SYM-AM-24-075

Excerpt from the Proceedings of the Twenty-First Annual Acquisition Research Symposium

Acquisition Research: Creating Synergy for Informed Change

May 8–9, 2024

Published: April 30, 2024

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

The research presented in this report was supported by the Acquisition Research Program at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).

ACQUISITION RESEARCH PROGRAM
DEPARTMENT OF DEFENSE MANAGEMENT
NAVAL POSTGRADUATE SCHOOL

Statistical Procedures for Validation of a Computer Model with Multimodal Output When the Observation is a Single Time Series

Patricia (P.A.) Jacobs—is a retired Distinguished Professor of Operations Research at the Naval Postgraduate School. She is a fellow of the Royal Statistical Society (RSS), the American Statistical Association (ASA), and the American Association for the Advancement of Science (AAAS). [pajacobs@nps.edu]

Abstract

The AEGIS combat system is a system of systems (SoS) in which the systems and tactics for their use continually evolve. The Combat System Test Bed (CSTB) is a federation of computer simulation models (SMs) that represents the performance of the AEGIS SoS. Physical test events are conducted to assess how well the CSTB represents the performance of the evolving AEGIS SoS. Test event data, including time series, are compared to SMs' output. In this paper, simulation is used to study the efficacy of statistical procedures, including one currently in use, to obtain statistical evidence that a model's multimodal distribution does not include that of a time series observation.

Introduction

The AEGIS combat system is a system of systems (SoS) in which the systems and tactics for their use continually evolve. The Combat System Test Bed (CSTB) is a federation of computer simulation models (SMs) that represent the performance of the AEGIS SoS. Physical test events are conducted to assess how well the CSTB represents the performance of the evolving AEGIS SoS. Measurements from the test events are compared to CSTB output. Statistical evidence that the CSTB does not well summarize the test event's measurements can lead to CSTB modification or enhancement.

Su et al. (2022) present results comparing four statistical procedures, including one currently used, for computer model validation in the case that the test event measurement is a single time series. In this working paper, we consider additional statistical procedures to validate a computer model in the case the test event measurement is a single time series and the model output is multimodal. The validation of a computer model for test events resulting in a time series has been of interest in several areas, including meteorology (cf. Gneiting et al., 2008) and economics (cf. Diebold et al., 1998, 1999).

The next section, Procedures to Assess How Well a Multimodal Model Distribution Summarizes One Observed Time Series, presents three procedures to assess how well a computer model with multimodal output summarizes one observation time series. The procedures summarize the model replications at each time and compare the observation time series to the summaries to obtain statistical evidence that the model's multimodal distribution does not include that of the observation. One procedure considered is the currently used two-sided hypothesis procedure described in Su et al. (2022) that uses the sample mean and sample variance of the model replications at each time to create 2-sided confidence intervals; we call this procedure the 2.5-sigma procedure. The second procedure uses a 2-sided 99% percentile confidence interval of the model replications at each time. The third type of procedure uses a kernel density estimate (KDE) of the model's density function at each time and the value of the model's KDE of the observation at that time. The third section, Simulation Study, presents the results of a simulation study of the efficacy of the procedures when the model has a multimodal mixture distribution.

The simulation results suggest that the 2.5-sigma procedure is among the least likely to result in a false positive; a false positive occurs when the procedure results in statistical evidence that the multimodal model mixture distribution does not include the distribution of the observation when the model's distribution includes that of the observation. However, the 2.5-sigma procedure is among the least likely of the procedures considered to result in correct statistical evidence that the model's distribution does not include that of the observation. The percentile confidence procedure is the most likely to result in a false positive; it is more likely than the 2.5-sigma procedure to result in correct statistical evidence that multimodal model distribution does not include that of the observation. The KDE procedure is better at balancing incorrect and correct statistical evidence that the model distribution does not include that of the observation.

Procedures to Assess How Well a Multimodal Model Distribution Summarizes One Observed Time Series

Let $X(t_k)$ be the value of the test event's observed time series at time $t_k, k = 1,...,n_T$. Let $Y_i(t_k)$ be the value of the *ith* model replication at time t_k for $i = 1,...,n_M$ where n_M is the number of model replications.

The 2.5-Sigma Procedure and Percentile Interval Procedure

The two-sided hypothesis procedure of Su et al. (2022) is as follows: for each time t_k, the sample mean and sample standard deviation of the n_M model replications $\{Y_i(t_k); i = 1,....,n_M\}$ are computed. The 2.5-sigma confidence interval has lower bound equal to the sample mean minus 2.5 times the sample standard deviation and upper bound equal to the sample mean plus 2.5 times the sample standard deviation. There is said to be statistical evidence that the model's multimodal distribution does not contain that of the observation if the fraction of times the observed time series is outside of the confidence intervals is greater than 0.1. We call this two-sided hypothesis procedure the 2.5-sigma procedure. If $\{Y_i(t_k); i = 1,...,n_M\}$ are independent and identically distributed having a Gaussian distribution, then the 2.5-sigma confidence interval is an approximate 99% prediction interval. This procedure is an approved simulation validation method at the Johns Hopkins University Applied Physics Laboratory (APL; cf. Su et al., 2022).

For each time t_k, let $q_L(t_k)$ (respectively $q_H(t_k)$) be the 0.005 quantile (respectively 0.995 quantile) of the n_M model replications at time t_k, $\{Y_i(t_k); i = 1,...,n_M\}$. The 99% percentile interval is $[q_L(t_k), q_H(t_k)]$. There is said to be statistical evidence that the model's multimodal distribution does not contain that of the observation if the fraction of times the observed time series is outside of the confidence intervals is greater than 0.1.

Gaussian Kernel Density Estimation (KDE) Procedure

Introduction

Let $F_{Y(t)}(y) = P\{Y(t) \le y\} = \int_{-\infty}^{y} f_{Y(t)}(z)\,dz$ for $-\infty < y < \infty$ be the model's probability cumulative distribution function for time t; $f_{Y(t)}(\bullet)$ is the probability density function of the model at time *t*. A Gaussian kernel density smoothing of the model replications at time t,

$\kappa_{Y(t)}(\bullet)$, is an estimate of $f_{Y(t)}(\bullet)$ (cf. Silverman, 1986). Let $\kappa_{Y(t)}(X(t))$ be the value of the model's kernel density estimate (KDE) evaluated at the value of the observed time series at time t. For small positive *h*, $\kappa_{Y(t)}(X(t))h$ is an estimate of $F_{Y(t)}(X(t)+(h/2))-F_{Y(t)}(X(t)-(h/2))$; the model's probability that the value of the observed time series at time t occurs in the interval $[X(t)-(h/2), X(t)+(h/2)]$. There is statistical evidence that the distribution of the observed time series is not included in that of the model if the model's KDE for the observation at time t_k, $\kappa_{Y(t_k)}(X(t_k))$, is too small for too many times $t_k, k=1,...,n_T$.

A Kernel Density Estimate (KDE) Procedure

For the *ith* model replication at time t_k, $Y_i(t_k)$, let $\kappa_M(\square, i, t_k)$ be the Gaussian kernel density estimate (KDE) obtained using model replications at time t_k **without** the ith model replication, $\{Y_j(t_k); j \in \{1,...,i-1,i+1,...,n_M\}\}$. Let $\kappa_M(Y_i(t_k); i, t_k)$ be the value of the KDE for time t_k at $Y_i(t_k)$. Let $q_M(t_k;\alpha)$ be the $\alpha-$quantile of $\{\kappa_M(Y_i(t_k); i, t_k); i=1,...,n_M\}$, the KDEs of the n_M model replication values at time t_k. Let $Q_M(i;\alpha)=\sum_{k=1}^{n_T} I(\kappa_M(Y_i(t_k); i, t_k) < q_M(t_k;\alpha))$, the number of times the value of the KDEs for the *ith* model replication are less than the model values' KDE $\alpha-$quantiles; $I(A)$ equals 1 if event *A* occurs and 0 otherwise.

For each time t_k, let $\kappa_A(\square, t_k)$ be the KDE obtained using all the model replications at time t_k, $\{Y_i(t_k); i=1,...,n_M\}$. Let $\kappa_O(t_k)=\kappa_A(X(t_k); t_k)$, the value of the KDE, $\kappa_A(\square, t_k)$, at $X(t_k)$, the value of the observed time series at time t_k. Let $Q_O(\alpha)=\sum_{k=1}^{n_T} I(\kappa_O(t_k) < q_M(t_k;\alpha))$, the number of times the observed time series KDEs are less than the model values' KDE $\alpha-$quantiles. $Q_O(\alpha)$ is compared to a lower bound obtained from the model replications. The most conservative lower bound considered is $B_M(\alpha;1)=\max_{i=1,...,n_M}(Q_M(i;\alpha))$, the maximum number of times a model's replication value KDEs are less than the model values' KDE $\alpha-$quantiles. Other considered lower bounds are $B_M(\alpha;n)=$ the nth largest number of times a model's replication value KDEs are less than the model values' KDE $\alpha-$quantiles for *n*=2, 3, 4.

For a chosen lower bound, there is statistical evidence that the multimodal distribution of the model does not include that of the observation if the number of times the observed time series KDEs are less than the model values' KDE $\alpha-$quantiles, $Q_O(\alpha)$, is greater than the chosen lower bound.

Simulation Study

There are 500 simulation replications. Each simulation replication generates 300 model replications. The model replications have a mixture distribution. Nine observation time series

each having different parameter values are also generated. All the considered model validation procedures compare each of the 9 observed time series to the same model replications.

Simulated Model and Observation

The ith replication of the simulation model satisfies

$$\boldsymbol{Y}_i\left(t_k\right)=1-\exp\left\{-\boldsymbol{\theta}_i\left(M\right)\times\frac{k}{100}\right\}+\boldsymbol{\varepsilon}_i\left(k;M\right)\ \text{for}\ k=1,...,100 \tag{1}$$

where $\left\{\boldsymbol{\theta}_i\left(M\right);i=1,...,n_M\right\}$ are independent identically distributed random variables having a mixture distribution; with probability 0.5, $\boldsymbol{\theta}_i\left(M\right)$ has a gamma distribution with shape parameter $\beta_i\left(M\right)=50$ and mean $\beta_i\left(M\right)/\alpha_i\left(M\right)=2$; with probability 0.5, $\boldsymbol{\theta}_i\left(M\right)$ has a gamma distribution with shape parameter $\beta_i\left(M\right)=50$ and mean $\beta_i\left(M\right)/\alpha_i\left(M\right)=6$; $\left\{\boldsymbol{\varepsilon}_i\left(k;M\right);k=1,...,100\right\}$ are independent identically distributed random variables having a normal distribution with mean $\mu_M=0$ and standard deviation $\sigma_M=0.02$.

The observed time series satisfies

$$\boldsymbol{X}\left(t_k\right)=1-\exp\left\{-\boldsymbol{\theta}\left(O\right)\times\frac{k}{100}\right\}+\boldsymbol{\varepsilon}\left(k;O\right)\ \text{for}\ k=1,...,100 \tag{2}$$

where $\boldsymbol{\theta}\left(O\right)$ is a random variable having a gamma distribution with shape parameter $\beta\left(O\right)=50$ and mean $\frac{\beta\left(O\right)}{\alpha\left(O\right)}$; $\left\{\boldsymbol{\varepsilon}\left(k;O\right);k=1,...,100\right\}$ are independent identically distributed random variables having a normal distribution with mean $\mu_O=0$ and variance σ_O^2. The values of the parameters, $\left(\frac{\beta\left(O\right)}{\alpha\left(O\right)},\sigma_O\right)$, considered are (1, 0.02), **(2, 0.02)**, (2, 0.1), (3, 0.02), (4, 0.02), (5, 0.02), **(6, 0.02)**, (6, 0.1) and (10, 0.02); the bold values correspond to parameters included in the model mixture distribution.

Kernel Density Estimation

The software R function called density with Gaussian kernel and bandwidth ucv (unbiased cross validation) is used to obtain the kernel density estimates (cf. R Core Team, 2021).

Figure 1 displays fifty model replications as a function of time. Figure 2 displays a histogram of 300 model replications at time 0.5. Figure 3 displays the kernel density estimate using the model replications at time 0.5. The three Figures suggest that the model replications have 2 modes. Figure 4 displays the number of times the model replications value KDEs, $\left\{\kappa_M\left(Y_i\left(t_k\right);i,t_k\right);k=1,...,100\right\}$, are less than the model replication values' KDE 0.001 quantiles, $\left\{q_M\left(t_k;0.001\right);k=1,...100\right\}$;

$$Q_M\left(i;0.001\right)=\sum_{k=1}^{n_T}I\left(\kappa_M\left(Y_i\left(t_k\right);i,t_k\right)<q_M\left(t_k;0.001\right)\right)\ \text{for}\ i=1,...,300 \tag{3}$$

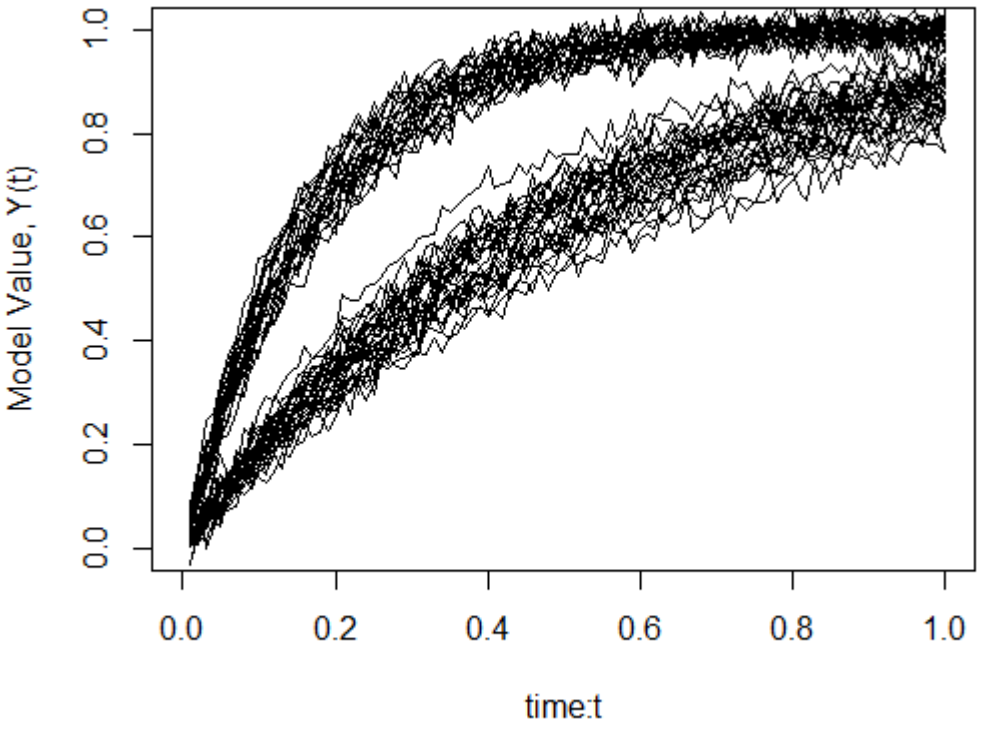

Figure 1. 50 Model Replications

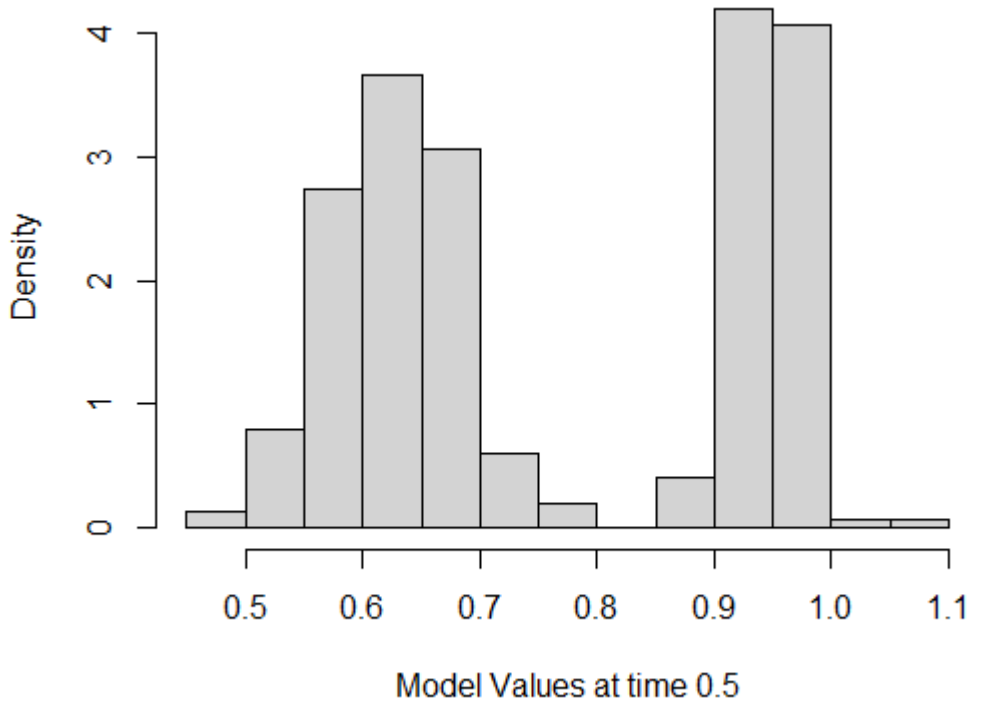

Figure 2. Model Replications at Time 0.5

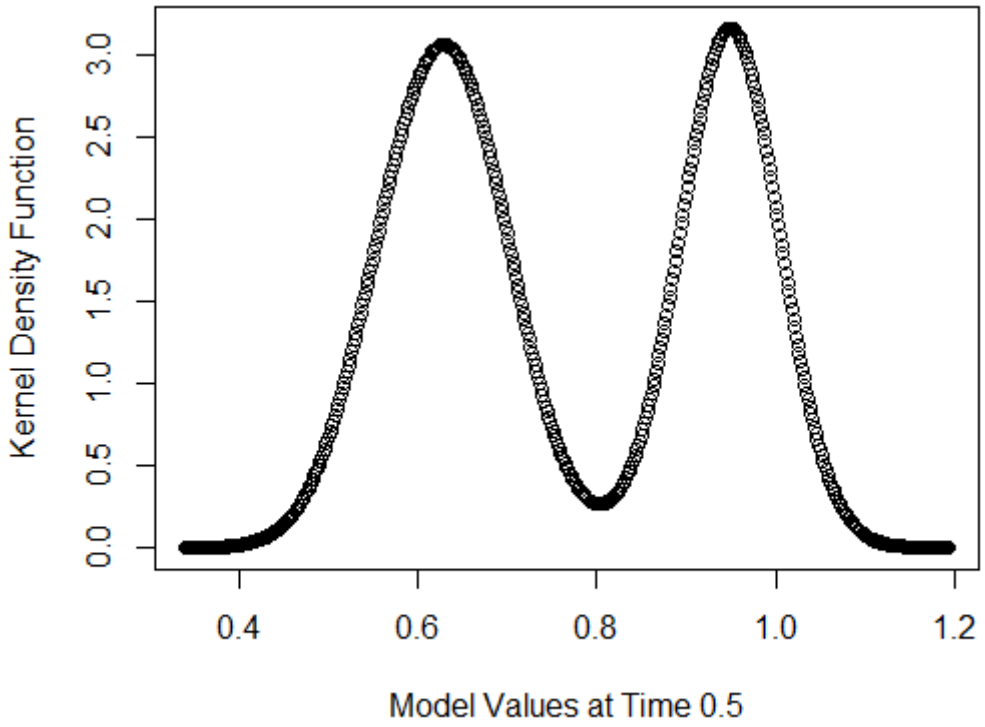

Figure 3. Model Replications at Time 0.5 Gaussian Kernel Density With Bandwidth=ucv (Unbiased Cross Validation)

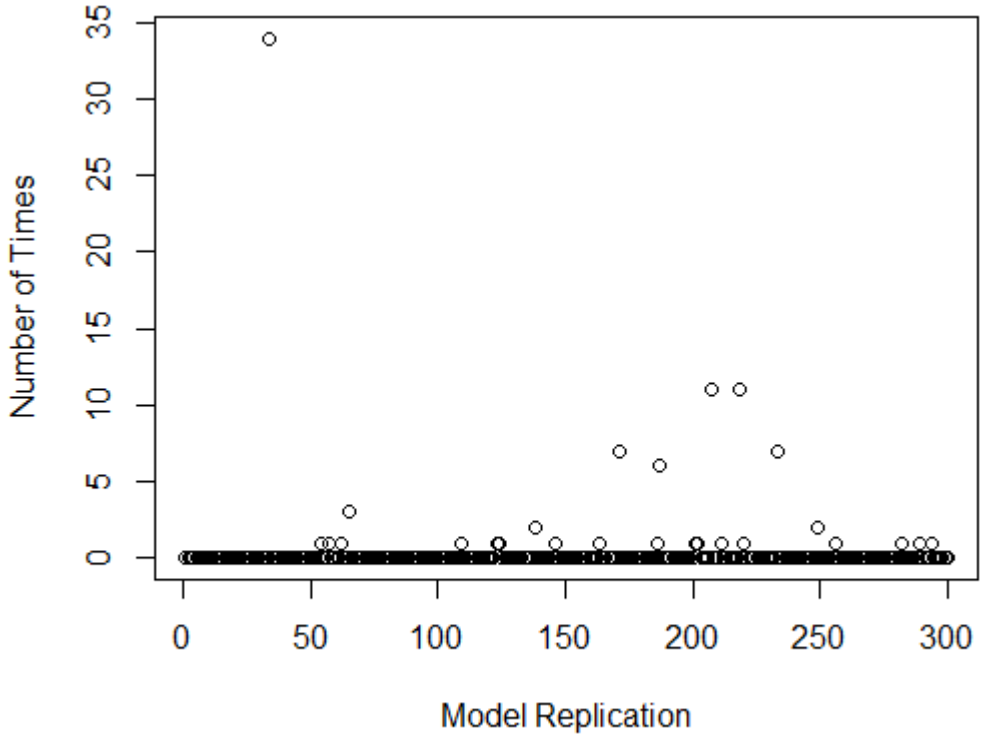

Figure 4. Number of Times Model Replication's Value Kernel Density Estimates (KDEs) Are Less Than the Model Values KDE 0.001 Quantiles

Simulation Results

The 2.5-sigma, 99% percentile confidence interval, and kernel density estimate (KDE) procedures are used to obtain statistical evidence that the multimodal mixture model distribution does not include that of the observation. Several values for the KDE procedure quantile and lower bound are considered; the quantile values are $\alpha \in \{0.01, 0.005, 0.001, 0.0005\}$; the lower bounds are the maximum number of times a model replication value KDEs are less than the model values' KDE $\alpha-$quantile, $B_M(\alpha;1)$; the second largest number of times, $B_M(\alpha;2)$; the third largest number of times, $B_M(\alpha;3)$; and the fourth largest number of times, $B_M(\alpha;4)$. There is statistical evidence that the mixture model distribution does not include that of the observation if the number of times the observation's KDEs are less than the model values' KDE quantiles is greater than the lower bound.

Table 1 displays the fraction of the 1,000 cases that result in a false positive (incorrect statistical evidence that the observation distribution is not included in that of the model when the model distribution includes that of the observation). Each of the 500 simulation replications has 2 cases in which the model distribution includes that of the observation; $\left(\frac{\alpha(O)}{\beta(O)}, \sigma_O\right) \in \{(2, 0.02), (6, 0.02)\}$.

Table 1. Fraction of Cases With Incorrect Statistical Evidence That the Multimodal Mixture Model Distribution Does NOT Include That of the Observation When the Model Distribution Does Include That of the Observation

Procedure		Fraction of Cases		Fraction of Cases		Fraction of Cases		Fraction of Cases
2.5-sigma		0.001						
99% Percentile Confidence Interval		0.040						
KDE								
	KDE Quantile α	Observation Number Greater Than Maximum Model Number, $B_M(\alpha;1)$		Observation Number Greater Than 2nd Largest Model Number, $B_M(\alpha;2)$		Observation Number Greater Than 3rd Largest Model Number, $B_M(\alpha;3)$		Observation Number Greater Than 4th Largest Model Number, $B_M(\alpha;4)$
	0.01	0.002		0.006		0.010		0.013
	0.005	0.001		0.003		0.003		0.004
	0.001	0.002		0.005		0.007		0.014
	0.0005	0.002		0.005		0.007		0.010

Table 2 displays the fraction of the 3,500 cases that result in correct statistical evidence that the observation's distribution is not included in that of the model; there are 7 such cases in each simulation replication.

Table 2. Fraction of Cases With Correct Statistical Evidence That the Multimodal Mixture Model Distribution Does NOT Include That of the Observation When the Model Distribution Does NOT Include That of the Observation

Procedure		Fraction of Cases		Fraction of Cases		Fraction of Cases		Fraction of Cases
2.5-sigma		0.173						
99% Percentile Confidence Interval		0.526						
KDE								
	KDE Quantile, α	Observation Number Greater Than Maximum Model Number, $B_M(\alpha;1)$		Observation Number Greater Than 2nd Largest Model Number, $B_M(\alpha;2)$		Observation Number Greater Than 3rd Largest Model Number, $B_M(\alpha;3)$		Observation Number Greater Than 4th Largest Model Number, $B_M(\alpha;4)$
	0.01	0.222		0.368		0.513		0.616
	0.005	0.174		0.300		0.438		0.544
	0.001	0.285		0.519		0.641		0.698
	0.0005	0.269		0.497		0.624		0.681

The results displayed in Tables 1 and 2 suggest the following concerning the ability of the considered statistical procedures to result in correct statistical evidence that the model distribution does not include that of the observation.

False Positive: Incorrect Statistical Evidence That the Mixture Model Distribution Does Not Include That of the Observation When the Model Distribution Does Include That of the Observation

The 2.5-sigma procedure and the KDE procedure with lower bound the maximum number of times a model replication's value KDEs are less than the model's value KDE 0.005-quantiles, $B_M(0.005;1)$, result in 1 false positive in the 500 simulation replications. The KDE procedure with 0.001- quantile and lower bound the maximum number of times a model replication's value KDEs are less than the model values' KDE quantiles, $B_M(0.001;1)$, results in 2 false positives. The KDE procedure with lower bounds $B_M(0.005;2)$ and $B_M(0.005;3)$ result in 3 false positives. The percentile confidence interval procedure results in the most false positives.

True Positive: Correct Statistical Evidence the Observation Distribution Is Not Included in That of the Model

The 2.5-sigma procedure and the KDE procedure with lower bound the maximum number of times a model replication's value KDEs are less than the model KDE values' 0.005-quantiles, $B_M(0.005;1)$, result in correct statistical evidence the model distribution does not include that of the observation in less than 18% of the cases. The KDE procedure with lower

bound $B_M(0.001,1)$ results in one more false positive than the 2.5-sigma procedure and results in correct statistical evidence that the model distribution does not include that of the observation in slightly over 28% of the cases. The KDE procedure with lower bound $B_M(0.005,3)$ results in 2 more false positives than the 2.5-sigma procedure but results in correct statistical evidence that the model distribution does not include that of the observation in slightly over 40% of the cases. The KDE procedure with lower bound $B_M(0.001;4)$, the fourth largest number of times a model replication value KDEs are less than the model values' KDE 0.001-quantiles, results in the most cases with correct statistical evidence the model distribution does not include that of the observation; however, it also results in the highest number of false positives.

Tables 3 and 4 display the fraction of the 500 simulation replications that result in statistical evidence that the observation distribution is not included in that of the model. Table 3 (respectively 4) displays the results using the 0.005 (respectively 0.001) model values' KDE quantile for the KDE procedure. The bold entries correspond to results in the case that the observation distribution is included in that of the mixture model.

Table 3. Fraction of Simulation Replications Resulting in Statistical Evidence That the Model Mixture Distribution Does Not Include That of the Observation

Model Values' KDE quantile $\alpha = 0.005$ Bold Entries: Observation Distribution is included in the Model Mixture Distribution							
Observation parameters		Procedure					
Gamma mean $\frac{50}{\alpha(O)}$	Normal standard deviation σ_O	2.5-Sigma	Percen-tile	KDE Obser-vation number > model's max $B_M(\alpha;1)$	KDE Obser-vation number > model's 2nd max $B_M(\alpha;2)$	KDE Obser-vation number > model's 3rd max $B_M(\alpha;3)$	KDE Obser-vation number > model's 4th max $B_M(\alpha;4)$
1	0.02	0.974	0.998	0.874	0.942	0.978	0.986
2	**0.02**	**0.002**	**0.042**	**0.002**	**0.006**	**0.006**	**0.008**
2	0.1	0.082	0.736	0.020	0.100	0.232	0.422
3	0.02	0	0	0.130	0.280	0.376	0.472
4	0.02	0	0	0.092	0.206	0.302	0.382
5	0.02	0	0.002	0.010	0.018	0.034	0.050
6	**0.02**	**0**	**0.038**	**0**	**0**	**0**	**0**
6	0.1	0.002	0.988	0.046	0.260	0.608	0.830
10	0.02	0.154	0.958	0.050	0.292	0.534	0.666

Table 4. Fraction of Simulation Replications Resulting in Statistical Evidence That the Model Mixture Distribution Does Not Include That of the Observation

Model Values' KDE quantile $\alpha = 0.001$ Bold Entries: Observation Distribution is included in the Model Mixture Distribution							
Observation parameters		Procedure					
Gamma mean $\frac{50}{\alpha(O)}$	Normal standard deviation σ_O	2.5-Sigma	Percen-tile	Obser-vation number > model's max $B_M(\alpha;1)$	Obser-vation number > model's 2nd max $B_M(\alpha;2)$	Obser-vation number > model's 3rd max $B_M(\alpha;3)$	Obser-vation number > model's 4th max $B_M(\alpha;4)$
1	0.02	0.974	0.998	0.920	0.986	0.990	0.996
2	**0.02**	**0.002**	**0.042**	**0.004**	**0.008**	**0.010**	**0.020**
2	0.1	0.082	0.736	0.106	0.402	0.708	0.878
3	0.02	0	0	0.242	0.450	0.554	0.608
4	0.02	0	0	0.182	0.374	0.464	0.532
5	0.02	0	0.002	0.018	0.048	0.060	0.064
6	**0.02**	**0**	**0.038**	**0**	**0.002**	**0.004**	**0.008**
6	0.1	0.002	0.988	0.264	0.744	0.956	0.994
10	0.02	0.154	0.958	0.266	0.626	0.758	0.814

The results displayed in Tables 3 and 4 suggest that the KDE procedure with lower bound the maximum number of times a model replication's value KDEs are less than that of the model values' KDE 0.001-quantiles, $B_M(0.001;1)$, is the best at balancing incorrect and correct statistical evidence the model distribution does not include that of the observation; it results in one more false positive than the 2.5-sigma procedure; it results in correct statistical evidence that the mixed model distribution does not include that of the observation in slightly more than 28% of the cases compared to the 2.5-sigma procedure's less than 18%. However, if one can tolerate 2 more false positives than the 2.5-sigma procedure, than the lower bound $B_M(0.005;3)$ results in correct statistical evidence that the model distribution does not include that of the observation in slightly over 40% of the cases.

The percentile confidence interval procedure is the most likely to result in correct statistical evidence that the observation distribution is not included in that of the model mixture distribution if the observed time series values tend to always lie above or always lie below the model replications. The 2.5-sigma procedure tends to result in correct statistical evidence that the observation distribution is not included in that of the model mixture distribution when the observed time series tends to lie above the model replications. The 2.5-sigma procedure and percentile confidence interval tend not to result in correct statistical evidence that the model mixture distribution does not include that of the observation when the observed time series tends to lie between the two model distribution modes. The KDE procedure can result in correct statistical evidence that the model distribution does not include that of the observation when the

observed time series tends to lie between the two model distribution modes and when the observed time series tends to always lie above or to always lie below the model replications.

Conclusions

The simulation results reported here, and the results of Su et al. (2002) suggest that the currently used 2.5-sigma procedure is unlikely to result in a false positive (incorrect statistical evidence that the mixture model distribution does not include that of the observation when the observation distribution is included in that of the model). However, it is also among the least likely of the procedures considered here to result in correct statistical evidence that the model distribution does not include that of the observation. The efficacy of the KDE procedure depends on the quantile and lower bound chosen. Increased ability to result in correct statistical evidence that the model distribution does not include that of the observation can be associated with an increased chance of incorrect statistical evidence that the model distribution does not include that of the observation when the model mixture distribution does include that of the observation. The simulation results presented here suggest that the KDE procedure with lower bound, $B_M\left(0.001;1\right)$, the maximum number of times a model replication's value KDEs are less than the model replication values' 0.001-quantile or lower bound $B_M\left(0.005;3\right)$ are also unlikely to result in a false positive but are more likely than the 2.5-sigma procedure to result in correct statistical evidence that the mixture model distribution does not include that of the observation.

References

Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, *39*, 863–883. https://www.sas.upenn.edu/~fdiebold/papers/paper16/paper16.pdf

Diebold, F. X., Hahn, J., & Tay, A. S. (1999). Multivariate density forecast evaluation and calibration in financial risk management: High-frequency returns on foreign exchange. *The Review of Economics and Statistics*, *81*(4), 661–673.

Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L., & Johnson, N. A. (2008, June). *Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds* (Technical Report No. 537). Department of Statistics, University of Washington.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. URI. http://www.R-project/org/.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall/CRC.

Su, S. Y., McCarty, S. K., Warfield, J. D., Jr., Uthoff, E. J., & Youngblood, S. M. (2022). Evaluation framework for assessing validation methods on modeling and simulation models. *Johns Hopkins APL Technical Digest*, *36*(3), 280–287.

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net