2005 American Control Conference
June 8-10, 2005. Portland, OR, USA

Unscented Kalman Filtering for Additive Noise Case: Augmented vs. Non-augmented

Yuanxin Wu, Dewen Hu, *Member, IEEE*, Meiping Wu and Xiaoping Hu

Abstract—This paper concerns the unscented Kalman filtering (UKF) for the nonlinear dynamic systems with additive process and measurement noises. We find that under some condition, the basic difference between them is that the augmented UKF draws sigma set only once within a filtering recursion while the non-augmented UKF has to redraw a new set of sigma points to incorporate the effect of additive process noise. This difference generally favors the augmented UKF. The analyses are supported by a representative example.

***Index Terms*—unscented transformation, unscented Kalman filtering, dynamic system**

I. INTRODUCTION

In light of the intuition that to approximate a probability distribution is easier than to approximate an arbitrary nonlinear transformation, Julier and Uhlmann [1, 2] invented the unscented transformation (UT) to make probabilistic inference. Eliminating the cumbersome derivation and evaluation of Jacobian/Hessian matrices, the UT-based unscented Kalman filter (UKF) is much easier to implement and performs better than the EKF. The original UKF was first formulated in its augmented form [1-4]. It is believed that for the special (but often found) case where process and measurement noises are additive, the computational complexity can be reduced by using the non-augmented form, which presumably yields similar results [4], if not the same. The non-augmented UKF has been accepted and employed to analyze the practical systems [5, 6]. In this paper we will show that it is not a quite right belief and the non-augmented UKF usage can lead to noticeable losses in accuracy. The contents are organized as follows. Section II shows in a rigorous way the conditionally equivalent relationship between the non-augmented and augmented UTs. This will facilitate the discussions about the UKF, which is essentially a natural extension of the UT to recursive estimation. Section III analyzes and compares the non-augmented and augmented UKFs. Section IV examines a representative example in the signal processing community to support our findings and the conclusions are drawn in Section V.

This work was supported in part by National Natural Science Foundation of China (60374006, 60234030 and 60171003), Distinguished Young Scholars Fund of China (60225015), and Ministry of Education of China (TRAPOYT Project).

Authors' address: Laboratory of Inertial Technology, Department of Automatic Control, College of Mechatronics and Automation, National University of Defense Technology, Changsha, Hunan, P. R. China, 410073 (E-mail: yuanx_wu@hotmail.com).

II. UNSCENTED TRANSFORMATION

Consider a one-step nonlinear transformation with additive noise

$$y = f(x) + w \tag{1}$$

where x is an $n\times 1$ random vector with mean $\hat{x}$ and covariance P_x and w is an $m\times 1$ zero-mean noise vector with covariance Q that is uncorrelated with x. The problem is to calculate the mean $\hat{y}$ and covariance P_y of y. Note that here either x or w is not restricted to be Gaussian as long as they are well characterized by mean and covariance information. Equation (1) can be reformulated through the state augmentation method as

$$y^a = f^a(x^a) \tag{2}$$

where the augmented random vector is $x^a = \begin{bmatrix} x^T & w^T \end{bmatrix}^T$ and the new nonlinear transformation is defined as $f^a(x^a) = f^a\left(\begin{bmatrix} x^T & w^T \end{bmatrix}^T\right) = f(x) + w$. The problem now is to calculate the mean $\hat{y}^a$ and covariance P_{y^a} of y^a.

A. The Non-augmented UT

1) The random vector x is approximated by $2n+1$ symmetric sigma points

$$\begin{aligned} &\chi_0 = \hat{x} && W_0 = \kappa/(n+\kappa) \\ &\chi_i = \hat{x} + \left(\sqrt{(n+\kappa)P_x}\right)_i && W_i = 1/2(n+\kappa) \\ &\chi_{i+n} = \hat{x} - \left(\sqrt{(n+\kappa)P_x}\right)_i && W_{i+n} = 1/2(n+\kappa) \\ &i = 1,\ldots,n \end{aligned} \tag{3}$$

where $\left(\sqrt{P}\right)_i$ is the i^{th} column of the matrix square root of P and W_i is the weight associated with the i^{th} sigma point. The scalar κ is a scaling parameter which is usually set to 0 or $3-n$ [2, 4]. Note that if κ is set to 0, the sigma points and their weights will be related to n, the dimension of x. $\kappa = 3-n$ is selected so that the fourth-order moment information is mostly captured in the true Gaussian case [2]. In general other choices of κ would lead to better or worse results depending on specific characteristics of the integrand [7].

2) Instantiate each point through the function to yield a set of transformed sigma points

$$\gamma_i = f(\chi_i). \tag{4}$$

3) The mean $\hat{y}$ is given by the weighted average of the transformed points

$$\hat{y} = \sum_{i=0}^{2n} W_i \gamma_i. \tag{5}$$

4) The covariance P_y is the weighted outer product of the transformed points plus the noise covariance

$$P_y = \sum_{i=0}^{2n} W_i (\gamma_i - \hat{y})(\gamma_i - \hat{y})^T + Q. \tag{6}$$

B. The Augmented UT

1) The augmented random vector x^a are approximated by $2(n+m)+1$ symmetric sigma points

$$\begin{aligned}
&\chi_0^a = \hat{x}^a && W_0^a = \kappa^a / (n+m+\kappa^a) \\
&\chi_i^a = \hat{x}^a + \left(\sqrt{(n+m+\kappa^a)P_{x^a}}\right)_i && W_i^a = 1/2(n+m+\kappa^a) \\
&\chi_{i+n+m}^a = \hat{x}^a - \left(\sqrt{(n+m+\kappa^a)P_{x^a}}\right)_i && W_{i+n+m}^a = 1/2(n+m+\kappa^a) \\
&i = 1,\ldots,n+m
\end{aligned} \tag{7}$$

where the weight W_i^a and scalar κ^a are counterparts of W_i and κ in (3). Note that $\hat{x}^a = \begin{bmatrix} \hat{x}^T & 0_{1\times m} \end{bmatrix}^T$ and

$$\begin{aligned}
\left(\sqrt{P_{x^a}}\right)_i &= \left(\sqrt{\begin{bmatrix} P_x & 0_{n\times m} \\ 0_{m\times n} & Q \end{bmatrix}}\right)_i \\
&= \begin{cases} \begin{bmatrix} \left(\sqrt{P_x}\right)_i \\ 0_{m\times 1} \end{bmatrix}, & i = 1,\ldots,n \\ \begin{bmatrix} 0_{n\times 1} \\ \left(\sqrt{Q}\right)_{i-n} \end{bmatrix}, & i = n+1,\ldots,n+m \end{cases}
\end{aligned} \tag{8}$$

Substituting (8) into (7) yields

$$\begin{aligned}
&\chi_0^a = \begin{bmatrix} \hat{x} \\ 0_{m\times 1} \end{bmatrix} && W_0^a = \kappa^a / (n+m+\kappa^a), \\
&\chi_i^a = \begin{bmatrix} \hat{x} + \left(\sqrt{(n+m+\kappa^a)P_x}\right)_i \\ 0_{m\times 1} \end{bmatrix} && W_i^a = 1/2(n+m+\kappa^a) \\
&\chi_{i+n}^a = \begin{bmatrix} \hat{x} - \left(\sqrt{(n+m+\kappa^a)P_x}\right)_i \\ 0_{m\times 1} \end{bmatrix} && W_{i+n}^a = 1/2(n+m+\kappa^a) \\
&\chi_{j+2n}^a = \begin{bmatrix} \hat{x} \\ \left(\sqrt{(n+m+\kappa^a)Q}\right)_j \end{bmatrix} && W_{j+2n}^a = 1/2(n+m+\kappa^a) \\
&\chi_{j+2n+m}^a = \begin{bmatrix} \hat{x} \\ -\left(\sqrt{(n+m+\kappa^a)Q}\right)_j \end{bmatrix} && W_{j+2n+m}^a = 1/2(n+m+\kappa^a) \\
&i = 1,\ldots,n \quad j = 1,\ldots,m
\end{aligned} \tag{9}$$

2) Instantiate each point through the new function to yield a set of transformed sigma points

$$\gamma_i^a = f^a(\chi_i^a) \tag{10}$$

3) The mean $\hat{y}^a$ is the weighted average of the transformed points

$$\hat{y}^a = \sum_{i=0}^{2(n+m)} W_i^a \gamma_i^a \tag{11}$$

Substituting (9) and (10) into (11) yields

$$\begin{aligned}
\hat{y}^a &= W_0^a f^a(\chi_0^a) + \sum_{i=1}^{n} W_i^a \left[f^a(\chi_i^a) + f^a(\chi_{i+n}^a) \right] \\
&\quad + \sum_{i=1}^{m} W_i^a \left[f^a(\chi_{i+2n}^a) + f^a(\chi_{i+2n+m}^a) \right] \\
&= W_0^a f(\hat{x}) + W_i^a \sum_{i=1}^{n} \left[f\left(\hat{x} + \left(\sqrt{(n+m+\kappa^a)P_x}\right)_i\right) \right. \\
&\quad \left. + f\left(\hat{x} - \left(\sqrt{(n+m+\kappa^a)P_x}\right)_i\right) \right] + 2mW_i^a f(\hat{x}) \\
&= \frac{m+\kappa^a}{n+m+\kappa^a} f(\hat{x}) + \frac{1}{2(n+m+\kappa^a)} \\
&\quad \sum_{i=1}^{n} \left[f\left(\hat{x} + \left(\sqrt{(n+m+\kappa^a)P_x}\right)_i\right) + f\left(\hat{x} - \left(\sqrt{(n+m+\kappa^a)P_x}\right)_i\right) \right]
\end{aligned} \tag{12}$$

As compared with (5), the following equation

$$\begin{aligned}
\hat{y}^a \equiv \hat{y} = \sum_{i=0}^{2n} W_i \gamma_i &= \frac{\kappa}{n+\kappa} f(\hat{x}) + \frac{1}{2(n+\kappa)} \\
&\quad \sum_{i=1}^{n} \left[f\left(\hat{x} + \left(\sqrt{(n+\kappa)P_x}\right)_i\right) + f\left(\hat{x} - \left(\sqrt{(n+\kappa)P_x}\right)_i\right) \right]
\end{aligned} \tag{13}$$

is satisfied if and only if

$$n + \kappa = n + m + \kappa^a \triangleq C. \tag{14}$$

That is to say, the sums $n+\kappa$ and $n+m+\kappa^a$ are identical and independent of the state dimension.

4) The covariance P_y^a is the weighted outer product of the transformed points

$$\begin{aligned}
P_y^a &= \sum_{i=0}^{2(n+m)} W_i^a (\gamma_i^a - \hat{y}^a)(\gamma_i^a - \hat{y}^a)^T \\
&= \sum_{i=0}^{2n} W_i^a (\gamma_i^a - \hat{y}^a)(\gamma_i^a - \hat{y}^a)^T + \sum_{i=2n+1}^{2(n+m)} W_i^a (\gamma_i^a - \hat{y}^a)(\gamma_i^a - \hat{y}^a)^T
\end{aligned} \tag{15}$$

With (14) assumed, the first term on the right side becomes

$$\begin{aligned}
\sum_{i=0}^{2n} W_i^a (\gamma_i^a - \hat{y}^a)(\gamma_i^a - \hat{y}^a)^T &= \frac{C-m-n}{C} \left(f(\hat{x}) - \hat{y}\right)\left(f(\hat{x}) - \hat{y}\right)^T \\
&\quad + \sum_{i=1}^{2n} W_i \left(f(\chi_i) - \hat{y}\right)\left(f(\chi_i) - \hat{y}\right)^T,
\end{aligned} \tag{16}$$

and the second term is

$$\sum_{i=2n+1}^{2(n+m)} W_i^a \left(\gamma_i^a - \hat{y}^a\right)\left(\gamma_i^a - \hat{y}^a\right)^T$$
$$= \frac{1}{2C}\sum_{i=1}^{m}\left\{\left[f(\hat{x}) - \hat{y} + \left(\sqrt{CQ}\right)_i\right]\left[f(\hat{x}) - \hat{y} + \left(\sqrt{CQ}\right)_i\right]^T\right.$$
$$\left. + \left[f(\hat{x}) - \hat{y} - \left(\sqrt{CQ}\right)_i\right]\left[f(\hat{x}) - \hat{y} - \left(\sqrt{CQ}\right)_i\right]^T\right\} \quad (17)$$
$$= \frac{1}{2C}\sum_{i=1}^{m} 2\left\{\left[f(\hat{x}) - \hat{y}\right]\left[f(\hat{x}) - \hat{y}\right]^T + C\left(\sqrt{Q}\right)_i\left(\sqrt{Q}\right)_i^T\right\}$$
$$= \frac{m}{C}\left(f(\hat{x}) - \hat{y}\right)\left(f(\hat{x}) - \hat{y}\right)^T + Q.$$

Therefore

$$P_y^a = \sum_{i=0}^{2n} W_i \left(f(\chi_i) - \hat{y}\right)\left(f(\chi_i) - \hat{y}\right)^T + Q \equiv P_y. \quad (18)$$

In summary, the UT is configurable with the free parameter κ. Except the condition that (14) is satisfied the non-augmented UT will be different from the augmented UT. Coincidently, the common version of the UT selects $C = 3$ [2, 8], which leads to equivalence of the non-augmented and augmented UTs. In this case, $\kappa = 3 - n$ and $\kappa^a = 3 - n - m$. For other versions, such as the simplex UT ([2], Appendix III) and the UT that uses $\kappa = 0$ ([9] and [2], Section III), it will be another story. The discussions in the next section are made about the common UT.

III. Discussions on Unscented Kalman Filter

The UKF is a straightforward extension of the UT to the recursive estimation [1, 2]. For the sake of brevity, we prefer to treat the UT as a "black box" rather than get involved in details again as in Section II.

The prediction of the resulting UKF, whether based on the non-augmented UT or the augmented UT, consists of two concatenated UTs: one for the process function (the 1st UT) directly followed by the other for the measurement function (the 2nd UT). The 2nd UT for the measurement function makes a difference between the non-augmented UKF and augmented UKF. The non-augmented UKF has to[1] redraw a new set of sigma points to incorporate the effect of additive process noise ([4], Table 7.3.2). Specifically, readers are referred to the addition of the noise covariance in (6). Regarding the augmented UKF, the propagated sigma points in the 1st UT for the process function can be reused to propagate further through the measurement equation ([2], Step 5 of Fig. 7). By doing so, the computation of redrawing sigma points is spared and, more importantly, the odd-moment information is well propagated throughout one filtering recursion. In contrast, because it is a must to use the redrawn symmetric sigma points the non-augmented UKF is unable to propagate odd-moment information. Although the propagated sigma points of the 1st UT do capture the odd-moment information, the indispensable regeneration of a new sigma set for the 2nd UT interrupts its propagation. Expectably, the augmented UKF would be equivalent to the non-augmented UKF if the odd-moment information was intentionally abandoned through redrawing a new sigma set in the 2nd UT.

Referring back to Section II, the principle of the UT is to capture the first two moments of the random vector (x or x^a) via a set of sigma points. However, it should be made clear that with another set of sigma points capturing extra statistical information other than mean and covariance instead, the UT would yield better results. As a matter of fact, this is a natural conclusion that can be readily deduced from the Monte-Carlo method [10]. Therefore, the difference in drawing sigma points between the two UKFs generally favors the augmented UKF in that the extra odd-order moment information is captured by the nonlinearly transformed sigma points in the 1st UT and propagated throughout the whole recursion. This statement will be supported by a representative example in the next section.

Remark: It may be possible to prove that the augmented UKF accurately captures more statistical information than the non-augmented UKF. Additionally for the UKFs based on the other UTs, such as the simplex UT, the above analyses would have been even cumbersome, if not impossible.

IV. Example

Both non-augmented and augmented UKFs are applied to the univariate nonstationary growth model (UNGM), which is very popular in econometrics and has been previously used in [11-14]. This model is highly nonlinear and is bimodal in nature. The discrete-time dynamic system equation for this model can be written as

$$\begin{aligned} x_n &= 0.5x_{n-1} + 25\frac{x_{n-1}}{1 + x_{n-1}^2} + 8\cos\left(1.2(n-1)\right) + u_n \\ y_n &= \frac{x_n^2}{20} + v_n, \qquad n = 1, \ldots, N \end{aligned} \quad (19)$$

where the process noise u_n and measurement noise v_n are both Gaussian noises with zero mean and unity variance. The reference data were generated using $x_0 = 0.1$ and $N = 500$. The likelihood of measurement conditioned on the system state has bimodal nature when $y_n > 0$ and is unimodal when $y_n < 0$. The bimodality makes this problem more difficult to address using conventional methods[2].

[1] Unfortunately, the sigma points redrawing were neglected in addressing practical systems [5, 6]. It was argued that the propagated sigma points might be augmented with additional points derived from the matrix square root of the process noise covariance (see fn. 6, [4]). Although careful examination was not presented, the arguments there mentioned the possibility of capturing odd-moment information.

[2] Admittedly, the UKF's performance for this example may be unsatisfactory since a Gaussian approximation is implicitly made to the

The initial conditions were $\hat{x}_0 = 0, P_0 = 1$. The performance of the two UKFs was compared using the mean squared error (MSE) defined by

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \hat{x}_i\right)^2. \quad (20)$$

A large number of computer runs were carried out. The state estimates for the first 50 points in an exemplar run are shown in Fig. 1. The bimodality is so influential that both UKFs go in the opposite direction when $x_n^2/20$ is small. This behavior is often observed in all runs. Figure 1 indicates that the augmented UKF has a better estimation than the non-augmented UKF does. In Fig. 2-3, the errors and estimated 3σ confidence intervals are plotted for the non-augmented UKF and the augmented UKF, respectively. It is shown that the augmented UKF performs noticeably better, i.e., more accurate and more consistent. As discussed above, the superiorities of the augmented UKF are mainly owed to its capability in capturing and propagating odd-moment information throughout one filtering recursion. The augmented UKF with an unnecessary sigma set regeneration inserted was also simulated and yielded, as expected, the same results as the non-augmented UKF. Figure 4 plots the MSEs for 50 random runs. The non-augmented UKF's MSE is nearly one times larger than that of the augmented UKF. The mean and standard variance of MSEs are plotted in Fig. 5, which shows that the standard variance is also evidently smaller for the augmented UKF.

The computation time of the non-augmented UKF is half of that cost by the augmented UKF in our simulation. One of the main computational disadvantages of the augmented UKF is that there are more sigma points that have to be propagated through the nonlinear process and measurement equations. Referring to (3) and (9), however, the arithmetic operations for the last extra $2m$ sigma points in (9) can be efficiently reduced using the results of the first sigma point χ_0^a. On the other hand, for the dynamic system with uncorrelated process and measurement noises, in the above example for instance, the computation complexity can be further lowered by calculating low-dimensional matrix square roots instead. Equipped with the optimized implementation, the augmented UKF promises to yield better performance with comparable computational expense.

V. Conclusions

In this paper, we have analyzed and compared two alternative versions of UT-based filters for the nonlinear dynamic system with additive noises: the non-augmented UKF and the augmented UKF. In a rigorous manner, we proved that the non-augmented UT is equivalent to the augmented counterpart only if $n + \kappa = \text{const}$ is satisfied. We pointed out that the basic difference between the augmented and non-augmented UKFs is that the former draws sigma points only once in a recursion while the latter has to redraw a new set of sigma points to incorporate the effect of additive process noise. This difference generally favors the augmented UKF in that the odd-order moment information is captured by the propagated sigma points and well propagated within one recursion. On the other hand, if a new (but unnecessary) set of sigma points were redrawn in the augmented UKF, it would be equivalent to and yield exactly the same results as the non-augmented UKF. The simulation results of a representative example agree well with our conclusions.

Acknowledgment

The first author would like to thank Dr. Simon J. Julier for his inspiring discussions on several related topics.

References:

[1] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45, pp. 477-482, 2000.

[2] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, pp. 401-422, 2004.

[3] S. Julier and J. Uhlmann, "Comment on 'A new method for the nonlinear transformation of means and covariances in filters and estimators' - Reply," *IEEE Transactions on Automatic Control*, vol. 47, pp. 1408-1409, 2002.

[4] E. A. Wan and R. v. d. Merwe, "The unscented Kalman filter," in *Kalman Filtering and Neural Networks, Chapter 7*, S. Haykin, Ed.: Wiley Publishing, 2001, pp. 221-280.

[5] A. Farina, B. Ristic, and D. Benvenuti, "Tracking a ballistic target: Comparison of several nonlinear filters," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, pp. 854-867, 2002.

[6] J. L. Crassidis and F. L. Markley, "Unscented filtering for spacecraft attitude estimation," *Journal of Guidance Control and Dynamics*, vol. 26, pp. 536-542, 2003.

[7] U. N. Lerner, "Hybrid Bayesian networks for reasoning about complex systems," PhD thesis, Stanford University, 2002.

[8] R. v. d. Merwe, "Sigma-point Kalman filters for probabilistic inference in dynamic state-space models," in *Electrical and Computer Engineering*: PhD thesis, Oregon Health & Science University, 2004.

[9] J. K. Uhlmann, "Simultanenous map building and localization for real time applications." Oxford U.K.: Transfer thesis, Univ. Oxford, 1994.

[10] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, pp. 174-188, 2002.

[11] G. Kitagawa, "Non-Gaussian state-space modeling of nonstationary time-series," *Journal of the American Statistical Association*, vol. 82, pp. 1032-1063, 1987.

[12] N. Gordon, D. Salmond, and C. Ewing, "Bayesian state estimation for tracking and guidance using the bootstrap filter," *Journal of Guidance Control and Dynamics*, vol. 18, pp. 1434-1443, 1995.

[13] E. R. Beadle and P. M. Djuric, "A fast weighted Bayesian bootstrap filter for nonlinear model state estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, pp. 338-343, 1997.

[14] J. H. Kotecha and P. A. Djuric, "Gaussian particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, pp. 2592-2601, 2003.

posterior distribution. But it did not hinder the comparison of two versions of UKFs hereafter.

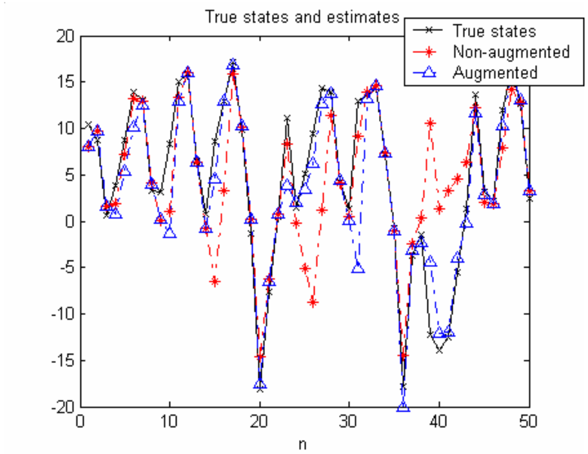

Figure. 1. True states and estimates: non-augmented UKF vs. augmented UKF.

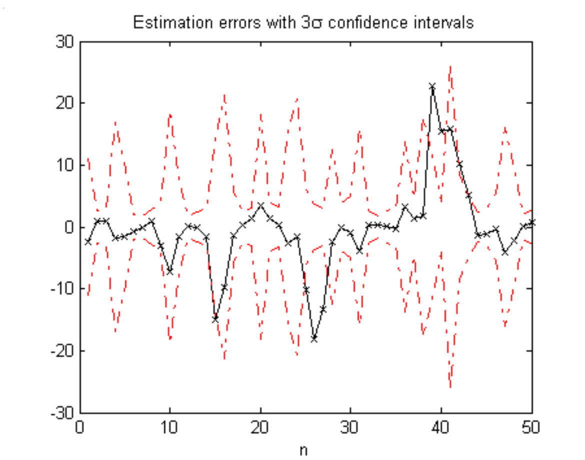

Figure. 2. Estimation errors and 3σ confidence intervals for the non-augmented UKF.

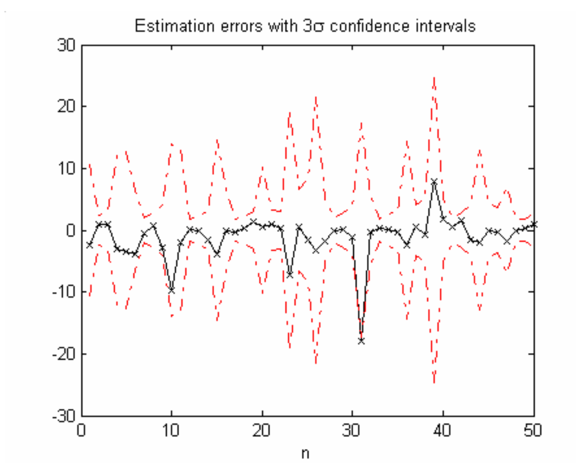

Figure. 3. Estimation errors and 3σ confidence intervals for the augmented UKF.

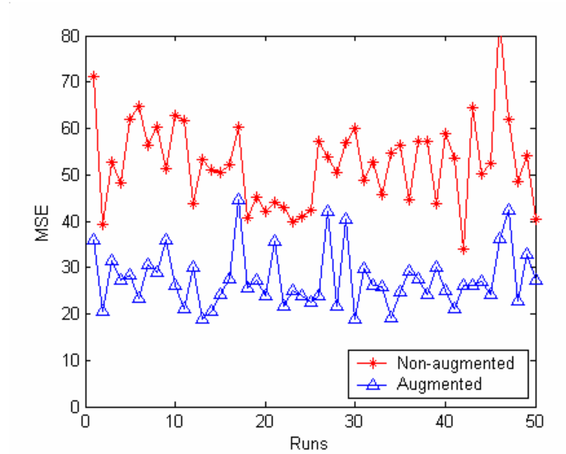

Figure. 4. Performance comparison of the non-augmented and augmented UKF filters. The MSEs across 50 random runs.

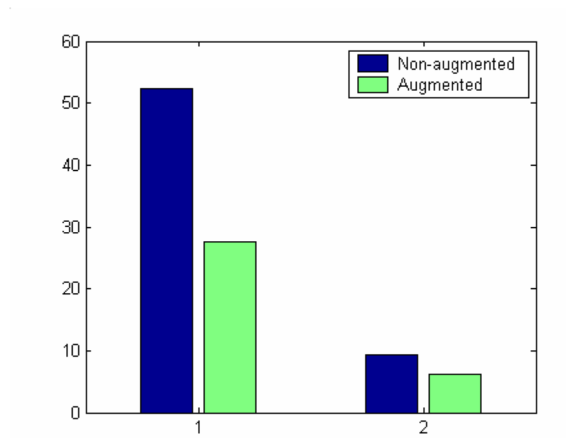

Figure. 5. Performance comparison of the non-augmented and augmented UKF filters. The mean and standard variance of MSEs.