

Applicability evaluation of Akaike's Bayesian information criterion to covariance modeling in the cross-section adjustment method

Shuhei Maruyama^{1*}, Tomohiro Endo², and Akio Yamamoto²

¹Japan Atomic Energy Agency, Reactor Core and Plant System Evaluation Group, Fast Reactor Cycle System Research and Development Center, 4002 Narita-cho, Oarai-machi, Higashi-Ibaraki-gun, Ibaraki, Japan

²Nagoya University, Department of Applied Energy, Graduate School of Engineering, Furo-cho, Chikusa-ku, Nagoya, Japan

Abstract. The applicability of Akaike's Bayesian Information Criterion (ABIC) to the covariance modeling in the cross-section adjustment method has been investigated. In the conventional cross-section adjustment method, the covariance matrices are assumed to be true. However, this assumption is not always appropriate. To improve the reliability of the cross-section adjustment method, the estimation of the covariance model using the metric ABIC has been introduced, and the performance of ABIC has been investigated through simple numerical experiments. This paper derives the formula to efficiently evaluate ABIC which is represented by a lower rank matrix to enable numerical experiments with large samples in a realistic computation time. From the results of the numerical experiments, it has been confirmed that ABIC tends to select a covariance model with fewer hyperparameters and a smaller variance for the estimation error. However, it has also been found that this desirable property of ABIC will be lost when the structure of the covariance model is far from the true one.

1 Background

Every measurement and analysis values always have uncertainties. Namely, the data follow a probability distribution with parameters such as population mean and covariance. Unfortunately, we cannot know the true parameters. Therefore, when we analyze the data, we usually estimate and/or assume these parameters. The appropriateness of the covariance matrix set (e.g., nuclear-data covariance) by analysts has been widely discussed in the framework of Sub-Groups (SGs) under the Working Party on International Nuclear Data Evaluation Cooperation (WPEC) in OECD/NEA (e.g., [1-4]). Recently, the discussion had been also done in Japan through the activity of the covariance data utilization and promotion working group organized in the JENDL committee [5].

In the cross-section adjustment method, covariance matrices are also used. One of the most important things for a reliable cross-section adjustment method is giving suitable covariance matrices close enough to the true covariance matrices. To judge the goodness of the covariance modeling, a metric is desirable. As a candidate for this metric, we focus on Akaike's Bayesian Information Criterion (ABIC) [6] which is one of the information criteria in Bayesian inference, because the cross-section adjustment method is often discussed within the framework of Bayesian inference.

In the conventional cross-section adjustment method, incorporation of the analysis method errors

(errors due to the core calculation method, e.g., discretizing error in a deterministic code) as a covariance matrix still requires ad hoc treatment. In JAEA, the integral experimental database for fast reactors has been developed and the adjusted cross-section set ADJ2017 [7, 8] has been created based on this database. Many of the core characteristics in the database have been analyzed by a deterministic method. Therefore, the predicted core characteristics have non-negligible uncertainties with correlations due to some numerical approximations. However, evaluating the uncertainties and their correlations is still a challenging issue. In addition, there would be unknown uncertainties that experimenters and analysts of reactor physics experiments were not able to recognize.

In order to address the difficulties in the conventional adjustment method, we will try to incorporate ABIC into the adjustment method. ABIC is expected to work as a metric for evaluating the goodness of covariance matrix modeling related to the uncertainties. We aim to improve the reliability of the conventional cross-section adjustment method by ABIC. This paper investigates the applicability of ABIC through several numerical experiments using random sampling techniques.

We introduce ABIC and incorporate this into the conventional cross-section adjustment in the next section. The applicability is evaluated through the simple numerical experiments in section 3. The conclusion of this paper will be described in section 4.

* Corresponding author: maruyama.shuhei@jaea.go.jp

2 Methodology

Before we get to the main subject, we introduce the notations that will be used in this paper. The covariance matrix with the tilde (“~”) means a covariance model with hyperparameters determined later. The covariance matrix with a hat (“^”) means the selected covariance by determining the hyperparameters. These covariances are not always “true.” The true covariance is represented by the characters without a tilde or a hat. For example,

- \mathbf{V} : a true covariance matrix,
- $\tilde{\mathbf{V}}$: a covariance model with undetermined hyperparameters,
- $\hat{\mathbf{V}}$: the covariance selected among a covariance model $\tilde{\mathbf{V}}$ by determining hyperparameters.

2.1 Review of conventional cross-section adjustment method

When true nuclear data set \mathbf{T} follows a multivariate normal distribution with the mean \mathbf{T}_0 and covariance \mathbf{M} , the probability distribution of \mathbf{T} is

$$\mathcal{P}(\mathbf{T}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{M}|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (\mathbf{T} - \mathbf{T}_0)^T \cdot \mathbf{M}^{-1} (\mathbf{T} - \mathbf{T}_0) \right], \quad (1)$$

where n is the dimension of \mathbf{T} and \mathbf{T}_0 , $|\mathbf{X}|$ is the determinant of a matrix \mathbf{X} , and the superscripts -1 and T represent the inverse and transpose of the matrix respectively. When integral experimental data sets are obtained, the likelihood function is represented as follows:

$$\mathcal{P}(\mathbf{R}_e | \mathbf{T}) = \frac{1}{(2\pi)^{\frac{r}{2}} |\mathbf{V}_e + \mathbf{V}_m|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (\mathbf{R}_e - \mathbf{R}_c(\mathbf{T}))^T \cdot (\mathbf{V}_e + \mathbf{V}_m)^{-1} (\mathbf{R}_e - \mathbf{R}_c(\mathbf{T})) \right]. \quad (2)$$

\mathbf{R}_e is a set of experimental values and \mathbf{V}_e is the experimental covariance matrix. $\mathbf{R}_c(\mathbf{T})$ is a set of calculational values if the true nuclear data set \mathbf{T} is given, and \mathbf{V}_m is the covariance matrix due to calculation methods. Here, r denotes the dimension of \mathbf{R}_e and \mathbf{R}_c . From Bayes' theorem,

$$\mathcal{P}(\mathbf{T} | \mathbf{R}_e) \propto \mathcal{P}(\mathbf{R}_e | \mathbf{T}) \mathcal{P}(\mathbf{T}), \quad (3)$$

the posterior probability is represented as follows:

$$\mathcal{P}(\mathbf{T} | \mathbf{R}_e) \propto \exp \left[-\frac{1}{2} (\mathbf{T} - \mathbf{T}_0)^T \mathbf{M}^{-1} (\mathbf{T} - \mathbf{T}_0) - \frac{1}{2} (\mathbf{R}_e - \mathbf{R}_c(\mathbf{T}))^T (\mathbf{V}_e + \mathbf{V}_m)^{-1} (\mathbf{R}_e - \mathbf{R}_c(\mathbf{T})) \right]. \quad (4)$$

The cross-section and its covariance after adjustment are derived from the condition that maximizes the posterior probability of Eq. (4) as follows:

$$\mathbf{T}' = \mathbf{T}_0 + \mathbf{M} \mathbf{G}^T \mathbf{V}_t^{-1} (\mathbf{R}_e - \mathbf{R}_c(\mathbf{T}_0)), \quad (5)$$

$$\mathbf{M}' = \mathbf{M} - \mathbf{M} \mathbf{G}^T \mathbf{V}_t^{-1} \mathbf{G} \mathbf{M}, \quad (6)$$

where

$$\mathbf{V}_t \equiv \mathbf{G} \mathbf{M} \mathbf{G}^T + \mathbf{V}_e + \mathbf{V}_m. \quad (7)$$

Here, \mathbf{G} means nuclear-data sensitivity coefficients. Note that the linear approximation for $\mathbf{R}_c(\mathbf{T})$, *i.e.*,

$$\mathbf{R}_c(\mathbf{T}) = \mathbf{R}_c(\mathbf{T}_0) + \mathbf{G}(\mathbf{T} - \mathbf{T}_0), \quad (8)$$

is assumed to derive Eqs. (5) and (6). The concept of the cross-section adjustment method is shown in Fig. 1.

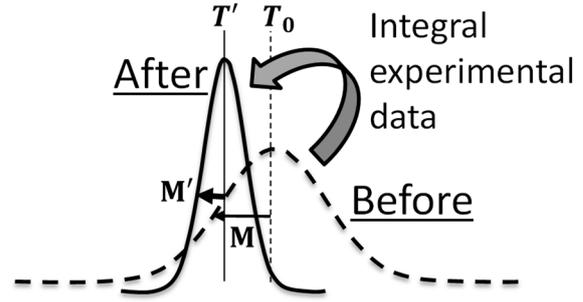


Fig. 1. Concept of cross-section adjustment

2.2 Akaike's Bayesian information criterion

In the conventional cross-section adjustment method, the covariance matrices \mathbf{M} , \mathbf{V}_e , and \mathbf{V}_m used for the adjustment are assumed to be true. However, these are not always appropriate. Therefore, let us assume the covariance matrices have hyperparameters that are determined to suit the observed data. To find a better covariance model by tuning the hyperparameters, we propose to use ABIC in this paper. As mentioned in the previous section, since the setting of the covariance matrix \mathbf{V}_m due to a calculation method is a challenging issue in the conventional adjustment method, only \mathbf{V}_m is assumed to be unknown and to be the estimation target using ABIC in this study. In other words, \mathbf{M} and \mathbf{V}_e in this study are assumed to be true.

If the covariance model $\tilde{\mathbf{V}}_m$ has unknown positive hyperparameters $p_1^2, p_2^2, \dots, p_q^2$, the likelihood function of Eq. (2) is represented as follows:

$$\mathcal{P}(\mathbf{R}_e | \mathbf{T}; p_1^2, p_2^2, \dots, p_q^2) \propto \exp \left[-\frac{1}{2} (\mathbf{R}_e - \mathbf{R}_c(\mathbf{T}))^T \cdot (\mathbf{V}_e + \tilde{\mathbf{V}}_m(p_1^2, p_2^2, \dots, p_q^2))^{-1} (\mathbf{R}_e - \mathbf{R}_c(\mathbf{T})) \right], \quad (9)$$

where q is the number of hyperparameters. As shown later (subsection 3.1.3), we prepare positive hyperparameters expressed as squared values to guarantee the positive definiteness of the covariance matrix $\tilde{\mathbf{V}}_m$. As a metric of the goodness of an inference model including the covariance modeling, we employ an information criterion called ABIC,

$$\text{ABIC} = -2 \ln l(p_1^2, p_2^2, \dots, p_q^2) + 2q. \quad (10)$$

Here, l is the marginal likelihood defined as

$$l(p_1^2, p_2^2, \dots, p_q^2) = \int d\mathbf{T} \mathcal{P}(\mathbf{R}_e | \mathbf{T}; p_1^2, p_2^2, \dots, p_q^2) \mathcal{P}(\mathbf{T}). \quad (11)$$

ABIC consists of the two terms related to l and q . The marginal likelihood l is proportional to the generation probability of the observed data (integral experimental data set \mathbf{R}_e). By choosing a good inference model that leads to this probability, *i.e.*, a suitable covariance model, ABIC is decreased. On the other hand, the number of hyperparameters q works penalty in terms of the complexity of an inference model. Therefore, ABIC prefers a simpler model having fewer hyperparameters. In a previous study, a cross-section

adjustment method based on the optimization of marginal likelihood l was presented [9]. Our proposed method using ABIC proposed in this paper is an extension of the previous study.

Performing the integral in Eq. (11) under the assumption of Eq. (8), the l becomes (e.g., [10])

$$-2 \ln l = \ln |\mathbf{M}| + \ln |\mathbf{V}_e + \tilde{\mathbf{V}}_m(p_1^2, p_2^2, \dots, p_q^2)| - \ln |\tilde{\mathbf{M}}'(p_1^2, p_2^2, \dots, p_q^2)| + \text{tr}[\mathbf{S}\tilde{\mathbf{V}}_t^{-1}(p_1^2, p_2^2, \dots, p_q^2)] + r \ln(2\pi), \quad (12)$$

where

$$\tilde{\mathbf{M}}'(p_1^2, p_2^2, \dots, p_q^2) = \left[\mathbf{M}^{-1} + \mathbf{G}^T (\mathbf{V}_e + \tilde{\mathbf{V}}_m(p_1^2, p_2^2, \dots, p_q^2))^{-1} \mathbf{G} \right]^{-1}, \quad (13)$$

$$\tilde{\mathbf{V}}_t(p_1^2, p_2^2, \dots, p_q^2) \equiv \mathbf{G}\mathbf{M}\mathbf{G}^T + \mathbf{V}_e + \tilde{\mathbf{V}}_m(p_1^2, p_2^2, \dots, p_q^2), \quad (14)$$

$$\mathbf{S} \equiv (\mathbf{R}_e - \mathbf{R}_c(\mathbf{T}_0))(\mathbf{R}_e - \mathbf{R}_c(\mathbf{T}_0))^T. \quad (15)$$

Here, $\text{tr}(\mathbf{X})$ is the trace of a matrix \mathbf{X} . In general, the rank of $\tilde{\mathbf{M}}'$, i.e., n is very large in the cross-section adjustment method. So, the evaluation of $\ln|\tilde{\mathbf{M}}'|$ becomes expensive without any treatments. In the hyperparameters tuning using ABIC, the iterative calculation due to the determination of hyperparameters is needed. Moreover, this iterative calculation will be repeated for large samples in this paper. Thus, the numerical experiments cannot be performed within a realistic computational cost. To avoid this issue, ABIC is represented as follows (see Appendix):

$$\text{ABIC} = \mathcal{L} + 2q, \quad (16)$$

where

$$\mathcal{L} \equiv \ln |\tilde{\mathbf{V}}_t(p_1^2, p_2^2, \dots, p_q^2)| + \text{tr}[\mathbf{S}\tilde{\mathbf{V}}_t^{-1}(p_1^2, p_2^2, \dots, p_q^2)] + r \ln(2\pi). \quad (17)$$

The equivalence of Eq. (12) and Eq. (17) was confirmed by preliminary numerical experiments.

Generally, the dimension of $\tilde{\mathbf{V}}_t$, i.e., r is much smaller than n . As described in Sec. 3, the total number of experimental data r is 33 in the present numerical experiments and sufficiently smaller than the total number of nuclear data $n = 14230$. Therefore, ABIC is evaluated from Eqs. (16) and (17) instead of Eqs. (10) and (12) in this study.

The covariance $\tilde{\mathbf{V}}_t$ generally has a wide range of values with different orders. Then, the determinant of such matrices can easily cause a numerical underflow/overflow even if the rank of the matrix is not so high. Therefore, some treatment should be needed to stably evaluate ABIC. Various methods can be considered to avoid this numerical difficulty, but from the viewpoint of computation time, we adopted the “linalg.slogdet” method of the NumPy library [11]. This method computes a determinant via LU factorization using a LAPACK [12] routine.

The procedure of the cross-section adjustment using ABIC consists of two steps as shown in Fig. 2. The new procedure of covariance modeling is added before the conventional cross-section adjustment procedure. In this procedure, the covariance model with hyperparameters is assumed first. Then, the hyperparameters will be determined so that ABIC is minimized. As the solver of the ABIC minimization, the SciPy library [13] will be used. Finally, based on the hyperparameters determined

by ABIC, the suitable covariance matrix $\hat{\mathbf{V}}_m(\hat{p}_1^2, \hat{p}_2^2, \dots, \hat{p}_q^2)$ is selected. The cross-section adjustment is performed using this selected covariance.

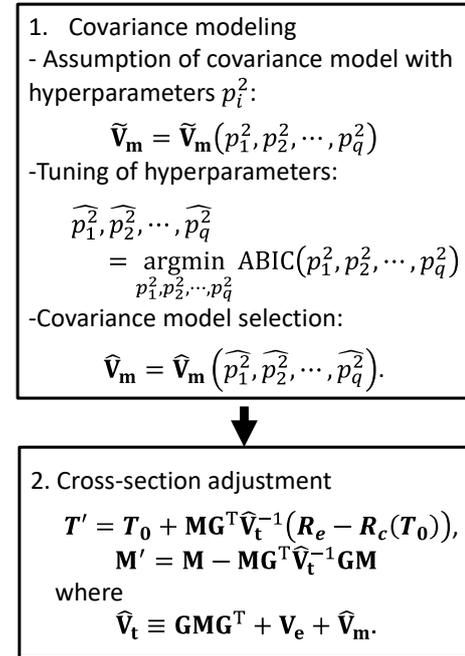


Fig. 2. The procedure of cross-section adjustment using ABIC

3 Numerical experiments

To evaluate the applicability of ABIC to the cross-section adjustment method, some numerical experiments using the random sampling method were performed in this section. Thereby, we generated virtual nuclear data sets and integral experimental data sets based on the virtual true covariances. Using the random sampling data sets, the cross-section adjustment was performed. After that, the performance of ABIC will be discussed from the relationship between ABIC and the variance of an estimation error resulting from the cross-section adjustment method.

3.1 Conditions

3.1.1 Procedure of random sampling

First, to perform the random sampling, we virtually assumed the true covariance matrices \mathbf{M} , \mathbf{V}_e , and \mathbf{V}_m . We employed the covariance matrices in our integral experimental database [7, 8] for fast reactors as the true covariance matrices. Sensitivity coefficients \mathbf{G} were also prepared from the database. As the integral experimental data for the numerical experiments, the experimental data related to criticality measurements used for the development of ADJ2017 were selected. Thirty-three experimental data listed in Table 1 were used.

It was assumed that these covariance matrices were all positive definite matrices. The singular value decomposition was performed on the original

covariance matrix, and all the singular values which are smaller than a setting value Λ were replaced with Λ in advance. We set $\Lambda = 1.0 \times 10^{-8}$ here. The rank of \mathbf{M} is the same as the dimension of the nuclear data set, $n = 14,230$, and the ranks of \mathbf{V}_e and \mathbf{V}_m are the same as the dimension of the integral experimental data set, $r = 33$.

The procedure of the random sampling based on these prepared covariance matrices is described below. The i -th sample related to a nuclear data set $\Delta \mathbf{T}_i \equiv [\mathbf{T} - \mathbf{T}_0]_i$ was generated from the probability distribution of Eq. (1). Namely, the obtained sample follows the following multivariate normal distribution:

$$\Delta \mathbf{T}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{M}), \quad (18)$$

and the i th sample related to an integral experimental data set $\Delta \mathbf{R}_i(\mathbf{T}) \equiv [\mathbf{R}_e - \mathbf{R}_c(\mathbf{T})]_i$ was generated from the probability distribution of Eq. (2), Namely, the obtained sample obeys the following multivariate normal distribution:

$$\Delta \mathbf{R}_i(\mathbf{T}) \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_e + \mathbf{V}_m). \quad (19)$$

Here, $i = 1, 2, \dots, N$. We set sample size $N = 10,000$ in this paper. The sample of $\Delta \mathbf{R}_i(\mathbf{T}_0) \equiv [\mathbf{R}_e - \mathbf{R}_c(\mathbf{T}_0)]_i$ which was used for the cross-section adjustment was obtained from $\Delta \mathbf{R}_i(\mathbf{T})$ correcting $\mathbf{G}\Delta \mathbf{T}_i$ based on the linear relationship of Eq. (8), *i.e.*,

$$\Delta \mathbf{R}_i(\mathbf{T}_0) = \Delta \mathbf{R}_i(\mathbf{T}) + \mathbf{G}\Delta \mathbf{T}_i. \quad (20)$$

Using this sample to the adjustment formula of Eq. (5), $\Delta \mathbf{T}'_i \equiv [\mathbf{T}' - \mathbf{T}_0]_i$ was evaluated as

$$\Delta \mathbf{T}'_i = \mathbf{M}\mathbf{G}^T \hat{\mathbf{V}}_t^{-1} \Delta \mathbf{R}_i(\mathbf{T}_0). \quad (21)$$

Note that, since there is no way to know the true covariance matrix in the real world, $\hat{\mathbf{V}}_m$ was determined by minimizing ABIC and $\Delta \mathbf{T}'_i$ were estimated by Eq. (21) using $\hat{\mathbf{V}}_t$ which includes $\hat{\mathbf{V}}_m$.

To evaluate the performance of ABIC, the estimation error of the cross-section adjustment was defined as $\Delta \mathbf{T}'_i - \Delta \mathbf{T}_i$. Here, $\Delta \mathbf{T}_i$ was generated from the true covariance \mathbf{M} , and $\Delta \mathbf{T}'_i$ was evaluated from the sample $\Delta \mathbf{R}_i(\mathbf{T}_0)$ generated from the true covariances $\mathbf{V}_e + \mathbf{V}_m$ and the estimated covariance $\hat{\mathbf{V}}_t$. To facilitate the interpretation of results, the estimation error was projected onto a sensitivity vector \mathbf{g} as

$$\Delta r_i \equiv \mathbf{g}(\Delta \mathbf{T}'_i - \Delta \mathbf{T}_i). \quad (22)$$

Here, \mathbf{g} was the sensitivity of the criticality of the JSFR core listed in Table 1. This core had been proposed as a design example for a next-generation sodium-cooled fast reactor core. Δr_i means the estimation error of the target core characteristic.

Moreover, the statistics for Δr_i were investigated. The following sample mean $\bar{\Delta r}$ and sample variance u were also calculated:

$$\bar{\Delta r} = \frac{1}{N} \sum_{i=1}^N \Delta r_i \quad (23)$$

$$u = \frac{1}{N-1} \sum_{i=1}^N (\Delta r_i - \bar{\Delta r})^2 \quad (24)$$

Table 1. Cores of integral experimental data and target core

Facility	Core name	Feature
Cores of integral experimental data		
ZPPR	9, 10A, 10B, 10C	600-800 MWe-class, two-region homogeneous MOX cores.
	13A	650 MWe-class, radially-heterogeneous MOX cores.
	18A, 18C, 19B	1000 MWe-class, two-region homogeneous MOX cores with enriched-uranium regions.
ZEBRA	MZA	550 liter-sized one-region MOX core as a clean benchmark.
	MZB, MZC	2,300 liter-sized two-region homogeneous MOX cores to simulate the prototype fast reactor MONJU
JOYO	MK-I (64, 70 assemblies)	300 liter-sized 50/75MWt fast power reactor core with MOX and enriched-uranium fuel with blanket
MONJU	1994, 2010	280MWe prototype fast breeder reactor with two-region homogeneous MOX core
BFS	62s, 66-1	3,400 liter-sized three or four-region enriched-uranium and/or MOX fuel cores with or without radial blankets.
	67s, 69s, 66-2s	10kg of NpO ₂ loading cores in central MOX region with various grade Pu.
LANL	FLATTOP -Pu, -25, GODIVA, JEZEBEL, JEZEBEL -Pu240	Sphere-shaped cores of approx. 10 cm in diameter with metallic fuel consisting of Pu-239, degraded Pu or U-235.
Target core		
JSFR	U-Pu	750MWe next-generation fast reactor with two-region homogeneous MOX core

3.1.2 Method to evaluate the performance of ABIC

We describe the method to evaluate the performance of ABIC using the results obtained from random sampling here. The performance was evaluated using the property of u described below.

If the selected covariance matrix $\hat{\mathbf{V}}_m$ is equal to the true one, *i.e.*, if $\hat{\mathbf{V}}_m = \mathbf{V}_m$, it is expected that the sample Δr_i follows Gauss distribution with a mean of 0 and a variance of $v = \mathbf{g}\mathbf{M}'\mathbf{g}^T$. Note \mathbf{M}' was nuclear data covariance after adjustment evaluated from the true covariance matrix \mathbf{V}_m . Hereafter, we call v as "reference variance." The distribution of Δr_i is shown in Fig. 3. It is found that the distribution follows the theoretical distribution as expected. The values of variance u and v are also consistent when $\hat{\mathbf{V}}_m = \mathbf{V}_m$, as shown in Table 2.

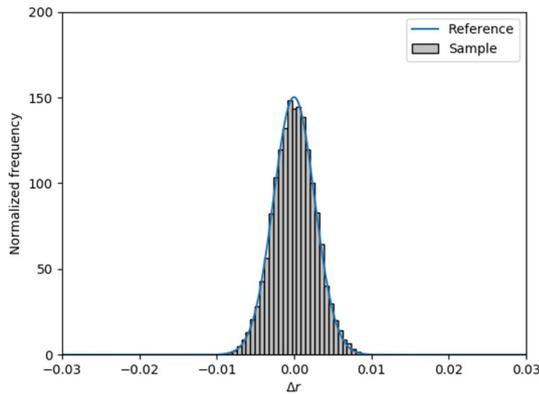


Fig. 3. Histogram of Δr (when $\hat{\mathbf{V}}_m = \mathbf{V}_m$)

Table 2. Comparison example of sample variances u with the reference variance v

$\hat{\mathbf{V}}_m$	Sample variance u [$\times 10^{-6}$]	Reference variance v [$\times 10^{-6}$]
\mathbf{V}_m	7.2	7.1
$0.1^2\mathbf{V}_m$	24.0	
$10^2\mathbf{V}_m$	10.7	

If we misestimate the evaluation of \mathbf{V}_m , for example, when $\hat{\mathbf{V}}_m = 0.1^2\mathbf{V}_m$ and $\hat{\mathbf{V}}_m = 10^2\mathbf{V}_m$, the value of u becomes larger than the reference variance v as shown in Table 2 and Figs. 4 and 5. The value of u becomes minimum when $\hat{\mathbf{V}}_m = \mathbf{V}_m$. It is expected that the value of u becomes larger as $\hat{\mathbf{V}}_m$ gets further away from the true one. Especially, the underestimation of \mathbf{V}_m has a larger impact on u than the overestimation of \mathbf{V}_m .

Based on this property, we interpret the covariance model showing the smallest u (i.e., closest u to the reference variance v) as the best model. We investigate the performance of ABIC based on whether the best model can be found by minimizing ABIC or not. Namely, we check for the consistency of the minimum point between ABIC and u .

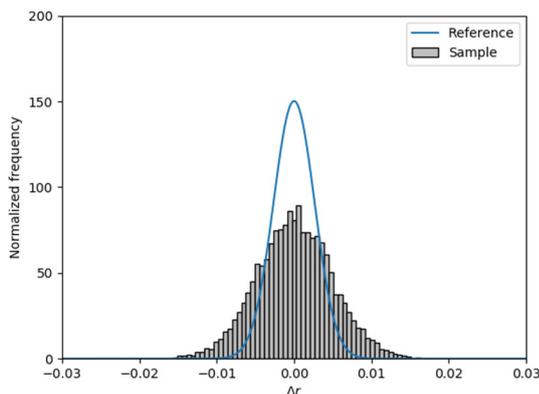


Fig. 4. Histogram of Δr (when $\hat{\mathbf{V}}_m = 0.1^2\mathbf{V}_m$)

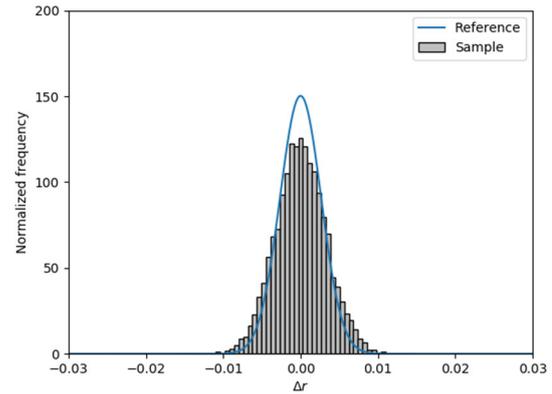


Fig. 5. Histogram of Δr when ($\hat{\mathbf{V}}_m = 10^2\mathbf{V}_m$)

3.1.3 Assumption of covariance models $\tilde{\mathbf{V}}_m$

The assumptions of the covariance models $\tilde{\mathbf{V}}_m$ corresponding to the unknown true covariance \mathbf{V}_m are listed in Table 3. The three kinds of covariance models $\tilde{\mathbf{V}}_m$ were prepared for the numerical experiments. Since the true covariance \mathbf{V}_m is unknown, assumptions for the covariance model $\tilde{\mathbf{V}}_m$ are necessary. Therefore, in the numerical tests, the covariance models $\tilde{\mathbf{V}}_m$ have been prepared with some modifications to the true covariance \mathbf{V}_m .

Table 3. Assumptions of $\tilde{\mathbf{V}}_m$

$\tilde{\mathbf{V}}_m$	Number of hyper-parameters q	Feature of the cases
$p^2\mathbf{V}_m$	1 (const.)	$\tilde{\mathbf{V}}_m$ can reproduce \mathbf{V}_m by hyperparameter tuning.
$p^2(\mathbf{V}_m + a\mathbf{I})$ $a\mathbf{I}$: noise matrix a : positive const.	1 (const.)	$\tilde{\mathbf{V}}_m$ cannot reproduce \mathbf{V}_m by hyperparameter tuning.
$\sum_{k=1}^K p_k^2 \mathbf{u}_k \mathbf{u}_k^T$ \mathbf{u}_k : singular vectors of \mathbf{V}_m	1 to K	Number of hyperparameters q is not constant.

The difference between the first and second cases is whether $\tilde{\mathbf{V}}_m$ can reproduce the true covariance \mathbf{V}_m or not by tuning the hyperparameter. In these cases, the number of hyperparameters was assumed to be constant. On the other hand, the third case had multiple hyperparameters.

The covariance structure and the singular values of the true covariance matrix \mathbf{V}_m used in the numerical experiments are shown in Figs. 6 and 7, respectively.

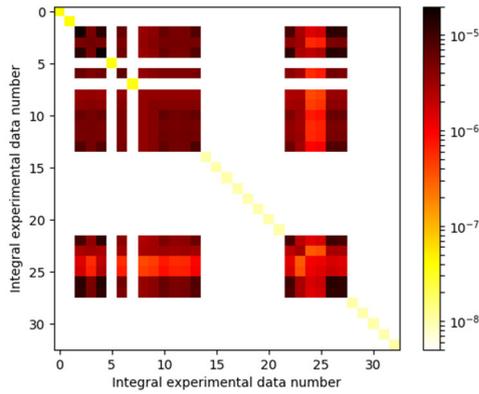


Fig. 6. Covariance structure of \mathbf{V}_m

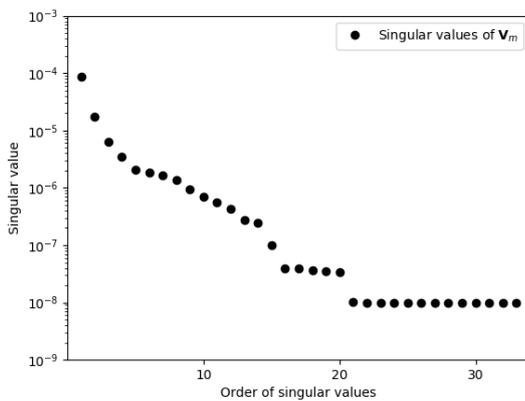


Fig. 7. Distribution for the singular values of \mathbf{V}_m

3.2 Results

3.2.1 The case of $\tilde{\mathbf{V}}_m = p^2 \mathbf{V}_m$

As described above, in the first case, the covariance model $\tilde{\mathbf{V}}_m$ can reproduce the true covariance \mathbf{V}_m by tuning the hyperparameter.

The histogram of the determined hyperparameter for each sample is shown in Fig. 8. The singular distribution was observed around $\hat{p}^2 = 0$. The mean of \hat{p}^2 became somewhat smaller than the expected value, $p^2 = 1$. These could be caused by the assumption of a positive value as the range of the hyperparameter.

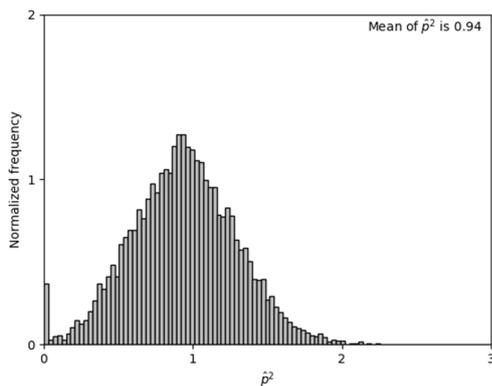


Fig. 8. Histogram of \hat{p}^2 (when $\tilde{\mathbf{V}}_m = \hat{p}^2 \mathbf{V}_m$)

The evaluated value of the sample variance u is listed in Table 4 with the reference variance v . The histogram of Δr_i is shown in Fig. 9. The shape of the histogram of Δr_i is very close to that of the reference one. In fact, the sample variance u and the mean of the reference variance v are close. Consequently, in the first numerical experiment, ABIC showed good performance to find an appropriate covariance model.

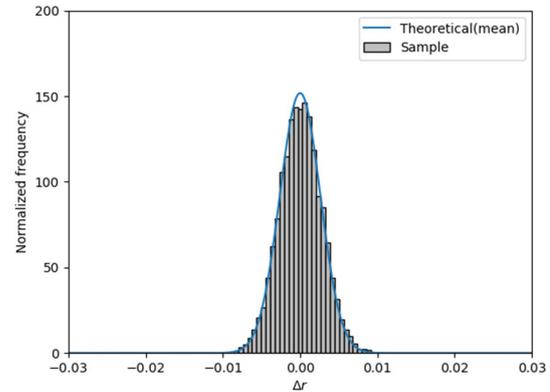


Fig. 9. Histogram of Δr (when $\tilde{\mathbf{V}}_m = \hat{p}^2 \mathbf{V}_m$)

Table 4. Comparison between sample variance u and the reference variance v in the first case.

$\tilde{\mathbf{V}}_m$	Sample variance u [$\times 10^{-6}$]	Reference variance v [$\times 10^{-6}$]
$\hat{p}^2 \mathbf{V}_m$	7.4	7.1

3.2.2 The case of $\tilde{\mathbf{V}}_m = p^2(\mathbf{V}_m + a\mathbf{I})$

In the second case, the noise matrix $a\mathbf{I}$ was employed in the $\tilde{\mathbf{V}}_m$. Unlike $\tilde{\mathbf{V}}_m$ given in the first case, $\tilde{\mathbf{V}}_m$ cannot reproduce the true covariance matrix \mathbf{V}_m due to the noise matrix $a\mathbf{I}$, even if the hyperparameter is tuned. Here, \mathbf{I} is a unit matrix, and a is a constant value to be $a = 8.6 \times 10^{-5}$. The value of a was set equal to the maximum singular value of \mathbf{V}_m .

The histogram of Δr_i is shown in Fig. 10 and Table 5 shows the evaluated values of the mean of ABIC and u . The result of the first case of $\tilde{\mathbf{V}}_m = p^2 \mathbf{V}_m$ is also shown in this table for comparison. The agreement of the histogram with the reference distribution becomes worse compared to the first case. The sample variance also becomes larger than that of the first case. As if to correspond to this phenomenon, the mean value of ABIC becomes larger. From this result, ABIC can reflect the badness of the covariance modeling of the second case compared to that of the first case. As far as this result is concerned, ABIC seems to adequately reflect the goodness of covariance modeling.

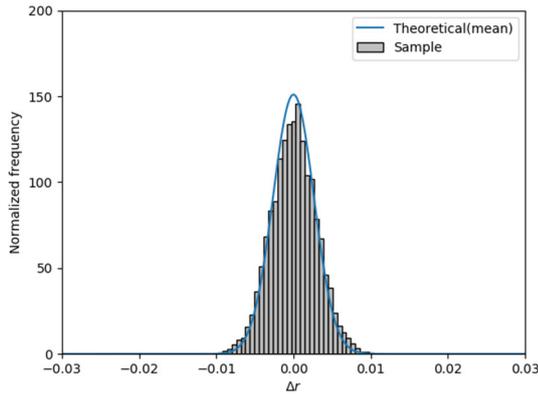


Fig. 10. Histogram of Δr (when $\hat{\mathbf{V}}_m = \hat{p}^2(\mathbf{V}_m + a\mathbf{I})$)

Table 5. Comparison of the mean of ABIC and sample variance u between the first and the second cases.

$\hat{\mathbf{V}}_m$	Mean of ABIC	Sample variance u [$\times 10^{-6}$]	Reference variance v [$\times 10^{-6}$]
$\hat{p}^2(\mathbf{V}_m + a\mathbf{I})$	-352	13.6	7.1
$\hat{p}^2\mathbf{V}_m$	-362	7.4	

However, a more detailed analysis reveals that the performance of ABIC will become poor when the covariance structure differs greatly from the true one. For the first and second cases, Figs. 11 and 12 show the changes in the mean of ABIC and the sample variance u according to the hyperparameter \hat{p}^2 change, respectively. The minimum points of the mean of ABIC and the sample variance u are shown as the white and black arrows in these figures. Figure 11 represents the results when we assumed $\hat{\mathbf{V}}_m = \hat{p}^2\mathbf{V}_m$. Both minimum points are matching between the sample mean of ABIC and sample variance u in the first case. This matching result shows that ABIC can find the optimal hyperparameter. On the other hand, Fig. 12 represents the results when we assumed $\hat{\mathbf{V}}_m = \hat{p}^2(\mathbf{V}_m + a\mathbf{I})$. In contrast to the previous result in Fig. 11, the mismatching of the minimum point was observed in the second case. This mismatching represents the difficulty in finding the optimal hyperparameters using ABIC. The result of the second numerical experiment implies that the mismatch tends to increase as the covariance structure differs.

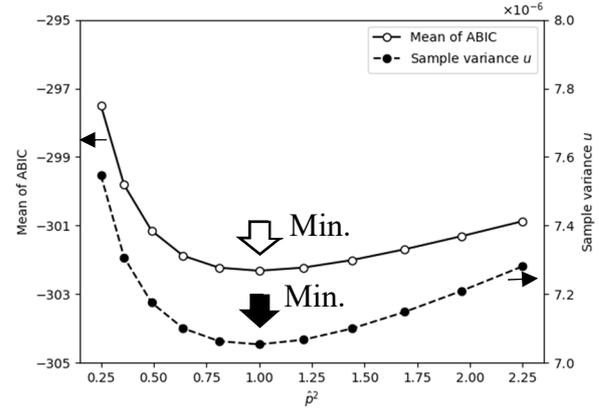


Fig. 11. Relationship between changes in the mean of ABIC and sample variance u with changes in the hyperparameters (when $\hat{\mathbf{V}}_m = \hat{p}^2\mathbf{V}_m$)

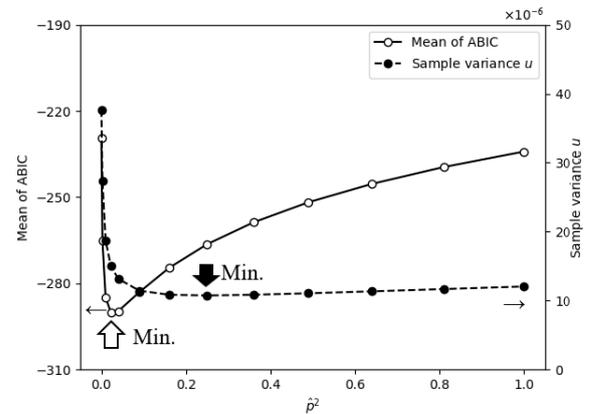


Fig. 12. Relationship between changes in the mean of ABIC and sample variance u with changes in the hyperparameters (when $\hat{\mathbf{V}}_m = \hat{p}^2(\mathbf{V}_m + a\mathbf{I})$)

3.2.3 The case of $\tilde{\mathbf{V}}_m = \sum_{k=1}^K p_k^2 \mathbf{u}_k \mathbf{u}_k^T$

In the third case, we investigated whether ABIC can choose an optimal number of hyperparameters. To clarify this, the singular vectors \mathbf{u}_k of \mathbf{V}_m were used for $\tilde{\mathbf{V}}_m$. Here, \mathbf{u}_k having a smaller number of k corresponds to the larger singular value.

The results of the sample mean of ABIC and sample variance u of when $K = 1-8$ are shown in Fig. 13. In addition, the result of the case $\tilde{\mathbf{V}}_m = \mathbf{O}$ is also shown in this figure as the case of $K = 0$. As K increases, the mean of ABIC decreases rapidly at first due to the decrease of \mathcal{L} (the first log likelihood term in the right-hand side of Eq. (16)), and then increases slowly due to the increase of $2q$ (the last penalty term in the right-hand side of Eq. (16)). The mean of ABIC becomes minimum when the reduction of the value of \mathcal{L} is saturated. The sample mean of ABIC becomes minimum when $K = 4$. The sample variance u becomes minimum when $K = 3$. The minimum points of the mean of ABIC and u were almost consistent. Thus, the third numerical experiment demonstrates that ABIC can appropriately suggest the number of hyperparameters. In this case, ABIC and u are larger than in the first case. This is due to the following reasons: If the covariance

modeling $\tilde{\mathbf{V}}_m$ is equal to the true covariance matrix \mathbf{V}_m , the sample variance u is consistent with the reference variance v . However, for smaller K , the true covariance matrix \mathbf{V}_m cannot be perfectly reproduced by $\tilde{\mathbf{V}}_m$ even if the hyperparameters are tuned. For larger K , there are too many hyperparameters to accurately estimate \mathbf{V}_m by $\tilde{\mathbf{V}}_m$.

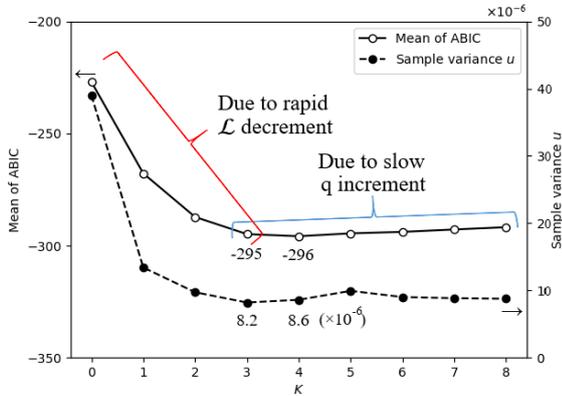


Fig. 13. Relationship between changes in the mean of ABIC and sample variance u

4 Conclusions

To improve the reliability of the cross-section adjustment method, the use of ABIC was proposed. We derived the formula to efficiently evaluate ABIC which was represented by a lower-rank matrix to enable numerical experiments with large samples in a realistic computation time.

To evaluate the applicability of ABIC to the cross-section adjustment method, some simple numerical experiments were performed in this study. Through these numerical experiments, the relationship between ABIC and u (*i.e.*, the variance of the estimation error of a core characteristic) was investigated. More specifically, the applicability of ABIC was discussed based on the consistency of the minimum points of ABIC and u . Consequently, it was confirmed that ABIC tends to select a covariance model with fewer hyperparameters and a smaller u . However, it was also found that this desirable property of ABIC will be lost if the structure of the covariance model was far from the true one. This is a future research problem to be solved when ABIC is applied to the cross-section adjustment method.

This work was supported by the Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (C) [Grant Number 21K04940].

References

1. M. Salvatores, G. Palmiotti, G. Aliberti, R. McKnight, P. Archier, C. De Saint Jean, E. Dupont, M. Herman, M. Ishikawa, K. Sugino, T. Ivanova, E. Ivanov, S.-J. Kim, I. Kodeli, A. Trkov, G. Manturov, S. Pelloni, C. Perfett, B.T. Rearden, A. Plompen, D.

Rochman, W. Wang, H. Wu, W.-S. Yang, NEA/NSC/WPEC/DOC(2013)445, OECD/NEA (2013)

2. M. Salvatores, G. Aliberti, M. Dunn, A. Hogenbirk, A. Ignatyuk, M. Ishikawa, I. Kodeli, A.J. Koning, R. McKnight, R.W. Mills, P. Oblozinsky, G. Palmiotti, A. Plompen, G. Rimpault, Y. Rugama, P. Talou, W. S. Yang, Nuclear Science NEA/WPEC-26, OECD/NEA (2008)
3. G. Palmiotti, M. Salvatores, K. Yokoyama, M. Ishikawa, NEA/NSC/R(2016)6, OECD/NEA (2017)
4. M. Salvatores, G. Palmiotti, Proc. of Int. Conf. on Nuclear Data for Science and Technology, ND2019, EPJ Web of Conferences, **239**, 13001 (2020)
5. Covariance Data Utilization and Promotion Working Group, JAEA-Review 2021-014, Japan Atomic Energy Agency (2021) [In Japanese]
6. H. Akaike, Proceedings of the 2nd International Symposium on Information Theory, pp. 267-281, Budapest (1973)
7. K. Yokoyama, K. Sugino, M. Ishikawa, S. Maruyama, Y. Nagaya, K. Numata, T. Jin, JAEA-Research 2018-011 (2019) [in Japanese]
8. K. Yokoyama, S. Maruyama, H. Taninaka, S. Ohki, JAEA-Data/Code 2021-019 (2022) [in Japanese]
9. D. Siefman, M. Hursin, G. Schnabel, H. Sjöstrand, Annals of nuclear energy **159**, 108255 (2021)
10. C. M. Bishop, Pattern Recognition and Machine Learning (2006)
11. NumPy reference guide, release 1.19 [Internet]. NumPy.org; [updated 2020 Jun 29; cited 2022 Sep 27] Available from: <https://numpy.org/doc/1.19/numpy-ref.pdf>
12. LAPACK Documentation, release 3.10.1 [Internet]. Netlib.org; [updated 2022 Sep 27; cited 2022 Sep 27] Available from: <https://netlib.org/lapack/explore-html/index.html>
13. SciPy reference guide, release 1.4.1 [Internet]. SciPy.org; [updated 2019 Dec 19; cited 2022 Sep 27] Available from: <https://docs.scipy.org/doc/scipy-1.4.1/scipy-ref-1.4.1.pdf>
14. Gene H. Golub and Charles F. van Loan, Matrix Computations (1996)

Appendix

We prove the equivalence of Eqs. (12) and (17). Namely, the equivalence is represented by the following relationship:

$$\ln|\tilde{\mathbf{M}}'| = \ln|\mathbf{M}| + \ln|\mathbf{V}_e + \tilde{\mathbf{V}}_m| - \ln|\tilde{\mathbf{V}}_t|. \quad (\text{A.1})$$

$\ln|\tilde{\mathbf{M}}'|$ in Eq. (12) is represented as

$$\ln|\tilde{\mathbf{M}}'| = \ln \left| \left[\mathbf{M}^{-1} + \mathbf{G}^T (\mathbf{V}_e + \tilde{\mathbf{V}}_m)^{-1} \mathbf{G} \right]^{-1} \right|, \quad (\text{A.2})$$

using Eq. (13). From the Woodbury identity [8]:

$$(\mathbf{X} + \mathbf{YZY}^T)^{-1} = \mathbf{X}^{-1}$$

$$-\mathbf{X}^{-1}\mathbf{Y}(\mathbf{Z}^{-1} + \mathbf{Y}^T\mathbf{X}^{-1}\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{X}^{-1}, \quad (\text{A.3})$$

Eq. (A.2) becomes

$$\begin{aligned} \ln|\tilde{\mathbf{M}}'| &= \ln|\mathbf{M} - \mathbf{M}\mathbf{G}^T\tilde{\mathbf{V}}_t^{-1}\mathbf{G}\mathbf{M}| \\ &= \ln|\mathbf{M}| + \ln|(\mathbf{I} - \mathbf{G}^T\tilde{\mathbf{V}}_t^{-1}\mathbf{G}\mathbf{M})|, \end{aligned} \quad (\text{A.4})$$

where \mathbf{I} is the unit matrix. Using the following relationship:

$$\ln|\mathbf{X}| = \text{tr}[\ln(\mathbf{X})], \quad (\text{A.5})$$

to the second term of Eq. (A.4), Eq. (A.4) becomes

$$\ln|\tilde{\mathbf{M}}'| = \ln|\mathbf{M}| + \text{tr}[\ln(\mathbf{I} - \mathbf{G}^T\tilde{\mathbf{V}}_t^{-1}\mathbf{G}\mathbf{M})]. \quad (\text{A.6})$$

Furthermore, using the relationships:

$$\ln(\mathbf{I} - \mathbf{X}) = \sum_{m=1}^{\infty} \frac{1}{m} \mathbf{X}^m, \quad (\text{A.7})$$

and

$$\text{tr}(\mathbf{X}\mathbf{Y}\mathbf{Z}) = \text{tr}(\mathbf{Y}\mathbf{Z}\mathbf{X}) = \text{tr}(\mathbf{Z}\mathbf{X}\mathbf{Y}), \quad (\text{A.8})$$

Eq. (A.6) is represented as

$$\ln|\tilde{\mathbf{M}}'| = \ln|\mathbf{M}| + \text{tr} \left[\sum_{m=1}^{\infty} \frac{1}{m} (\tilde{\mathbf{V}}_t^{-1}\mathbf{G}\mathbf{M}\mathbf{G}^T)^m \right]. \quad (\text{A.9})$$

Again, using the relationship of Eqs. (A.7) and (A.8) to Eq. (A.9), we obtain

$$\begin{aligned} \ln|\tilde{\mathbf{M}}'| &= \ln|\mathbf{M}| + \ln|\mathbf{I} - \tilde{\mathbf{V}}_t^{-1}\mathbf{G}\mathbf{M}\mathbf{G}^T| \\ &= \ln|\mathbf{M}| + \ln|\mathbf{V}_e + \tilde{\mathbf{V}}_m| - \ln|\tilde{\mathbf{V}}_t|. \end{aligned} \quad (\text{A.10})$$