduced Computation

Haojing Shen [1],*, Haksu Lee [2] and Dong-Jun Seo [1]

[1] Department of Civil Engineering, The University of Texas at Arlington, Arlington, TX 76019, USA; djseo@uta.edu
[2] Len Technologies, Oak Hill, VA 20171, USA; haksulee@hanmail.net
* Correspondence: haojing.shen@mavs.uta.edu; Tel.: +1-817-272-5063; Fax: +1-817-272-2630

Abstract: Kalman filter (KF) and its variants and extensions are wildly used for hydrologic prediction in environmental science and engineering. In many data assimilation applications of Kalman filter (KF) and its variants and extensions, accurate estimation of extreme states is often of great importance. When the observations used are uncertain, however, KF suffers from conditional bias (CB) which results in consistent under- and overestimation of extremes in the right and left tails, respectively. Recently, CB-penalized KF, or CBPKF, has been developed to address CB. In this paper, we present an alternative formulation based on variance-inflated KF to reduce computation and algorithmic complexity, and describe adaptive implementation to improve unconditional performance. For theoretical basis and context, we also provide a complete self-contained description of CB-penalized Fisher-like estimation and CBPKF. The results from one-dimensional synthetic experiments for a linear system with varying degrees of nonstationarity show that adaptive CBPKF reduces the root-mean-square error at the extreme tail ends by 20 to 30% over KF while performing comparably to KF in the unconditional sense. The alternative formulation is found to approximate the original formulation very closely while reducing computing time to 1.5 to 3.5 times of that for KF depending on the dimensionality of the problem. Hence, adaptive CBPKF offers a significant addition to the dynamic filtering methods for general application in data assimilation when the accurate estimation of extremes is of importance.

Keywords: state estimation; extremes; conditional bias; Kalman filter; adaptive filtering

1. Introduction

Streamflow prediction is subject to uncertainties from multiple sources. These include uncertainties in the input (e.g., mean areal precipitation (MAP) and mean areal potential evapotranspiration (MAPE)), hydrologic model uncertainties, and the uncertainties in the initial conditions. With climate change and rapid urbanization, the uncertainty in the input and parametric uncertainties will increase; thus, reducing them has become increasingly challenging. Recently, data assimilation is widely used to reduce the uncertainty in initial conditions [1,2].

In many data assimilation applications in hydrologic predictions, Kalman filter (KF) and its variants and extensions are widely used to fuse observations with model predictions in a wide range of applications [3–23]. In geophysics and environmental science and engineering, often the main objective of data assimilation is to improve the estimation and prediction of states in their extremes rather than in normal ranges. In hydrologic forecasting, for example, the accurate prediction of floods and droughts is far more important than that of streamflow and soil moisture in normal conditions. Because KF minimizes unconditional error variance, its solution tends to improve the estimation near median where the state of the dynamic system resides most of the times while often leaving significant biases in

the extremes. Such conditional biases (CB) [24] generally result in consistent under- and overestimation of the true states in the upper and lower tails of the distribution, respectively. To address CB, CB-penalized Fisher-like estimation and CB-penalized KF (CBPKF) [25] have recently been developed which jointly minimize error variance and expectation of the Type-II CB squared for the improved estimation and prediction of extremes. The Type-II CB, defined as $E[\hat{X}|X = x] - x$, is associated with failure to detect the event where x denotes the realization of X where X, $\hat{X}$ and $\hat{x}$ denote the unknown truth, the estimate, and the realization of $\hat{X}$, respectively [26]. The original formulation of CBPKF, however, is computationally extremely expensive for high-dimensional problems. Additionally, whereas CBPKF improves performance in the tails, it deteriorates performance in the normal ranges. In this work, we approximate CBPKF with forecast error covariance-inflated KF, referred to hereafter as the variance-inflated KF (VIKF) formulation, as a computationally less expensive and algorithmically simpler alternative, and implement adaptive CBPKF to improve unconditional performance.

Elements of CB-penalized Fisher-like estimation have been described in the forms of CB-penalized indicator cokriging for fusion of predicted streamflow from multiple models and observed streamflow [27], CB-penalized kriging for spatial estimation [28] and rainfall estimation [29], and CB-penalized cokriging for fusion of radar rainfall and rain gauge data [30]. The original formulation of CBPKF have been described in [25], respectively. Its ensemble extension, CB-penalized ensemble KF, or CEnKF, is described in [31] in the context of ensemble data assimilation for flood forecasting. In this paper, we provide in the context of data assimilation a complete self-contained description of CBPKF for theoretical background for the alternative formulation and adaptive implementation. Whereas CBPKF was initially motivated for environmental and geophysical state estimation and prediction, it is broadly applicable to a wide range of applications for which improved performance in the extremes is desired. This paper is organized as follows. Sections 2 and 3 describe CB-penalized Fisher-like solution and CBPKF, respectively. Section 4 describes approximation of CBPKF. Section 5 describe the evaluation experiments and results, respectively. Section 6 describes adaptive CBPKF. Section 7 provides the conclusions.

2. Conditional Bias-Penalized Fisher-like Solution

As in Fisher estimation [32], the estimator sought for CB-penalized Fisher-like estimation is $X^* = WZ$ where X^* denotes the (m × 1) vector of the estimated states, **W** denotes the (m × (n + m)) weight matrix, and **Z** denotes the ((n + m) × 1) augmented observation vector. In the above, n denotes the number of observations, m denotes the number of state variables, and (n + m) represents the dimensionality of the augmented vector of the observations and the model-predicted states to be fused for the estimation of the true state X. The purpose of augmentation is to relate directly to CBPKF in Section 3 without introducing additional notations. Throughout this paper, we use regular and bold letters to differentiate the non-augmented and augmented variables, respectively. The linear observation equation is given by:

$$\mathbf{Z} = \mathbf{HX} + \mathbf{V} \tag{1}$$

where X denotes the (m × 1) vector of the true state with E[X] = M_X and Cov[X,X^T] = Ψ_{XX}, **H** denotes the ((n + m) × m) augmented linear observation equation matrix, and **V** denotes the ((n + m) × 1) augmented zero-mean observation error vector with $\text{Cov}[V, V^T] = R$. Assuming independence between X and **V**, we write the Bayesian estimator [30] for X, or X^*, as:

$$X^* = M_X + \mathbf{W}(\mathbf{Z} - \mathbf{H}M_x) \tag{2}$$

The error covariance matrix for X^*, $E\left[(X - X^*)(X - X^*)^T\right]$, is given by:

$$\Sigma_{EV} = (\mathbf{I} - \mathbf{WH})\Psi_{XX}(\mathbf{I} - \mathbf{WH})^T + \mathbf{WRW}^T \tag{3}$$

With Equation (2), we may write Type-II CB as:

$$X - E[X^*|X] = (X - M_X) - WE[(\mathbf{Z} - HM_x)|X] \tag{4}$$

The observation equation for **Z** is obtained by inverting Equation (1):

$$\mathbf{X} = \mathbf{G^T Z} - \mathbf{G^T V} \tag{5}$$

The (mxn) matrix, $\mathbf{G^T}$, in Equation (5) is given by:

$$G^T = \left(U^T H\right)^{-1} U^T \tag{6}$$

where $\mathbf{U^T}$ is some (m $\times$ (n + m)) nonzero matrix. Using Equation (5) and the identity, $\mathbf{\Psi}_{ZZ} = H\Psi_{XX}H^T + R$, we may write the Bayesian estimate for E[**Z**|X] in Equation (4) as:

$$\hat{E}[\mathbf{Z}|X] = HM_X + C(X - M_X) \tag{7}$$

where

$$C = (H\Psi_{XX}H^T + R)G^T[G(H\Psi_{XX}H^T + 2R)G^T]^{-1} \tag{8}$$

Equations (7) and (8) state that the Bayesian estimate of **Z** given X is given by **H**X if the a priori state error covariance Ψ_{XX} is noninformative or there are no observation errors, but by the average of the a priori mean M_X and the observed true state X if the a priori Ψ_{XX} is perfectly informative or observations are information-less.

With Equation (4), we may write the quadratic penalty due to Type-II CB as:

$$\Sigma_{CB} = E[(X - E_{X^*}[X^*|X])(X - E_{X^*}[X^*|X])^T] = (I - WC)\Psi_{XX}(I - WC)^T \tag{9}$$

where I denotes the (m $\times$ m) identity matrix. Combining Σ_{EV} in Equation (3) and Σ_{CB} in Equation (9), we have the apparent error covariance, Σ_a, which reflects both the error covariance and Type-II CB:

$$\Sigma_a = (I - WH)\mathbf{\Psi_{XX}}(I - WH)^T + WRW^T + \alpha(I - WC)\Psi_{XX}(I - WC)^T \tag{10}$$

where α denotes the scaler weight given to the CB penalty term. Minimizing Equation (10) with respect to **W**, or by direct analogy with the Bayesian solution [32], we have:

$$\mathbf{W} = \Psi_{XX}\hat{H}^T[\hat{H}\Psi_{XX}\hat{H}^T + \mathbf{\Lambda}]^{-1} \tag{11}$$

The modified structure matrix $\hat{H}^T$ and observation error covariance matrix $\mathbf{\Lambda}$ in Equation (11) are given by:

$$\hat{H}^T = H^T + \alpha C^T \tag{12}$$

$$\mathbf{\Lambda} = R + \alpha(1 - \alpha)C\Psi_{XX}C^T - \alpha H\Psi_{XX}C^T - \alpha C\Psi_{XX}H^T \tag{13}$$

Using Equation (11) and the matrix inversion lemma [33], we have for Σ_a and X^* in Equations (10) and (2), respectively:

$$\Sigma_a = \alpha\Psi_{XX} + \left[\hat{H}\mathbf{\Lambda}^{-1}\hat{H}^T + \Psi_{XX}^{-1}\right]^{-1} \tag{14}$$

$$X^* = \left[\hat{H}^T\mathbf{\Lambda}^{-1}\hat{H} + \Psi_{XX}^{-1}\right]^{-1}\left\{\hat{H}^T\mathbf{\Lambda}^{-1}\mathbf{Z} + \Psi_{XX}^{-1}M_X\right\} + \Delta \tag{15}$$

where $\Delta = \alpha \Psi_{XX} \hat{H}^T [\hat{H} \Psi_{XX} \hat{H}^T + \Lambda]^{-1} C M_X$. To render the above Bayesian solution to a Fisher-like solution, we assume no a priori information in X and let Ψ_{XX}^{-1} in the brackets in Equations (14) and (15) vanish:

$$\Sigma_a = B [\hat{H} \Lambda^{-1} \hat{H}^T]^{-1} \tag{16}$$

$$X^* = [\hat{H}^T \Lambda^{-1} \hat{H}]^{-1} \hat{H}^T \Lambda^{-1} Z + \Delta \tag{17}$$

where the scaling matrix B is given by $B = \alpha \Psi_{XX} \hat{H}^T \Lambda^{-1} \hat{H} + I$. To obtain the estimator of the form, $X^* = WZ$, we impose the unbiasedness condition, $E[X^*] = X$, or equivalently, $WH = I$. The above condition is satisfied by replacing $[\hat{H}^T \Lambda^{-1} \hat{H}]^{-1}$ with $[\hat{H}^T \Lambda^{-1} H]^{-1}$ and dropping Δ in Equation (17):

$$\Sigma_a = B [\hat{H} \Lambda^{-1} H^T]^{-1} \tag{18}$$

$$X^* = [\hat{H}^T \Lambda^{-1} H]^{-1} \hat{H}^T \Lambda^{-1} Z \tag{19}$$

Finally, we obtain from Equation (3) the error covariance, Σ_{EV}, associated with X^* in Equation (19):

$$\Sigma_{EV} = W R W^T = [\hat{H}^T \Lambda^{-1} H]^{-1} \hat{H}^T \Lambda^{-1} R \Lambda^{-1} \hat{H} [\hat{H}^T \Lambda^{-1} H]^{-1} \tag{20}$$

Note that, if $\alpha = 0$, we have $\hat{H}^T = H$ and $\Lambda = R$, and hence the CB-penalized Fisher-like solution, Equations (19) and (20), is reduced to the Fisher solution [32].

3. Conditional Bias-Penalized Kalman Filter

CBPKF results directly from decomposing the augmented matrices and vectors in Equations (19) and (20) as KF does from the Fisher solution [32]. The CBPKF solution, however, is not very simple because the modified observation error covariance matrix, Λ, is no longer diagonal. An important consideration in casting the CB-penalized Fisher-like solution into CBPKF is to recognize that CB arises from the error-in-variable effects associated with uncertain observations [34], and that the a priori state, represented by the dynamical model forecast, is not subject to CB. We therefore apply the CB penalty to the observations only and reduce **C** in Equation (8) to $C^T = \left(C_{1,k}^T \ C_{2,k}^T \right) = \left(C_{1,k}^T \ 0 \right)$. Separating the observation and dynamical model components in $\hat{H}^T$ and Λ via the matrix inversion lemma, we have:

$$\hat{H}^T = \left(\hat{H}_{1,k}^T \ I \right) \tag{21}$$

$$\Lambda = \begin{bmatrix} \Lambda_{11,k} & \Lambda_{12,k} \\ \Lambda_{21,k} & \Lambda_{22,k} \end{bmatrix} \tag{22}$$

where

$$\hat{H}_{1,k}^T = H_k^T + \alpha C_{1,k}^T \tag{23}$$

$$\Lambda_{11,k} = R_k + \alpha(1-\alpha) C_{1,k} \Psi_{XX} C_{1,k}^T - \alpha H_k \Psi_{XX} C_{1,k}^T - \alpha C_{1,k} \Psi_{XX} H_k^T \tag{24}$$

$$\Lambda_{12,k} = -\alpha C_{1,k} \Psi_{XX} \tag{25}$$

$$\Lambda_{21,k} = \Lambda_{12,k}^T \tag{26}$$

$$\Lambda_{22,k} = \Sigma_{k|k-1} \tag{27}$$

In the above, H_k denotes the (n × m) observation matrix, and R_k denotes the (n × n) observation error covariance matrix. To evaluate the (m × n) matrix, $C_{1,k}$, it is necessary to specify U^T in Equation (6). We use $U^T = H^T$ which ensures invertibility of $U^T H$, but other choices are also possible. We then have for $C_{1,k}^T$:

$$C_{1,k} = \left[\left(H_k \Psi_{XX} H_k^T + R_k \right) G_{1,k} + H_k \Psi_{XX} G_{2,k} \right] L_k^{-1} \tag{28}$$

where

$$G_{2,k}^T = \left(H_k^T H_k + I \right)^{-1} \tag{29}$$

$$G_{1,k}^T = G_{2,k}^T H_k^T \tag{30}$$

$$L_k = G_{2,k}^T \left[H_k^T \left(H_k \Psi_{XX} H_k^T + 2R_k \right) H_k + H_k^T H_k \Psi_{XX} + \Psi_{XX} H_k^T H_k + \Psi_{XX} + 2\Sigma_{k|k-1} \right] G_{2,k}. \tag{31}$$

Expanding **W** in Equation (11) with $\mathbf{\Lambda^{-1}} = \mathbf{\Gamma} = \begin{bmatrix} \Gamma_{11,k} & \Gamma_{12,k} \\ \Gamma_{21,k} & \Gamma_{22,k} \end{bmatrix}$, we have:

$$\mathbf{W} = \left[\hat{\mathbf{H}}^T \mathbf{\Lambda^{-1}} \mathbf{H} \right]^{-1} \hat{\mathbf{H}}^T \mathbf{\Lambda^{-1}} = (\varpi_{1,k} H_k + \varpi_{2,k})^{-1} (\varpi_{1,k} \; \varpi_{2,k}) \tag{32}$$

In Equation (32), the (m × n) and (m × m) weight matrices for the observation and model prediction, $\varpi_{1,k}$ and $\varpi_{2,k}$, respectively, are given by:

$$\varpi_{1,k} = \hat{H}_{1,k}^T \Gamma_{11,k} + \Gamma_{21,k} \tag{33}$$

$$\varpi_{2,k} = \hat{H}_{1,k}^T \Gamma_{12,k} + \Gamma_{22,k} \tag{34}$$

where

$$\Gamma_{22,k} = \left[\Lambda_{22,k} - \Lambda_{21,k} \Lambda_{11,k}^{-1} \Lambda_{12,k} \right]^{-1} \tag{35}$$

$$\Gamma_{11,k} = \Lambda_{11,k}^{-1} + \Lambda_{11,k}^{-1} \Lambda_{12,k} \Gamma_{22,k} \Lambda_{21,k} \Lambda_{11,k}^{-1} \tag{36}$$

$$\Gamma_{12,k} = -\Lambda_{11,k}^{-1} \Lambda_{12,k} \Gamma_{22,k} \tag{37}$$

The apparent CBPKF error covariance, which reflects both Σ_{EV} and Σ_{CB}, is given by Equation (18) as:

$$\Sigma_{a,k|k} = \alpha \Sigma_{k|k-1} + [\varpi_{1,k} H_k + \varpi_{2,k}]^{-1} \tag{38}$$

The CBPKF error covariance, which reflects Σ_{EV} only, is given by Equation (20) as:

$$\Sigma_{k|k} = [\varpi_{1,k} H_k + \varpi_{2,k}]^{-1} \left(\varpi_{1,k} R_k \varpi_{1,k}^T + \varpi_{2,k} \Sigma_{k|k-1} \varpi_{2,k}^T \right) [\varpi_{1,k} H_k + \varpi_{2,k}]^{-1} \tag{39}$$

Because CBPKF minimizes $\Sigma_{a,k|k}$ rather than $\Sigma_{k|k}$, it is not guaranteed that Equation (39) satisfies $\Sigma_{k|k} \leq \Sigma_{k|k-1}$ a priori. If the above condition is not met, it is necessary to reduce α and repeat the calculations. If α is reduced all the way to zero, CBPKF collapses to KF. The CBPKF estimate may be rewritten into a more familiar form:

$$\hat{X}_{k|k} = [\varpi_{1,k} H_k + \varpi_{2,k}]^{-1} [\varpi_{1,k} Z_k + \varpi_{2,k} \hat{X}_{k|k-1}] = \hat{X}_{k|k-1} + K_k [Z_k - H_k \hat{X}_{k|k-1}] \tag{40}$$

In Equation (40), Z_k denotes the (n × 1) observation vector, and the (m × n) CB-penalized Kalman gain, K_k, is given by:

$$K_k = [\varpi_{1,k} H_k + \varpi_{2,k}]^{-1} \varpi_{1,k} \tag{41}$$

To operate the above as a sequential filter, it is necessary to prescribe Ψ_{XX} and α. An obvious choice for Ψ_{XX}, i.e., the a priori error covariance of the state, is $\Sigma_{k|k-1}$. Specifying α requires some care. In general, a larger α improves accuracy over the tails but at the expense of increasing unconditional error. Too small an α may not affect large enough CB penalty in which case the CBPKF and KF solutions would differ a little. Too large an α, on the other hand, may severely violate the $\Sigma_{k|k} \leq \Sigma_{k|k-1}$ condition in which case the filter may have to be iterated at an additional computational expense with successively reduced α. A reasonable strategy for reducing α is $\alpha_i = c\alpha_{i-1}$, $i = 1, 2, 3, \ldots$, with $0 < c < 1$ where α_i denotes the value of α at the i-th iteration [24,29]. For high-dimensional problems, CBPKF can be computationally very expensive. Whereas KF requires solving an (m × n) linear system only once per updating or fusion cycle, CBPKF additionally requires solving two

(m × m) linear systems (for $C_{1,k}$ and Γ_{22}), and a (n × n) system (for Λ_{11}), assuming that the structure of the observation equation does not change in time (in which case $G_{2,k}^T$ in Equation (29) may be evaluated only once). To reduce computation, below we approximate CBPKF with KF by inflating the forecast error covariance.

4. VIKF Approximation of CBPKF

The main idea behind this simplification is that, if the gain for the CB penalty, **C**, in Equation (10) can be linearly approximated with **H**, the apparent error covariance Σ_a becomes identical to Σ_{EV} in Equation (3) but with Ψ_{XX} inflated by a factor of $1 + \alpha$:

$$\Sigma_{(1+\alpha)} = (I - WH)(1 + \alpha)\Psi_{XX}(I - WH)^T + WR_{(1+\alpha)}W^T \tag{42}$$

where $R_{(1+\alpha)} = \begin{bmatrix} R & 0 \\ 0 & (1+\alpha)\Psi_{XX} \end{bmatrix}$. The KF solution for Equation (42) is identical to the standard KF solution but with $\Sigma_{k|k-1}$ replaced by $(1 + \alpha)\Sigma_{k|k-1}$:

$$\hat{X}_{k|k} = [H_k^T R_k^{-1} H_k + \{(1+\alpha)\Sigma_{k|k-1}\}^{-1}]^{-1}[H_k^T R_k^{-1} Z_k + \left\{(1+\alpha)\Sigma_{k|k-1}\right\}^{-1}\hat{X}_{k|k-1}] \tag{43}$$

With **WH**=I in Equation (43) for the VIKF solution, we have $\Sigma_{(1+\alpha)} = WR_{(1+\alpha)}W^T$ for the apparent filtered error variance of $\hat{X}_{k|k}$ in Equation (42). The error covariance of $\hat{X}_{k|k}$, $\Sigma_{k|k}$, is given by Equation (3) as:

$$\begin{aligned} \Sigma_{k|k} = WRW^T &= \left[H^T R_{(1+\alpha)}^{-1} H\right]^{-1} H^T R_{(1+\alpha)}^{-1} R R_{(1+\alpha)}^{-1} H \left[H^T R_{(1+\alpha)}^{-1} H\right]^{-1} \\ &= \Sigma_{(1+\alpha),k|k} \Sigma_{(1+\alpha)^2,k|k}^{-1} \Sigma_{(1+\alpha),k|k} \end{aligned} \tag{44}$$

In Equation (44), the inflated filtered error covariance, $\Sigma_{\beta,k|k}$, where β denotes the multiplicative inflation factor, is given by:

$$\begin{aligned} \Sigma_{\beta,k|k} &= \beta\Sigma_{k|k-1} - \beta\Sigma_{k|k-1} H_k^T \left[H_k \beta\Sigma_{k|k-1} H_k^T + R_k\right]^{-1} H_k \beta\Sigma_{k|k-1} \\ &= \left[H_k^T R_k^{-1} H_k + \left(\beta\Sigma_{k|k-1}\right)^{-1}\right]^{-1} \end{aligned} \tag{45}$$

Computationally, the evaluation of Equations (43) and (44) requires solving two (m × n) and a (m × m) linear systems. As in the original formulation of CBPKF, iterative reduction of α is necessary to ensure $\Sigma_{k|k} \leq \Sigma_{k|k-1}$.

The above approximation assumes that the CB penalty, Σ_{CB}, is proportional to the error covariance, Σ_{EV}. To help ascertain how KF, CBPKF and the VIKF approximation may differ, we compare in Table 1 their analytical solutions for gain κ_k, and filtered error variance $\sigma_{k|k}^2$ for the 1D case of m = n = 1. The table shows that the VIKF approximation and CBPKF are identical for the 1D problem except that the CB penalty for CBPKF is twice as large as that for the VIKF approximation. To visualize the differences, Figure 1 shows κ_k and $\sigma_{k|k}^2$ for KF, the VIKF approximation and CBPKF for the three cases of $\sigma_{k|k-1}^2 = 1$ and $\sigma_Z^2 = 1$ (left), $\sigma_{k|k-1}^2 = 1$ and $\sigma_Z^2 = 4$ (middle), and $\sigma_{k|k-1}^2 = 4$ and $\sigma_Z^2 = 1$ (right). For all cases, we set h to unity and varied α from 0 to 1. The figure indicates that, compared to KF, the VIKF approximation and CBPKF prescribe appreciably larger gains, that the increase in gain is larger for larger α, and that the CBPKF gain is larger than the gain in the VIKF approximation for the same value of α. The figure also indicates that, compared to KF error variance, CBPKF error variance is larger, and that the increase in error variance is larger for larger α. Note that the differences between the KF and CBPKF solutions are the smallest for $\sigma_{k|k-1}^2 > \sigma_Z^2$, a reflection of the diminished impact of CB owing to the comparatively smaller uncertainty in the observations. The above development suggests that one may be able to approximate CBPKF very closely with the VIKF-based formulation by adjusting α in the

latter. Below, we evaluate the performance of CBPKF relative to KF and the VIKF-based approximation of CBPKF.

Table 1. Comparison of gain and filtered error variance among KF, the VIKF approximation, and CBPKF.

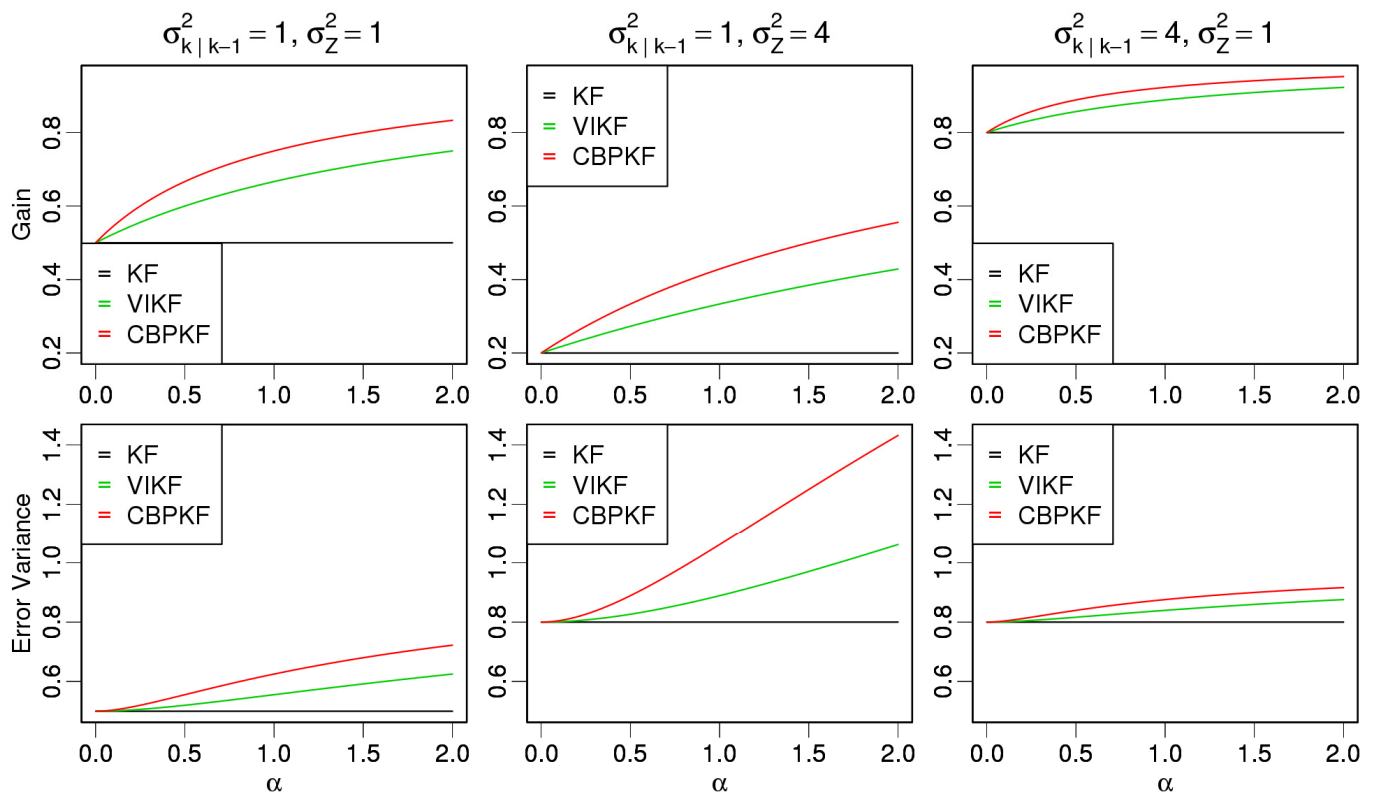|	Gain, κ_k	Filtered Error Variance, $\sigma^2_{k|k}$
KF	$\dfrac{h\sigma^2_{k|k-1}}{h^2\sigma^2_{k|k-1}+\sigma^2_Z}$	$\dfrac{\sigma^2_Z}{h^2\sigma^2_{k|k-1}+\sigma^2_Z}\sigma^2_{k|k-1}$
VIKF approx.	$\dfrac{h(1+\alpha)\sigma^2_{k|k-1}}{h^2(1+\alpha)\sigma^2_{k|k-1}+\sigma^2_Z}$	$\dfrac{\left\{(1+\alpha)^2h^2\sigma^2_{k|k-1}+\sigma^2_Z\right\}\sigma^2_Z}{\left\{(1+\alpha)h^2\sigma^2_{k|k-1}+\sigma^2_Z\right\}^2}\sigma^2_{k|k-1}$
CBPKF	$\dfrac{h(1+2\alpha)\sigma^2_{k|k-1}}{h^2(1+2\alpha)\sigma^2_{k|k-1}+\sigma^2_Z}$	$\dfrac{\left\{(1+2\alpha)^2h^2\sigma^2_{k|k-1}+\sigma^2_Z\right\}\sigma^2_Z}{\left\{(1+2\alpha)h^2\sigma^2_{k|k-1}+\sigma^2_Z\right\}^2}\sigma^2_{k|k-1}$

Figure 1. Comparison of κ_k and $\sigma^2_{k|k}$ for KF, the VIKF approximation and CBPKF for three different cases: $\sigma^2_{k|k-1}=1$ and $\sigma^2_Z=1$ (**left**), $\sigma^2_{k|k-1}=1$ and $\sigma^2_Z=4$ (**middle**), $\sigma^2_{k|k-1}=4$ and $\sigma^2_Z=1$ (**right**).

5. Evaluation and Results

For comparative evaluation, we carried out the synthetic experiments of [25]. We assume the following linear dynamical and observation models with perfectly known statistical parameters:

$$X_k = \Phi_{k-1}X_{k-1} + W_{k-1} \tag{46}$$

$$Z_k = H_kX_k + V_k \tag{47}$$

where X_k and X_{k-1} denote the state vectors at time steps k and k − 1, respectively, Φ_{k-1} denotes the state transition matrix at time step k − 1 assumed as $\Phi_{k-1} = \varphi_{k-1}I$, W_{k-1} denotes the white noise vector, $w_{j,k-1} \sim N\left(0, \sigma^2_{w_{k-1}}\right)$, j = 1, . . . ,m, with $Q_{k-1} = E\left[W_{k-1}W^T_{k-1}\right]$, and V_k denotes the observation error vector, $v_{i,k} \sim N\left(0, \sigma^2_{v_k}\right)$, I = 1, . . . , n. The number of observations, n, is assumed to be time-invariant. The observation errors are assumed to be

independent among themselves and of the true state. To assess comparative performance under widely varying conditions, we randomly perturbed φ_{k-1}, $\sigma_{w,k-1}$ and $\sigma_{v,k}$ above according to Equations (48) through (50) below, and used only those deviates that satisfy the bounds:

$$\varphi_{k-1}^{p} = \varphi_{k-1} + \gamma_{\varphi}\varepsilon_{\varphi} \quad 0.5 \leq \varphi_{k-1}^{p} \leq 0.95 \tag{48}$$

$$\sigma_{w,k-1}^{p} = \sigma_{w,k-1} + \gamma_{w}\varepsilon_{w} \quad \sigma_{w,k-1}^{p} \geq 0.01 \tag{49}$$

$$\sigma_{v,k}^{p} = \sigma_{v,k} + \gamma_{v}\varepsilon_{v} \quad \sigma_{v,k}^{p} \geq 0.01 \tag{50}$$

In the above, the superscript p signifies that the variable is a perturbation, ε_{φ} ε_{w} and ε_{v} denote the normally distributed white noise for the respective variables, and γ_{φ} γ_{w} and γ_{v} denote the standard deviations of the white noise added to φ_{k-1}, $\sigma_{w,k-1}$ and $\sigma_{v,k}$, respectively. The parameter settings (see Table 1) are chosen to encompass less predictable (small φ_{k-1}) to more predictable (large φ_{k-1}) processes, certain (small $\sigma_{w,k-1}$) to uncertain (large $\sigma_{w,k-1}$) model dynamics, and more informative (small $\sigma_{v,k}$) to less informative (large $\sigma_{v,k}$) observations. The bounds for φ_{k-1}^{p} in Equation (48) are based on the range of lag-1 serial correlation representing moderate to high predictability where CBPKF and KF are likely to differ the most. The bounding of the perturbed values $\sigma_{w,k-1}^{p}$ and $\sigma_{v,k}^{p}$ in Equations (49) and (50), respectively, is necessary to avoid the observational or model prediction uncertainty becoming unrealistically small. Very small $\sigma_{w,k-1}^{p}$ and $\sigma_{v,k}^{p}$ render the information content of the model prediction, $\Sigma_{k|k-1}$, and the observation, Z_{k}, respectively, very large, and hence keep the filters operating in unrealistically favorable conditions for extended periods of time. We then apply KF, CBPKF, and the VIKF approximation to obtain $\hat{X}_{k|k}$ and $\Sigma_{k|k}$, and verify them against the assumed truth. To evaluate the performance of CBPKF relative to KF, we calculate percent reduction in root-mean-square error (RMSE) by CBPKF over KF conditional on the true state exceeding some threshold between 0 and the largest truth.

Figure 2 shows the percent reduction in RMSE by CBPKF over KF for Cases 1 (left), 5 (middle) and 9 (right) representing Groups 1, 2 and 3 in Table 1, respectively. The three groups differ most significantly in the variability of the dynamical model error, γ_{w}, and may be characterized as nearly stationary (Group 1), nonstationary (Group 2), and highly nonstationary (Group 3). The range of α values used is [0.1, 1.2] with an increment of 0.1. The numbers of state variables, observations, and updating cycles used in Figure 2 are 1, 10, and 100,000 for all cases. The dotted line at 10% reduction in the figure serves as a reference for significant improvement. The figure shows that, at the extreme end of the tail, CBPKF with α of 0.7, 0.6, and 0.5 reduces RMSE by about 15, 25, and 30% for Cases 1, 5 and 9, respectively, but at the expense of increasing unconditional RMSE by about 5%. The general pattern of reduction in RMSE for other cases in Table 1 is similar within each group and is not shown. We only note here that larger variability in observational uncertainty (i.e., larger γ_{v}) reduces the relative performance of CBPKF somewhat, and that the magnitude of variability in predictability (i.e., γ_{φ}) has a relatively small impact on the relative performance.

It was seen in Table 1 that the VIKF approximation is identical to CBPKF for m = n = 1 but for the multiplicative scaler weight for the CB penalty. Numerical experiments indicate that, whereas the above relationship does not hold for other m or n, one may very closely approximate CBPKF with the VIKF-based formulation by adjusting α. For example, the VIKF approximation with α increased by a factor of 1.25 to 1.90 differ from CBPKF only by 1% or less for all 12 cases in Table 2 with m = 1 and n = 10. The above findings indicate that the VIKF approximation may be used as a computationally less expensive alternative for CBPKF. Table 3 compares the CPU time among KF, CBPKF, and the VIKF approximation for six different combinations of m and n based using Intel(R) Xeon(R) Gold 6152 CPU @ 2.10 GHz. The computing time is reported in multiples of the KF's. Note that the original formulation of CBPKF quickly becomes extremely expensive as the dimensionality of the

problem increases, whereas the CPU time of the VIKF approximation stays under 3.5 times that of KF for the size of the problems considered.
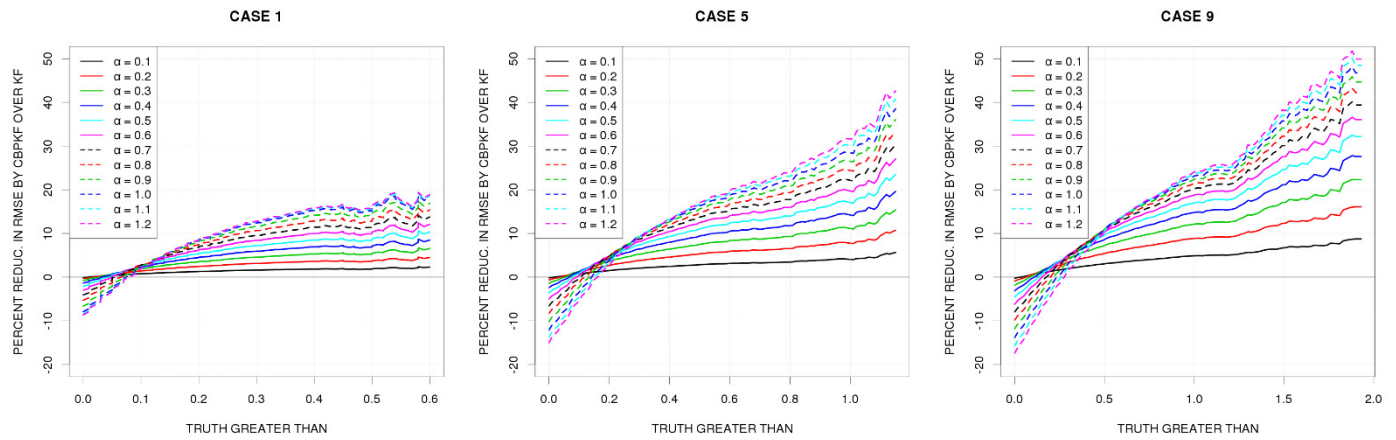

Figure 2. Percent reduction in RMSE by CBPKF over KF for a range of values of α for Cases 1 (**left**), 5 (**middle**), and 9 (**right**).

Table 2. Parameter settings for the 12 cases considered.

Group	Case	$\sigma_{w,k-1}$	γ_w	$\sigma_{v,k}$	γ_v	φ_{k-1}	γ_φ
	1	0.1	0.01	1.5	0.4	0.7	0.1
1	2	0.1	0.01	1.5	0.4	0.7	0.8
	3	0.1	0.01	1.5	1.2	0.7	0.1
	4	0.1	0.01	1.5	1.2	0.7	0.8
	5	0.1	0.1	1.5	0.4	0.7	0.1
2	6	0.1	0.1	1.5	0.4	0.7	0.8
	7	0.1	0.1	1.5	1.2	0.7	0.1
	8	0.1	0.1	1.5	1.2	0.7	0.8
	9	0.1	0.2	1.5	0.4	0.7	0.1
3	10	0.1	0.2	1.5	0.4	0.7	0.8
	11	0.1	0.2	1.5	1.2	0.7	0.1
	12	0.1	0.2	1.5	1.2	0.7	0.8

Table 3. Comparison of computing time among KF, CBPKF, and VIKF approximation.

Dimensionality		Normalized Computing Time		
m	n	KF	CBPKF	VIKF approx.
1	10	1	5.23	1.51
1	40	1	18.41	2.74
5	10	1	6.44	1.67
5	40	1	24.03	2.88
10	10	1	14.27	2.03
10	40	1	27.96	3.46

If the filtered error variance is unbiased, one would expect the mean of the actual error squared associated with the variance to be approximately the same as the variance itself. To verify this, we show in Figure 3 the filtered error variance vs. the actual error squared

for KF (left), the VIKF approximation (middle), and CBPKF (right) for all ranges of filtered error variance. For reference, we plot the one-to-one line representing the unbiased error variance conditional on the magnitude of the filtered error variance and overlay the local regression fit through the actual data points using the R package locfit [35]. The figure shows that all three provide conditionally unbiased estimates of filtered error variance as theoretically expected, and that the VIKF approximation and CBPKF results are extremely similar to each other.
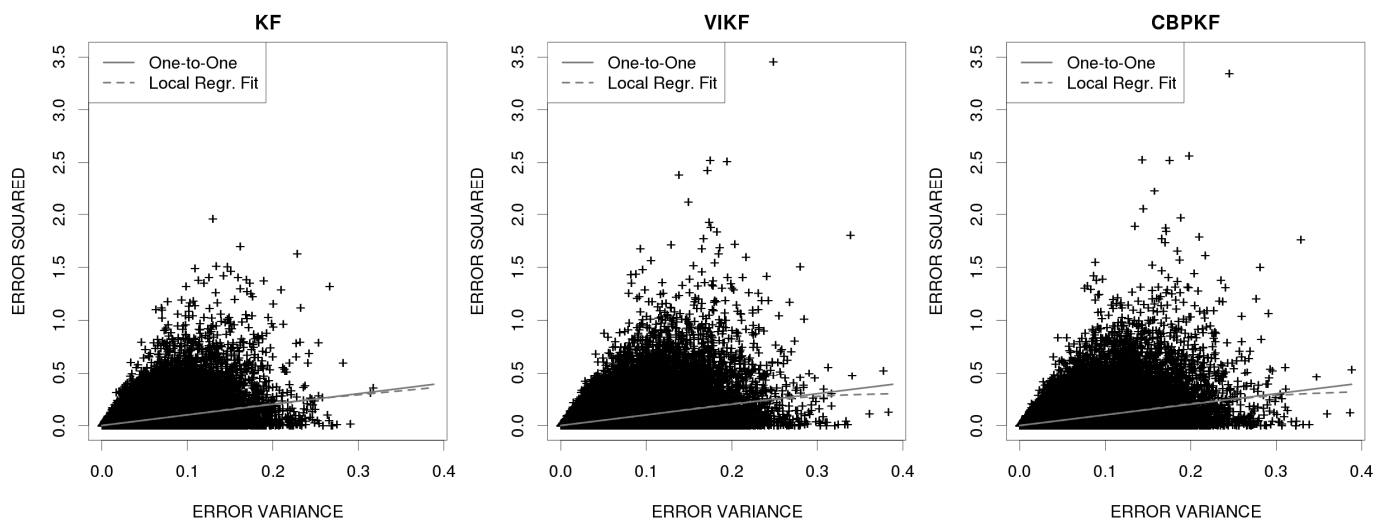

Figure 3. Filtered error variance vs. error squared for KF (**left**), the VIKF approximation (**middle**), and CBPKF (**right**). The one-to-one line is shown in black and the local regression fit is shown in green.

6. Adaptive CBPKF

Whereas CBPKF or the VIKF approximation significantly improves the accuracy of the estimates over the tails, it deteriorates performance near the median. Figure 2 suggests that if α can be prescribed adaptively such that a small/large CB penalty is affected when the system is in the normal/extreme state, the unconditional performance of CBPKF would improve. Because the true state of the system is not known, adaptively specifying α is necessarily an uncertain proposition. There are, however, certain applications in which the normal vs. extreme state of the system may be ascertained with higher accuracy than others. For example, the soil moisture state of a catchment may be estimated from assimilating precipitation and streamflow data into hydrologic models [36–41]. If α is prescribed adaptively based on the best available estimate of the state of the catchment, one may expect improved performance in hydrologic forecasting. In this section, we apply adaptive CBPKF in the synthetic experiment and assess its performance. An obvious strategy for adaptively filtering is to parameterize α in terms of the KF estimate (i.e., the CBPKF estimate with $\alpha = 0$) as the best guess for the true state. The premise of this strategy is that, though it may be conditionally biased, the KF estimate fuses the information available from both the observations and the dynamical model, and hence best captures the relationship between α and the departure of the state of the system from median. A similar approach has been used in fusing radar rainfall data and rain gauge observations for multisensor precipitation estimation in which an ordinary cokriging estimate was used to prescribe α in CB-penalized cokriging [30].

Necessarily, the effectiveness of the above strategy depends on the skill of the KF estimate; if the skill is very low, one may not expect significant improvement. Figure 2

suggests that, qualitatively, α should increase as the state becomes more extreme. To that end, we employed the following model for time-varying α:

$$\alpha_k = \gamma \| \hat{X}_{k|k}^{KF} \| \tag{51}$$

where α_k denotes the multiplicative CB penalty factor for CBPKF at time step k, $\| \hat{X}_{k|k}^{KF} \|$ denotes some norm of the KF estimate at time step k, and γ denotes the proportionality constant. Figure 4(left) shows the RMSE reduction by adaptive CBPKF over KF with $\alpha_k = \gamma \left| \hat{X}_{k|k}^{KF} \right|$ for the 12 cases in Table 2 m = 1 and n = 10. The γ values used were 3.0, 1.0, and 0.5 for Groups 1, 2, and 3 in Table 2, respectively. The figure shows that adaptive CBPKF performs comparably to KF in the unconditional sense while substantially improving performance in the tails. The rate of reduction in RMSE with respect to the increasing conditioning truth, however, is now slower than that seen in Figure 2 due to the occurrences of incorrectly specified α. To assess the uppermost bound of the feasible performance of adaptive CBPKF, we also specified α with perfect accuracy under Equation (51) via $\alpha_k = \gamma |X_k|$ where X_k denotes the true state. The results are shown in Figure 4(right) for which the γ values used were 3.0, 1.5, and 1.0 for Groups 1, 2, and 3 in Table 2, respectively. The figure indicates that adaptive CBPKF with perfectly prescribed α greatly improves performance, even outperforming KF in the unconditional sense. Figure 4 suggests that, if α can be prescribed more accurately with additional sources of information, the performance of adaptive CBPKF may be improved beyond the level seen in Figure 4(left). Finally, we show in Figure 5 the example scatter plots of the KF (black) and adaptive CBPKF (red) estimates vs. truth. They are for Cases 1 and 9 in Table 2 representing Groups 1 and 3, respectively. It is readily seen that the CBPKF significantly reduces CB in the tails while keeping its estimates close to the KF estimates in normal ranges.
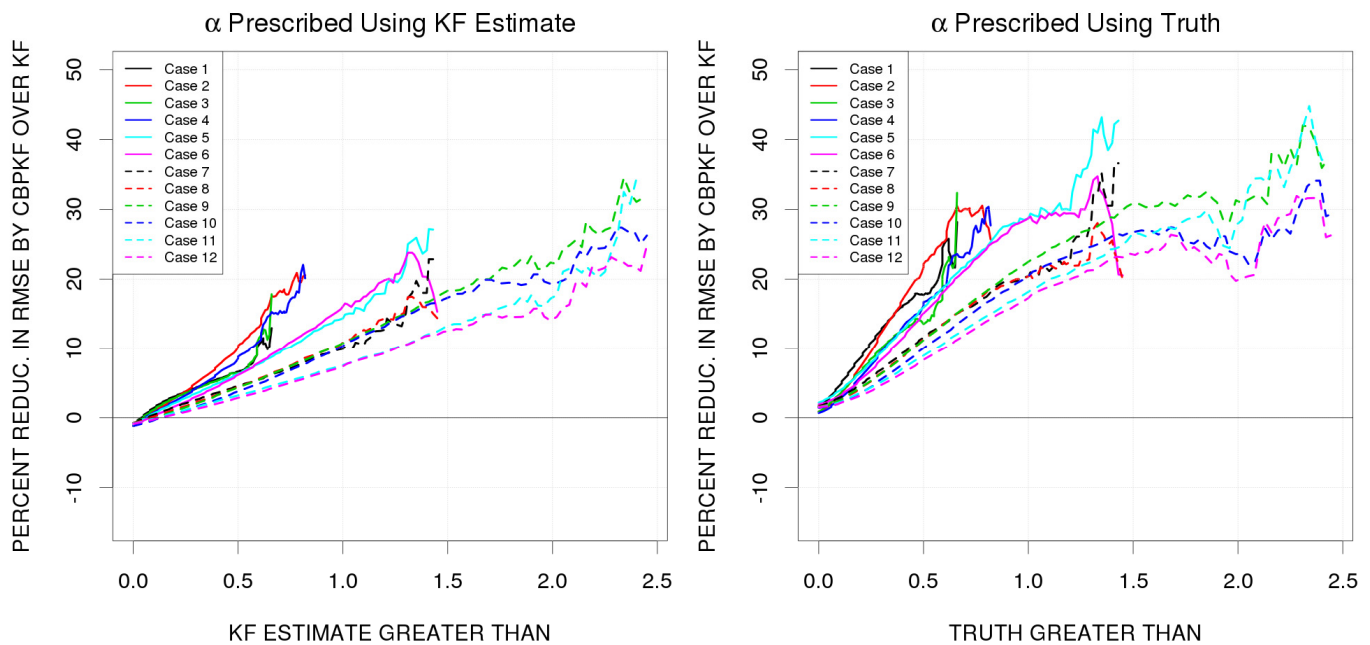

Figure 4. Percent reduction in RMSE by adaptive CBPKF over KF in which α is prescribed using the KF estimate (**left**) and the truth (**right**).
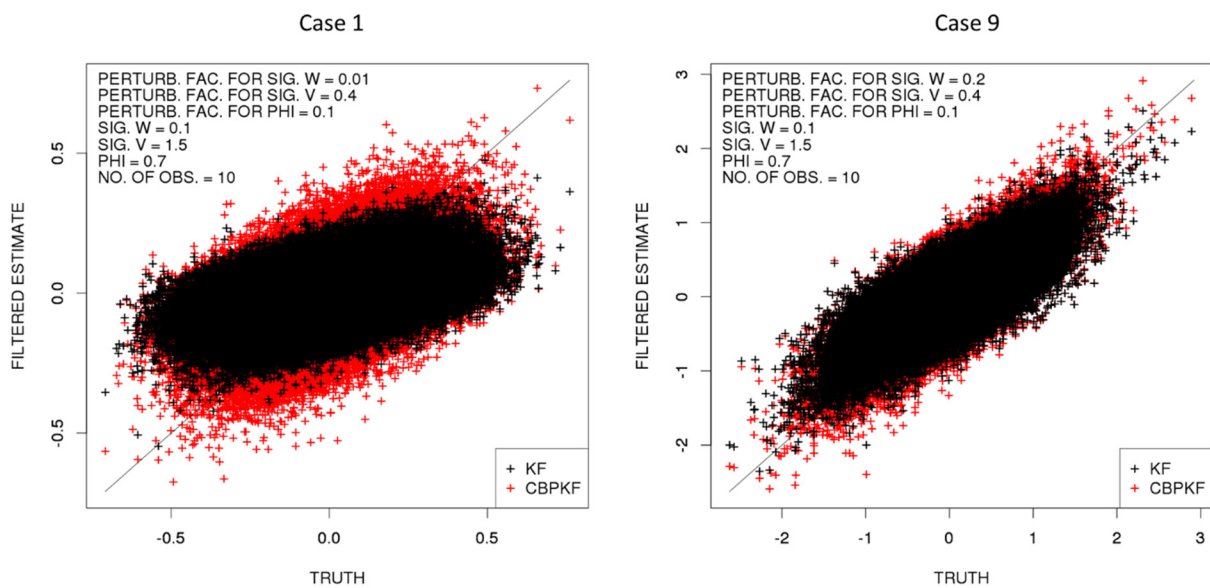
Figure 5. Example scatter plots of KF (black) and adaptive CBPKF (red) estimates vs. truth for Cases 1 (**left**) and 9 (**right**) in Table 2.

7. Conclusions

Conditional bias-penalized Kalman filter (CBPKF) has recently been developed to improve the estimation and prediction of extremes. The original formulation, however, is computationally very expensive, and deteriorates performance in the normal ranges relative to KF. In this work, we present a computationally less expensive alternative based on the variance-inflated KF (VIKF) approximation, and improve unconditional performance by adaptively prescribing the weight for the CB penalty. For evaluation, we carried out synthetic experiments using linear systems with varying degrees of dynamical model uncertainty, observational uncertainty, and predictability. The results indicate that the VIKF-based approximation of CBPKF provides a computationally much less expensive alternative to the original formulation, and that adaptive CBPKF performs comparably to KF in the unconditional sense while improving the estimation of extremes by about 20 to 30% over KF. It is also shown that additional improvement may be possible by improving adaptive prescription of the weight to the CB penalty using additional sources of information. The findings indicate that adaptive CBPKF offers a significant addition to the dynamic filtering methods for general application in data assimilation and, in particular, when or where the estimation of extremes is of importance. The findings in this work are based on idealized synthetic experiments that satisfy linearity and normality. Additional research is needed to assess performance for non-normal problems and for nonlinear problems using the ensemble extension [31], and to prescribe the weight for the CB penalty more skillfully.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wood, A.W.; Lettenmaier, D.P. An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophys. Res. Lett.* **2008**, *35*, L14401. [CrossRef]
2. Gupta, H.V.; Clark, M.P.; Vrugt, J.A.; Abramowitz, G.; Ye, M. Towards a comprehensive assessment of model structural adequacy. *Water Resour. Res.* **2012**, *48*, W08301. [CrossRef]
3. Antoniou, C.; Ben-Akiva, M.; Koutsopoulos, H.N. Nonlinear Kalman filtering algorithms for on-Line calibration of dynamic traffic assignment models. *IEEE Trans. Intell. Transp. Syst.* **2007**, *8*, 661–670. [CrossRef]
4. Bhotto, M.Z.A.; Bajić, I.V. Constant modulus blind adaptive beamforming based on unscented Kalman filtering. *IEEE Signal Process. Lett.* **2015**, *22*, 474–478. [CrossRef]
5. Bocher, M.; Fournier, A.; Coltice, N. Ensemble Kalman filter for the reconstruction of the Earth's mantle circulation. *Nonlinear Process. Geophys.* **2018**, *25*, 99–123. [CrossRef]
6. Chen, W.; Shen, H.; Huang, C.; Li, X. Improving soil moisture estimation with a dual ensemble Kalman smoother by jointly assimilating AMSR-E brightness temperature and MODIS LST. *Remote Sens.* **2017**, *9*, 273. [CrossRef]
7. Gao, Z.; Shen, W.; Zhang, H.; Ge, M.; Niu, X. Application of helmert variance component based adaptive Kalman filter in multi-GNSS PPP/INS tightly coupled integration. *Remote Sens.* **2016**, *8*, 553. [CrossRef]
8. Houtekamer, P.L.; Zhang, F. Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **2016**, *144*, 4489–4532. [CrossRef]
9. Jain, A.; Krishnamurthy, P.K. Phase noise tracking and compensation in coherent optical systems using Kalman filter. *IEEE Commun. Lett.* **2016**, *20*, 1072–1075. [CrossRef]
10. Jiang, Y.; Liao, M.; Zhou, Z.; Shi, X.; Zhang, L.; Balz, T. Landslide deformation analysis by coupling deformation time series from SAR data with hydrological factors through data assimilation. *Remote Sens.* **2016**, *8*, 179. [CrossRef]
11. Kurtz, W.; Franssen, H.-J.H.; Vereecken, H. Identification of time-variant river bed properties with the ensemble Kalman filter. *Water Resour. Res.* **2012**, *48*, W10534. [CrossRef]
12. Lu, X.; Wang, L.; Wang, H.; Wang, X. Kalman filtering for delayed singular systems with multiplicative noise. *IEEE/CAA J. Autom. Sin.* **2016**, *3*, 51–58. [CrossRef]
13. Lv, H.; Qi, F.; Zhang, Y.; Jiao, T.; Liang, F.; Li, Z.; Wang, J. Improved detection of human respiration using data fusion based on a multi-static UWB radar. *Remote Sens.* **2016**, *8*, 773. [CrossRef]
14. Ma, R.; Zhang, L.; Tian, X.; Zhang, J.; Yuan, W.; Zheng, Y.; Zhao, X.; Kato, T. Assimilation of remotely-sensed leaf area index into a dynamic vegetation model for gross primary productivity estimation. *Remote Sens.* **2017**, *9*, 188. [CrossRef]
15. Muñoz-Sabater, J. Incorporation of passive microwave brightness temperatures in the ECMWF soil moisture analysis. *Remote Sens.* **2015**, *7*, 5758–5784. [CrossRef]
16. Nair, A.; Indu, J. Enhancing Noah land surface model prediction skill over Indian subcontinent by assimilating SMOPS blended soil moisture. *Remote Sens.* **2016**, *8*, 976. [CrossRef]
17. Reichle, R.H.; McLaughlin, D.B.; Entekhabi, D. Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Weather Rev.* **2002**, *130*, 103–114. [CrossRef]
18. Wallace, L.; Lucieer, A.; Watson, C.; Turner, D. Development of a UAV-LiDAR system with application to forest inventory. *Remote Sens.* **2012**, *4*, 1519–1543. [CrossRef]
19. de Wit, A.J.W.; van Diepen, C.A. Crop model data assimilation with the ensemble Kalman filter for improving regional crop yield forecasts. *Agric. For. Meteorol.* **2007**, *146*, 38–56. [CrossRef]
20. Yan, M.; Tian, X.; Li, Z.; Chen, E.; Wang, X.; Han, Z.; Sun, H. Simulation of forest carbon fluxes using model incorporation and data assimilation. *Remote Sens.* **2016**, *8*, 567. [CrossRef]
21. Yu, K.K.C.; Watson, N.R.; Arrillaga, J. An adaptive Kalman filter for dynamic harmonic state estimation and harmonic injection tracking. *IEEE Trans. Power Del.* **2005**, *20*, 1577–1584. [CrossRef]
22. Dong, Z.; You, Z. Finite-horizon robust Kalman filtering for uncertain discrete time-varying systems with uncertain-covariance white noises. *IEEE Signal Process. Lett.* **2006**, *13*, 493–496. [CrossRef]
23. Zhou, H.; Huang, H.; Zhao, H.; Zhao, X.; Yin, X. Adaptive unscented Kalman filter for target tracking in the presence of nonlinear systems involving model mismatches. *Remote Sens.* **2017**, *9*, 657. [CrossRef]
24. Ciach, G.J.; Morrissey, M.L.; Krajewski, W.F.; Ciach, G.J.; Morrissey, M.L.; Krajewski, W.F. Conditional bias in radar rainfall estimation. *J. Appl. Meteorol.* **2000**, *39*, 1941–1946. [CrossRef]
25. Seo, D.-J.; Saifuddin, M.M.; Lee, H. Conditional bias-penalized Kalman filter for improved estimation and prediction of extremes. *Stochastic Environ. Res. Risk Assess* **2018**, *32*, 183–201, Erratum in *Stochastic Environ. Res. Risk Assess.* **2018**, *32*, 3561–3562. [CrossRef]
26. Jolliffe, I.T.; Stephenson, D.B. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*; John Wiley & Sons: Hoboken, NJ, USA, 2003.
27. Brown, J.D.; Seo, D.-J.; Brown, J.D.; Seo, D.-J. A nonparametric postprocessor for bias correction of hydrometeorological and hydrologic ensemble forecasts. *J. Hydrometeorol.* **2010**, *11*, 642–665. [CrossRef]
28. Seo, D.-J. Conditional bias-penalized kriging (CBPK). *Stochastic Environ. Res. Risk Assess.* **2013**, *27*, 43–58. [CrossRef]
29. Seo, D.-J.; Siddique, R.; Zhang, Y.; Kim, D. Improving real-time estimation of heavy-to-extreme precipitation using rain gauge data via conditional bias-penalized optimal estimation. *J. Hydrol.* **2014**, *519*, 1824–1835. [CrossRef]

30. Kim, B.; Seo, D.-J.; Noh, S.J.; Prat, O.P.; Nelson, B.R. Improving multi-sensor estimation of heavy-to-extreme precipitation via conditional bias-penalized optimal estimation. *J. Hydrol.* **2018**, *556*, 1096–1109. [CrossRef]

31. Lee, H.; Noh, S.J.; Kim, S.; Shen, H.; Seo, D.-J.; Zhang, Y. Improving flood forecasting using conditional bias-penalized ensemble Kalman filter. *J. Hydrol.* **2019**, *575*, 596–611. [CrossRef]

32. Schweppe, F.C. Uncertain Dynamic Systems, Prentice-Hall. 1973. Available online: https://openlibrary.org/books/OL5291577M/Uncertain_dynamic_systems (accessed on 12 January 2022).

33. Woodbury, M.A. *Inverting Modified Matrices*; Princeton University: Princeton, NJ, USA, 1950.

34. Hausman, J. Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *J. Econ. Perspect.* **2001**, *15*, 57–67. [CrossRef]

35. Catherine Loader, Locfit: Local Regression, Likelihood and Density Estimation. 2013. Available online: https://CRAN.Rproject.org/package=locfit (accessed on 12 January 2022).

36. Lee, H.; Seo, D.-J. Assimilation of hydrologic and hydrometeorological data into distributed hydrologic model: Effect of adjusting mean field bias in radar-based precipitation estimates. *Adv. Water Resour.* **2014**, *74*, 196–211. [CrossRef]

37. Lee, H.; Seo, D.-J.; Koren, V. Assimilation of streamflow and in situ soil moisture data into operational distributed hydrologic models: Effects of uncertainties in the data and initial model soil moisture states. *Adv. Water Resour.* **2011**, *34*, 1597–1615. [CrossRef]

38. Lee, H.; Seo, D.-J.; Liu, Y.; Koren, V.; McKee, P.; Corby, R. Variational assimilation of streamflow into operational distributed hydrologic models: Effect of spatiotemporal scale of adjustment. *Hydrol. Earth Syst. Sci.* **2012**, *16*, 2233–2251. [CrossRef]

39. Lee, H.; Zhang, Y.; Seo, D.-J.; Xie, P. Utilizing satellite precipitation estimates for streamflow forecasting via adjustment of mean field bias in precipitation data and assimilation of streamflow observations. *J. Hydrol.* **2015**, *529*, 779–794. [CrossRef]

40. Rafieeinasab, A.; Seo, D.-J.; Lee, H.; Kim, S. Comparative evaluation of maximum likelihood ensemble filter and ensemble Kalman filter for real-time assimilation of streamflow data into operational hydrologic models. *J. Hydrol.* **2014**, *519*, 2663–2675. [CrossRef]

41. Seo, D.-J.; Koren, V.; Cajina, N. Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. *J. Hydrometeor.* **2003**, *4*, 627–641. [CrossRef]