Scientific
Research
Publishing

Second-Order MaxEnt Predictive Modelling Methodology. II: Probabilistically Incorporated Computational Model (2nd-BERRU-PMP)

Dan Gabriel Cacuci

Center for Nuclear Science and Energy, Department of Mechanical Engineering, University of South Carolina, Columbia, SC, USA
Email: cacuci@cec.sc.edu

Abstract

This work presents a comprehensive second-order predictive modeling (PM) methodology based on the maximum entropy (MaxEnt) principle for obtaining best-estimate mean values and correlations for model responses and parameters. This methodology is designated by the acronym *2nd-BERRU-PMP*, where the attribute "2nd" indicates that this methodology incorporates second-order uncertainties (means and covariances) and second (and higher) order sensitivities of computed model responses to model parameters. The acronym BERRU stands for "Best-Estimate Results with Reduced Uncertainties" and the last letter ("*P*") in the acronym indicates "probabilistic," referring to the MaxEnt probabilistic inclusion of the computational model responses. This is in contradistinction to the *2nd-BERRU-PMD* methodology, which *deterministically* combines the computed model responses with the experimental information, as presented in the accompanying work (Part I). Although both the *2nd-BERRU-PMP* and the *2nd-BERRU-PMD* methodologies yield expressions that include second (and higher) order sensitivities of responses to model parameters, the respective expressions for the predicted responses, for the calibrated predicted parameters and for their predicted uncertainties (covariances), are not identical to each other. Nevertheless, the results predicted by both the *2nd-BERRU-PMP* and the *2nd-BERRU-PMD* methodologies encompass, as particular cases, the results produced by the extant data assimilation and data adjustment procedures, which rely on the minimization, in a least-square sense, of a user-defined functional meant to represent the discrepancies between measured and computed model responses.

Keywords

Second-Order Predictive Modeling, Data Assimilation, Data Adjustment,

Uncertainty Quantification, Reduced Predicted Uncertainties

1. Introduction

An accompanying work [1] has presented a comprehensive second-order predictive modeling (PM) methodology based on the maximum entropy (MaxEnt) principle [2], for obtaining best-estimate mean values and correlations for model responses and parameters, which was designated by the acronym 2nd-BERRU-PMD. The attribute "2nd" indicates that this methodology incorporates second-order uncertainties (means and covariances) and second-order sensitivities of computed model responses to model parameters. The acronym BERRU stands for "Best-Estimate Results with Reduced Uncertainties" and the last letter ("D") in the acronym indicates "deterministic," referring to the deterministic inclusion of the computational model responses.

Alternative to the 2nd-BERRU-PMD methodology, this work presents the 2nd-BERRU-PMP methodology for obtaining best-estimate mean values and correlations for model responses and parameters. The 2nd-BERRU-PMP methodology is also based on the MaxEnt principle but includes the computational model responses probabilistically-through a MaxEnt representation-in contradistinction to its deterministic inclusion within the 2nd-BERRU-PMD methodology. This (contra) distinction is indicated the last letter ("P") in the acronym, which refers to the "probabilistic" inclusion of the computed model responses.

This work is structured as follows: Section 2 presents the second-order MaxEnt probabilistic representation of the computational model. Section 3 presents the "second order predictive modeling methodology with probabilistically included computed responses" (2nd-BERRU-PMP) methodology. Although both the 2nd-BERRU-PMP and the 2nd-BERRU-PMD methodologies yield expressions that include second (and higher) order sensitivities of responses to model parameters, the respective expressions for the predicted responses, calibrated predicted parameters and their predicted uncertainties (covariances) are not identical to each other, although they encompass, as particular cases, the results produced by the extant data assimilation [3] [4] and data adjustment procedures [5] [6] [7] [8] [9]. Comparisons with the results produced by the first-order BERRU-PM methodology [10] [11] [12] are also presented. The advantages of the 2nd-BERRU-PMP methodology over the results produced by the data assimilation method [3] [4] are presented in Section 5. The discussion presented in Section 6 concludes this work. Illustrative applications of the 2nd-BERRU-PMP and the 2nd-BERRU-PMD methodologies to forward and inverse problems are currently in progress.

2. Second-Order MaxEnt Probabilistic Representation of the Computational Model

The modeling of a physical system requires consideration of the following mod-

eling components:

1) A mathematical model comprising linear and/or nonlinear equations that relate the system's independent variables and parameters to the system's state (*i.e.*, dependent) variables. In this work, matrices will be denoted using capital bold letters while vectors will be denoted using either capital or lower-case bold letters. The symbol " ≜ " will be used to denote "is defined as" or "is by definition equal to." Transposition will be indicated by a dagger (†) superscript. The equalities in this work are considered to hold in the weak ("distributional") sense.

2) Nominal values and uncertainties that characterize the model's parameters.

3) One or several computational results, customarily referred to as model responses (or objective functions, or indices of performance), which are computed using the mathematical model; and

4) External to the model: experimentally measured responses that correspond to the computed responses, with their respective experimentally determined nominal (mean) values and uncertainties (variances, covariances, skewness, kurtosis, etc.). Occasionally, measurements of correlations among the measured responses and the model parameters, as well as externally performed measurements of model parameters (in addition to the information about parameters used in the model-computations) might be available.

The model parameters usually stem from processes that are external to the physical system (and, hence, model) under consideration and their precise values are seldom, if ever, known. The known characteristics of the model parameters may include their nominal (expected/mean) values and, possibly, higher-order moments (*i.e.*, variance/covariances, skewness, kurtosis), which are usually determined from experimental data and/or processes external to the physical system under consideration. Occasionally, just the lower and the upper bounds may be known for some model parameters. Without loss of generality, the model parameters will be considered in this work to be real-valued scalars, and will be denoted as $\alpha_1, \cdots, \alpha_{TP}$, where the quantity "*TP*" denotes the "total number of model parameters." Mathematically, these parameters are considered as components of a *TP*-dimensional vector denoted as $\boldsymbol{\alpha} \triangleq (\alpha_1, \cdots, \alpha_{TP})^\dagger \in D_\alpha \in \mathbb{R}^{TP}$, defined over a domain D_α, which is included in a *TP*-dimensional subset of the $\mathbb{R}^{TP}$. The components of the *TP*-dimensional column vector $\boldsymbol{\alpha} \in \mathbb{R}^{TP}$ are considered to include imprecisely known geometrical parameters that characterize the physical system's boundaries in the phase-space of the model's independent variables. The model parameters can be considered to be quasi-random scalar-valued quantities which follow an unknown multivariate distribution denoted as $p_\alpha(\boldsymbol{\alpha})$. The mean values which will be called "nominal" values in the context of computational modeling of the model parameters will be denoted as α_i^0; the superscript "0" will be used throughout this work to denote "nominal values." These nominal values are formally defined as follows:

$$\alpha_i^0 \triangleq \int_{D_\alpha} \alpha_i p_\alpha(\boldsymbol{\alpha}) \mathrm{d}\boldsymbol{\alpha}, \ i = 1, \cdots, TP. \tag{1}$$

The expected values of the measured parameters will be considered to constitute the components of a "vector of nominal values" denoted as

$\boldsymbol{\alpha}^0 \triangleq \left[\alpha_1^0, \cdots, \alpha_i^0, \cdots, \alpha_{TP}^0 \right]^\dagger$.

The covariance, $\mathrm{cov}(\alpha_i, \alpha_j)$, of two model parameters, α_i and α_j, is defined as follows:

$$\mathrm{cov}(\alpha_i, \alpha_j) \triangleq \int_{D_\alpha} \left(\alpha_i - \alpha_i^0 \right)\left(\alpha_j - \alpha_j^0 \right) p_\alpha(\boldsymbol{\alpha}) \, \mathrm{d}\boldsymbol{\alpha}; \ i, j = 1, \cdots, TP. \tag{2}$$

The covariances $\mathrm{cov}(\alpha_i, \alpha_j)$ are considered to be the components of the parameter covariance matrix, denoted as $\boldsymbol{C}_{\alpha\alpha} \triangleq \left[\mathrm{cov}(\alpha_i, \alpha_j) \right]_{TP \times TP}$.

The results computed using a mathematical model are customarily called "model responses" (or "system responses" or "objective functions" or "indices of performance"). Consider that there are a total number of TR such model responses, each of which can be considered to be a component of the "vector of model responses" $\boldsymbol{r} \triangleq \left(r_1, \cdots, r_k, \cdots, r_{TR} \right)^\dagger$. Each of these model responses is formally a function (implicit and/or explicit) of the model parameters $\boldsymbol{\alpha}$, *i.e.*, $r_k = r_k(\boldsymbol{\alpha}); k = 1, \cdots, TR$. The uncertainties affecting the model parameters $\boldsymbol{\alpha}$ will "propagate" both directly and indirectly, through the model's dependent variables, to induce uncertainties in the computed responses, which will therefore be denoted as $r_k^c(\boldsymbol{\alpha})$. Each computed response can be formally expanded in a multivariate Taylor-series around the parameters' mean values. In particular, the fourth-order Taylor-series of a system response, denoted as $r_k(\boldsymbol{\alpha})$, around the expected (or nominal) parameter values $\boldsymbol{\alpha}^0$ has the following formal expression:

$$
\begin{aligned}
r_k(\boldsymbol{\alpha}) = r_k(\boldsymbol{\alpha}^0) &+ \sum_{j_1=1}^{TP} \left\{ \frac{\partial r_k(\boldsymbol{\alpha})}{\partial \alpha_{j_1}} \right\}_{\boldsymbol{\alpha}^0} \delta\alpha_{j_1} + \frac{1}{2} \sum_{j_1=1}^{TP} \sum_{j_2=1}^{TP} \left\{ \frac{\partial^2 r_k(\boldsymbol{\alpha})}{\partial \alpha_{j_1} \partial \alpha_{j_2}} \right\}_{\boldsymbol{\alpha}^0} \delta\alpha_{j_1} \delta\alpha_{j_2} \\
&+ \frac{1}{3!} \sum_{j_1=1}^{TP} \sum_{j_2=1}^{TP} \sum_{j_3=1}^{TP} \left\{ \frac{\partial^3 r_k(\boldsymbol{\alpha})}{\partial \alpha_{j_1} \partial \alpha_{j_2} \partial \alpha_{j_3}} \right\}_{\boldsymbol{\alpha}^0} \delta\alpha_{j_1} \delta\alpha_{j_2} \delta\alpha_{j_3} \\
&+ \frac{1}{4!} \sum_{j_1=1}^{TP} \sum_{j_2=1}^{TP} \sum_{j_3=1}^{TP} \sum_{j_4=1}^{TP} \left\{ \frac{\partial^4 r_k(\boldsymbol{\alpha})}{\partial \alpha_{j_1} \partial \alpha_{j_2} \partial \alpha_{j_3} \partial \alpha_{j_4}} \right\}_{\boldsymbol{\alpha}^0} \delta\alpha_{j_1} \delta\alpha_{j_2} \delta\alpha_{j_3} \delta\alpha_{j_4} + \varepsilon_k.
\end{aligned}
\tag{3}
$$

In Equation (3), the quantity $r_k(\boldsymbol{\alpha}^0)$ indicates the computed value of the response using the expected/nominal parameter values $\boldsymbol{\alpha}^0 \triangleq \left(\alpha_1^0, \cdots, \alpha_{TP}^0 \right)^\dagger$. The notation $\{\ \}_{\boldsymbol{\alpha}^0}$ indicates that the quantities within the braces are also computed using the expected/nominal parameter values. The quantity ε_k in Equation (3) comprises all quantifiable errors in the representation of the computed response as a function of the model parameters, including the truncation errors $O(\delta\alpha_j)^5$ of the Taylor-series expansion, possible bias-errors due to incompletely modeled physical phenomena, and possible random errors due to numerical approximations. The radius/domain of convergence of the series in Equation (3) determines the largest values of the parameter variations $\delta\alpha_j$ which are admissible before the respective series becomes divergent. In turn, these maximum admissible parameter variations limit, through Equation (3), the largest parameter co-

variances/standard deviations which can be considered for using the Taylor-expansion for the subsequent purposes of computing moments of the distribution of computed responses.

As is well known, and as indicated by Equation (3), the Taylor-series of a function of TP-variables [e.g., $r_k(\boldsymbol{\alpha})$] comprises TP 1st-order derivatives, $TP(TP+1)/2$ distinct 2nd-order derivatives, and so on. The computation by conventional methods of the nth-order functional derivatives (called "sensitivities" in the field of sensitivity analysis) of a response with respect to the TP-parameters (on which it depends) would require at least $O(TP^n)$ large-scale computations. The exponential increase with the order of response sensitivities of the number of large-scale computations needed to determine higher-order sensitivities is the manifestation of the "curse of dimensionality in sensitivity analysis," by analogy to the expression coined by Bellman [13] to express the difficulty of using "brute-force" grid search when optimizing a function with many input variables. The "nth-order Comprehensive Adjoint Sensitivity Analysis Methodology for Nonlinear Systems" (nth-CASAM-N) conceived by Cacuci [14] and the "nth-order Comprehensive Adjoint Sensitivity Analysis Methodology for Response-Coupled Forward/Adjoint Linear Systems" (nth-CASAM-L) conceived by Cacuci [15] are currently the only methodologies that enable the exact and efficient computation of arbitrarily high-order sensitivities while overcoming the curse of dimensionality.

Uncertainties in the model's parameters will evidently give rise to uncertainties in the computed model responses $r_k(\boldsymbol{\alpha})$. The computed model responses are considered to be distributed according to an unknown distribution denoted as $p_c(\boldsymbol{r})$. The unknown joint probability distribution of model parameters and responses will be denoted as $p_c(\boldsymbol{\alpha},\boldsymbol{r}) \triangleq p_\alpha(\boldsymbol{\alpha}) p_c(\boldsymbol{r})$. The approximate moments of the unknown distribution of $r_k(\boldsymbol{\alpha})$ are obtained by using the so-called "propagation of errors" methodology, which entails the formal integration over $p_c(\boldsymbol{\alpha},\boldsymbol{r}) \triangleq p_\alpha(\boldsymbol{\alpha}) p_c(\boldsymbol{r})$ of various expressions involving the truncated Taylor-series expansion of the response provided in Equation (3). This procedure was first used by Tukey [16]. Tukey's results were generalized to 6th-order by Cacuci [15].

The expectation value, $E_c(r_k)$, of a computed response $r_k(\boldsymbol{\alpha})$ is obtained by integrating formally Equation (3) over $p_c(\boldsymbol{\alpha},\boldsymbol{r})$, which yields the following expression:

$$
\begin{aligned}
E_c(r_k) &\triangleq \int r_k(\boldsymbol{\alpha}) p_c(\boldsymbol{\alpha},\boldsymbol{r}) \mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{r} \\
&= r_k(\boldsymbol{\alpha}^0) + \frac{1}{2}\sum_{j_1=1}^{TP}\sum_{j_2=1}^{TP}\left\{\frac{\partial^2 r_k(\boldsymbol{\alpha})}{\partial\alpha_{j_1}\partial\alpha_{j_2}}\right\}_{\boldsymbol{\alpha}^0} \mathrm{cov}\left(\alpha_{j_1},\alpha_{j_2}\right) \\
&\quad + \frac{1}{6}\sum_{j_1=1}^{TP}\sum_{j_2=1}^{TP}\sum_{j_3=1}^{TP}\left\{\frac{\partial^3 r_k(\boldsymbol{\alpha})}{\partial\alpha_{j_1}\partial\alpha_{j_2}\partial\alpha_{j_3}}\right\}_{\boldsymbol{\alpha}^0} t_{j_1 j_2 j_3}\sigma_{j_1}\sigma_{j_2}\sigma_{j_3} \\
&\quad + \frac{1}{24}\sum_{j_1=1}^{TP}\sum_{j_2=1}^{TP}\sum_{j_3=1}^{TP}\sum_{j_4=1}^{TP}\left\{\frac{\partial^4 r_k(\boldsymbol{\alpha})}{\partial\alpha_{j_1}\partial\alpha_{j_2}\partial\alpha_{j_3}\partial\alpha_{j_4}}\right\}_{\boldsymbol{\alpha}^0} q_{j_1 j_2 j_3 j_4}\sigma_{j_1}\sigma_{j_2}\sigma_{j_3}\sigma_{j_4} \\
&\quad + O\left(\sigma_{j_1}\sigma_{j_2}\sigma_{j_3}\sigma_{j_4}\sigma_{j_5}\right).
\end{aligned} \tag{4}
$$

The expectation values $E_c(r_k)$, $k = 1, \cdots, TR$, are considered to be the components of a vector defined as follows: $\boldsymbol{E}_c(\boldsymbol{r}) \triangleq [E_c(r_1), \cdots, E_c(r_k), \cdots, E_c(r_{TR})]^\dagger$.

The expression of the correlation between a computed responses and a parameter variance, which will be denoted as $\text{cor}(\alpha_i, r_k)$, is presented below:

$$
\begin{aligned}
\text{cor}(\alpha_i, r_k) &\triangleq \int \delta\alpha_i [r_k(\boldsymbol{\alpha}) - E_c(r_k)] p_c(\boldsymbol{\alpha}, \boldsymbol{r}) \mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{r} \\
&= \sum_{j=1}^{TP} \left\{ \frac{\partial r_k(\boldsymbol{\alpha})}{\partial \alpha_j} \right\}_{\boldsymbol{\alpha}^0} \text{cov}(\alpha_i, \alpha_j) + \frac{1}{2} \sum_{j_1=1}^{TP} \sum_{j_2=1}^{TP} \left\{ \frac{\partial^2 r_k(\boldsymbol{\alpha})}{\partial \alpha_{j_1} \partial \alpha_{j_2}} \right\}_{\boldsymbol{\alpha}^0} t_{i,j_1 j_2} \sigma_i \sigma_{j_1} \sigma_{j_2} \\
&\quad + \frac{1}{6} \sum_{j_1=1}^{TP} \sum_{j_2=1}^{TP} \sum_{j_3=1}^{TP} \left\{ \frac{\partial^3 r_k(\boldsymbol{\alpha})}{\partial \alpha_{j_1} \partial \alpha_{j_2} \partial \alpha_{j_3}} \right\}_{\boldsymbol{\alpha}^0} q_{i,j_1 j_2 j_3} \sigma_{j_1} \sigma_{j_2} \sigma_{j_3} \sigma_{j_4} \\
&\quad + O(\sigma_{j_1} \sigma_{j_2} \sigma_{j_3} \sigma_{j_4} \sigma_{j_5}).
\end{aligned} \tag{5}
$$

The correlations $\text{cor}(\alpha_i, r_k)$, $i = 1, \cdots, TP$, $k = 1, \cdots, TR$, are considered to be the components of a "parameter-response computed correlation matrix" denoted as $\boldsymbol{C}_{\alpha r}^c$ and defined as follows:

$$
\boldsymbol{C}_{\alpha r}^c \triangleq \begin{pmatrix} \text{cor}(\alpha_1, r_1) & \cdots & \text{cor}(\alpha_1, r_{TR}) \\ \vdots & \ddots & \vdots \\ \text{cor}(\alpha_{TP}, r_1) & \cdots & \text{cor}(\alpha_{TP}, r_{TR}) \end{pmatrix}. \tag{6}
$$

The expression of the covariance between two responses r_k and r_ℓ, denoted as $\text{cov}(r_k, r_\ell)$, is presented below:

$$
\begin{aligned}
\text{cov}(r_k, r_\ell) &\triangleq \int [r_k(\boldsymbol{\alpha}) - E_c(r_k)][r_\ell(\boldsymbol{\alpha}) - E_c(r_\ell)] p_c(\boldsymbol{\alpha}, \boldsymbol{r}) \mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{r} \\
&= \sum_{i=1}^{TP} \sum_{j=1}^{TP} \left\{ \frac{\partial r_k(\boldsymbol{\alpha})}{\partial \alpha_i} \frac{\partial r_\ell(\boldsymbol{\alpha})}{\partial \alpha_j} \right\}_{\boldsymbol{\alpha}^0} \text{cov}(\alpha_i, \alpha_j) \\
&\quad + \frac{1}{2} \sum_{j_1=1}^{TP} \sum_{j_2=1}^{TP} \sum_{j_3=1}^{TP} \left\{ \frac{\partial^2 r_k(\boldsymbol{\alpha})}{\partial \alpha_{j_1} \partial \alpha_{j_2}} \frac{\partial r_\ell(\boldsymbol{\alpha})}{\partial \alpha_{j_3}} + \frac{\partial r_k(\boldsymbol{\alpha})}{\partial \alpha_{j_1}} \frac{\partial^2 r_\ell(\boldsymbol{\alpha})}{\partial \alpha_{j_2} \partial \alpha_{j_3}} \right\}_{\boldsymbol{\alpha}^0} t_{j_1 j_2 j_3} \sigma_{j_1} \sigma_{j_2} \sigma_{j_3} \\
&\quad + \frac{1}{4} \sum_{j_1=1}^{TP} \sum_{j_2=1}^{TP} \sum_{j_3=1}^{TP} \sum_{j_4=1}^{TP} \left\{ \frac{\partial^2 r_k(\boldsymbol{\alpha})}{\partial \alpha_{j_1} \partial \alpha_{j_2}} \frac{\partial^2 r_\ell(\boldsymbol{\alpha})}{\partial \alpha_{j_2} \partial \alpha_{j_3}} \right\}_{\boldsymbol{\alpha}^0} (q_{j_1 j_2 j_3 j_4} - \rho_{j_1 j_2} \rho_{j_3 j_4}) \sigma_{j_1} \sigma_{j_2} \sigma_{j_3} \sigma_{j_4} \\
&\quad + \frac{1}{6} \sum_{j_1=1}^{TP} \sum_{j_2=1}^{TP} \sum_{j_3=1}^{TP} \sum_{j_4=1}^{TP} \left\{ \frac{\partial^3 r_k(\boldsymbol{\alpha})}{\partial \alpha_{j_1} \partial \alpha_{j_2} \partial \alpha_{j_3}} \frac{\partial r_\ell(\boldsymbol{\alpha})}{\partial \alpha_{j_4}} + \frac{\partial r_k(\boldsymbol{\alpha})}{\partial \alpha_{j_1}} \frac{\partial^3 r_k(\boldsymbol{\alpha})}{\partial \alpha_{j_2} \partial \alpha_{j_3} \partial \alpha_{j_4}} \right\}_{\boldsymbol{\alpha}^0} q_{j_1 j_2 j_3 j_4} \sigma_{j_1} \sigma_{j_2} \sigma_{j_3} \sigma_{j_4} \\
&\quad + O(\sigma_{j_1} \sigma_{j_2} \sigma_{j_3} \sigma_{j_4} \sigma_{j_5}).
\end{aligned} \tag{7}
$$

The covariances $\text{cov}(r_k, r_\ell)$, $k, \ell = 1, \cdots, TR$, are considered to be the components of a "computed-responses covariance matrix" denoted as $\boldsymbol{C}_{rr}^c$ and defined as follows:

$$
\boldsymbol{C}_{rr}^c \triangleq \begin{pmatrix} \text{cor}(r_1, r_1) & \cdots & \text{cor}(r_1, r_{TR}) \\ \vdots & \ddots & \vdots \\ \text{cor}(r_{TP}, r_1) & \cdots & \text{cor}(r_{TP}, r_{TR}) \end{pmatrix}. \tag{8}
$$

The information provided in Equations (1)-(8) regarding the mean values and correlations for the model parameters and computed model responses will be used to construct a second-order MaxEnt approximation, denoted as $p_c(z|z_c, \boldsymbol{C}_c)$, to represent the distribution $p_c(\boldsymbol{\alpha}, \boldsymbol{r}) \triangleq p_\alpha(\boldsymbol{\alpha}) p_c(\boldsymbol{r})$ of computed model res-

ponses and parameters. The distribution $p_c(z|z_c, C_c)$ is constructed by following the procedure outlined in Appendix, to obtain the following expression:

$$p_c(\boldsymbol{\alpha}, \boldsymbol{r}) \cong p_c(z|z_c, C_c) \triangleq \frac{(2\pi)^{-(TR+TP)/2}}{\sqrt{Det(C_c)}} \exp\left[-\frac{1}{2}(z-z_c)^\dagger C_c^{-1}(z-z_c)\right], \quad (9)$$

where:

$$C_c \triangleq \begin{pmatrix} C_{rr}^c & C_{r\alpha}^c \\ C_{\alpha r}^c & C_{\alpha\alpha}^c \end{pmatrix}; \quad C_{r\alpha}^c \triangleq (C_{\alpha r}^c)^\dagger; \quad z = \begin{pmatrix} \boldsymbol{r} \\ \boldsymbol{\alpha} \end{pmatrix}; \quad z_c = \begin{pmatrix} E_c(\boldsymbol{r}) \\ \boldsymbol{\alpha}^0 \end{pmatrix}. \quad (10)$$

It is important to note that even though the MaxEnt distribution represented in Equation (9) is a multivariate normal distribution characterized by just the first-order moments (mean values) and second-order moments (variance/covariances) of the full but unknown "joint distribution $p_c(\boldsymbol{\alpha}, \boldsymbol{r})$ of computed responses and parameters," the components of the vector z_c of mean values and the components of the covariance matrix C_c may comprise terms involving third- and higher-order response sensitivities and parameter correlations (e.g., triple and quadruple correlations), if available, as indicated in Equations (4), (5), and (7).

3. 2nd-BERRU-PMP: Second Order Predictive Modeling Methodology with Probabilistically Included Computed Responses

This Section presents the mathematical and physical considerations leading to the development of the second-order predictive modeling methodology in which the computational model is probabilistically incorporated with the measurements into the combined 2nd-order MaxEnt posterior distribution which represents all of the available computational and experimental information. This methodology will be designated using the acronym 2nd-BERRU-PMP, where the last letter ("P") indicates "probabilistic." Subsection 3.1 presents the general case, in which measurements (mean values and correlations) for parameters are available−in addition to the information used in the computational model−for incorporation together with the measured responses into the MaxEnt probabilistic representation of the computational model in order to finally obtain the joint posterior distribution that would represent all of the available second-order information. This general case seldom occurs in practice because additional measurements regarding the model parameters (outside of, and in addition to, their use in the model) are rarely available. Usually, only measurements for responses are available for incorporation into the MaxEnt representation of the computational model; this case is analyzed in Subsection 3.2.

3.1. General Case: Measurements for Both Responses and Parameters Are Available for Combination with the MaxEnt Probabilistic Representation of the Computational Model to Obtain Their Joint Posterior Distribution

In the most general case, measurements could be available not only for the res-

ponses (results) of interest but also about the parameters used in the model to compute the respective results. In such a case, the second-order MaxEnt distribution which will represent probabilistically the computed model responses and model parameters is to be combined with a second-order MaxEnt distribution which is to represent probabilistically the measured model responses and the additionally measured model parameters (*i.e.*, model parameters measured in addition to, and independently of, the computational model). Since only first- and second-order moments are considered to be available, each of these MaxEnt distributions will be multivariate normal distributions, so their combination will also yield a multivariate posterior MaxEnt distribution (for all of the first- and second-order available moments).

Consider that the total number of experimentally measured responses is *TR*. The information usually available regarding the distribution of such measured responses comprises the first-order moments (mean values), which will be denoted as r_i^e, $i = 1, \cdots, TR$, and the second-order moments (variances/covariances), which will be denoted as $\text{cov}(r_i, r_j)_e$, $i, j = 1, \cdots, TR$, for the measured responses. Occasionally, measured correlations between the model parameters and the measured responses could also be available. The letter "e" will be used either as a superscript or a superscript to indicate experimentally measured quantities. The expected values of the measured responses will be considered to constitute the components of a vector denoted as $r^e \triangleq (r_1^e, \cdots, r_{TR}^e)^\dagger$. The covariances of the measured responses are considered to be components of the $TR \times TR$-dimensional covariance matrix of measured responses, which will be denoted as $C_{rr}^e \triangleq [\text{cov}(r_i, r_j)_e]_{TR \times TR}$. In principle, it is also possible to obtain correlations between some measured responses and some model parameters. When such correlations between measured responses and measured model parameters are available, they are defined as $\text{cor}(\alpha_i, r_j)_e$, $i = 1, \cdots, TP$; $j = 1, \cdots, TR$, and can formally be considered to be elements of a rectangular correlation matrix which will be denoted as $C_{\alpha r}^e \triangleq [\text{cov}(\alpha_i, r_j)_e]_{TP \times TR}$. In the most general situation, it is possible to have not only measured responses but also new measurements for the parameters, in addition to and independently of, the parameters used in the mathematical/computational model (*i.e.*, where they are considered to have nominal values α^0 and covariance matrix $C_{\alpha\alpha}$). These measured parameters will be considered to have a vector of mean values denoted as $\alpha^e \triangleq (\alpha_1^e, \cdots, \alpha_{TP}^e)$, and a covariance matrix denoted as $C_{\alpha r}^e \triangleq [\text{cov}(\alpha_i, r_j)_e]_{TP \times TR}$.

The MaxEnt principle can now be applied, as described in Appendix, to construct the least informative (and hence, most conservative) distribution using the available experimental information, to obtain the following expression:

$$p_e(z|z_e, C_e) = \frac{(2\pi)^{-(TR+TP)/2}}{\sqrt{Det(C_e)}} \exp\left[-\frac{1}{2}(z - z_e)^\dagger C_e^{-1}(z - z_e)\right], \quad (11)$$

where:

$$C_e \triangleq \begin{pmatrix} C_{rr}^e & C_{r\alpha}^e \\ C_{\alpha r}^e & C_{\alpha\alpha}^e \end{pmatrix}; \; z = \begin{pmatrix} r \\ \alpha \end{pmatrix}; \; z_e = \begin{pmatrix} r^e \\ \alpha^e \end{pmatrix}. \quad (12)$$

The joint posterior probability distribution of all computed and experimentally measured quantities, which will be denoted as $p_p(z|z_p, C_p)$ where the subscript "p" indicates "posterior," is obtained as the properly normalized product of the distributions $p_e(z|z_e, C_e)$ and $p_c(z|z_c, C_c)$. Since both $p_e(z|z_e, C_e)$ and $p_c(z|z_c, C_c)$ are normally distributed, it follows that the posterior probability $p_p(z|z_p, C_p)$ will also be normally distributed and will be given by the following expression:

$$p_p(z|z_p, C_p) \triangleq p_c(z|z_c, C_c) p_e(z|z_e, C_e) = K \exp\left[-\frac{1}{2}Q(z|z_p, C_p)\right], \quad (13)$$

where the normalization constant K and the quadratic form $Q(z_p, C_p)$ have the following expressions, respectively:

$$K \triangleq (2\pi)^{-N/2} \left[Det(C_c + C_e)\right]^{-1/2} \exp\left[-\frac{1}{2}(z_c - z_e)^\dagger (C_c + C_e)^{-1}(z_c - z_e)\right]; \quad (14)$$

$$Q(z|z_p, C_p) \triangleq (z - z_p)^\dagger C_p^{-1}(z - z_p); \quad (15)$$

$$C_p \triangleq \left(C_c^{-1} + C_e^{-1}\right)^{-1} = C_c - C_c(C_c + C_e)^{-1}C_c = C_e - C_e(C_c + C_e)^{-1}C_e; \quad (16)$$

$$z_p \triangleq C_p\left(C_c^{-1}z_c + C_e^{-1}z_e\right). \quad (17)$$

The expression obtained in Equation (13) provides the exact first-order (mean values) and second-order (variances and co-variances) of the most comprehensive combined distribution of computations and measurements of responses and parameters. In practice, however, such a comprehensive amount of experimental information is highly unlikely to be available. Furthermore, even if such massive amount of experimental information were available, the inversion of the matrix $(C_c + C_e)^{-1}$ may be impractical since it would require massive computational resources when the physical systems involve many (thousands of) parameters.

3.2. Practical Case: Only Response Measurements Are Available for Combination with the MaxEnt Probabilistic Representation of the Computational Model to Obtain Their Joint Posterior Distribution

In practice, the information (mean values and covariances) about the model parameters indicated in Equations (1) and (2) is obtained and assessed prior to using this information in the mathematical/computational model used for computing responses. Thus, all of the information available regarding the model parameters would be used to construct the components of the vector $\boldsymbol{\alpha}^0 \triangleq \left(\alpha_1^0, \cdots, \alpha_{TP}^0\right)^\dagger$ of parameter nominal values and the components of the parameter covariance matrix $C_{\alpha\alpha}$. In practice, only information about measured responses would become additionally available, i.e., only the vector $\boldsymbol{r}^e \triangleq \left(r_1^e, \cdots, r_{TR}^e\right)^\dagger$ and the covariance matrix $C_{rr}^e \triangleq \left[\text{cov}(r_i, r_j)_e\right]_{TR \times TR}$ would become available for combination with the computational information. In such a case, the MaxEnt distribution corresponding to Equation (11) will reduce to the following expression:

$$p_e\left(\boldsymbol{r}\middle|\boldsymbol{r}^e,\boldsymbol{C}_{rr}^e\right)=\left(2\pi\right)^{-TR/2}\left[Det\left(\boldsymbol{C}_{rr}^e\right)\right]^{-1/2}\exp\left[-\frac{1}{2}\left(\boldsymbol{r}-\boldsymbol{r}^e\right)^\dagger\left(\boldsymbol{C}_{rr}^e\right)^{-1}\left(\boldsymbol{r}-\boldsymbol{r}^e\right)\right]. \quad (18)$$

Furthermore, when only the experimental information represented by the distribution $p_e\left(\boldsymbol{r}\middle|\boldsymbol{r}^e,\boldsymbol{C}_{rr}^e\right)$ is available, the posterior joint probability distribution of the computed and measured quantities, which will be denoted as $p_p(\boldsymbol{r},\boldsymbol{\alpha})$ will have the following form:

$$p_p(\boldsymbol{r},\boldsymbol{\alpha})=N\exp\left[-\frac{1}{2}Q(\boldsymbol{r},\boldsymbol{\alpha})\right];\ N\triangleq\left[\int\exp\left[-\frac{1}{2}Q(\boldsymbol{r},\boldsymbol{\alpha})\right]\mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{r}\right]^{-1};$$

$$Q(\boldsymbol{r},\boldsymbol{\alpha})\triangleq\left(\boldsymbol{z}-\boldsymbol{z}_c\right)^\dagger\boldsymbol{C}_c^{-1}\left(\boldsymbol{z}-\boldsymbol{z}_c\right)+\left(\boldsymbol{r}-\boldsymbol{r}^e\right)^\dagger\left(\boldsymbol{C}_{rr}^e\right)^{-1}\left(\boldsymbol{r}-\boldsymbol{r}^e\right). \quad (19)$$

The evaluation of the moments of the posterior distribution represented by Equation (19) will involve the evaluation of ratios of integrals having the following form:

$$I=\left[\int\exp\left[-g\left(\boldsymbol{z}\right)\right]\mathrm{d}\boldsymbol{z}\right]^{-1}\int f\left(\boldsymbol{z}\right)\exp\left[-g\left(\boldsymbol{z}\right)\right]\mathrm{d}\boldsymbol{z}. \quad (20)$$

The evaluations of expressions such as shown in Equation (20) can be performed to a high degree of accuracy, with a controlled error, by employing the saddle-point (Laplace) method, see e.g., [17] [18], which yields the following result:

$$I=\hat{f}-\frac{1}{2}\hat{f}_{i_1}\hat{g}_{i_2i_3i_4}\hat{g}^{i_1i_2}\hat{g}^{i_3i_4}+\frac{1}{2}\hat{f}_{j_1j_2}\hat{g}^{j_1j_2}+O\left(\hat{f}_{j_1j_2j_3}\right), \quad (21)$$

where:

1) The derivative of a function with respect to a component of $\boldsymbol{z}$ is denoted using a subscript, e.g., $f_i\triangleq\partial f/\partial z_i$, $f_{ij}\triangleq\partial^2 f/\partial z_i\partial z_j$, $i,j=1,\cdots,TI$, where TI denotes the total number of independent variables;

2) The superscripts denote the respective component of the inverse Hessian of the respective function, e.g., f^{ij} denotes the (i,j)-element of the inverse Hessian matrix $\left[f_{ij}^{-1}\right]$;

3) An index that appears as a subscript and a superscript implies a summation over all possible values of that index;

4) The "hat" denotes that the respective quantity is to be evaluated at the saddle point of $\exp\left[-g\left(\boldsymbol{z}\right)\right]$, which is defined as the point at which the gradient $g_z\left(\boldsymbol{z}\right)\triangleq\left(\partial g/\partial z_1,\cdots,\partial g/\partial z_{TI}\right)$ of $g\left(\boldsymbol{z}\right)$ vanishes, i.e., $g_z\left(\boldsymbol{z}\right)=\boldsymbol{0}$.

The saddle-point of the normalization integral N will be denoted as $\left(\boldsymbol{r}_s,\boldsymbol{\alpha}_s\right)$ and is defined by the following relations:

$$\frac{\partial Q(\boldsymbol{r},\boldsymbol{\alpha})}{\partial\boldsymbol{r}}=\boldsymbol{0},\ \frac{\partial Q(\boldsymbol{r},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}=\boldsymbol{0},\ \text{at}\ \left(\boldsymbol{r},\boldsymbol{\alpha}\right)=\left(\boldsymbol{r}_s,\boldsymbol{\alpha}_s\right). \quad (22)$$

To obtain the partial gradients (differentials) shown in Equation (22), it is convenient to write the matrix $\boldsymbol{C}_c^{-1}$ in the form $\boldsymbol{C}_c^{-1}=\begin{pmatrix}\boldsymbol{C}_{11}&\boldsymbol{C}_{12}\\\boldsymbol{C}_{21}&\boldsymbol{C}_{22}\end{pmatrix}$, and use this form together with the definitions provided in Equations (10) and (12) in Equation (19) to expand the functional $Q(\boldsymbol{r},\boldsymbol{\alpha})$ into the following form:

$$Q(r,\alpha) = \left[r - E_c(r)\right]^\dagger C_{11}\left[r - E_c(r)\right] + \left[r - E_c(r)\right]^\dagger C_{12}\left(\alpha - \alpha^0\right)$$
$$+ \left(\alpha - \alpha^0\right)^\dagger C_{21}\left[r - E_c(r)\right] + \left(\alpha - \alpha^0\right)^\dagger C_{22}\left(\alpha - \alpha^0\right) \qquad (23)$$
$$+ \left(r - r^e\right)^\dagger \left(C_{rr}^e\right)^{-1}\left(r - r^e\right).$$

Taking the partial differentials of the expression in Equation (23) yields the following equation for determining the coordinates of the saddle point $(r,\alpha) = (r_s,\alpha_s)$:

$$\begin{pmatrix} r_s - E_c(r) \\ \alpha_s - \alpha^0 \end{pmatrix} = -\begin{pmatrix} C_{rr}^c & C_{r\alpha}^c \\ C_{\alpha r}^c & C_{\alpha\alpha}^c \end{pmatrix}\begin{pmatrix} \left(C_{rr}^e\right)^{-1}\left(r_s - r^e\right) \\ \mathbf{0} \end{pmatrix}. \qquad (24)$$

Solving Equation (24) leads to the following expressions for the coordinates r_s and α_s of the saddle point:

$$r_s = r^e + C_{rr}^e\left(C_{rr}^e + C_{rr}^c\right)^{-1}\left[E_c(r) - r^e\right], \qquad (25)$$

$$\alpha_s = \alpha^0 - C_{\alpha r}^c\left(C_{rr}^e + C_{rr}^c\right)^{-1}\left[E_c(r) - r^e\right]. \qquad (26)$$

The best-estimate predicted mean values for the responses and parameters are defined below:

$$r^{bep} \triangleq \int_{D_p} r\, p_p(\alpha,r)\,d\alpha dr , \qquad (27)$$

$$\alpha^{bep} \triangleq \int_{D_p} \alpha\, p_p(\alpha,r)\,d\alpha dr . \qquad (28)$$

The superscript "bep" has been used in Equations (29) and (30), respectively, to indicate that the expressions obtained for the respective predicted responses and parameters are "best estimates with probabilistic computational model." The saddle-point method is applied to evaluate the integrals represented by Equations (27) and (28), respectively, to obtain the following expressions for the optimally-predicted "best-estimate" values for the responses and calibrated parameters:

$$r^{bep} = r_s = r^e + C_{rr}^e\left(C_{rr}^e + C_{rr}^c\right)^{-1}\left[E_c(r) - r^e\right], \qquad (29)$$

$$\alpha^{bep} = \alpha_s = \alpha^0 - C_{\alpha r}^c\left(C_{rr}^e + C_{rr}^c\right)^{-1}\left[E_c(r) - r^e\right]. \qquad (30)$$

Since the components of the vector $E_c(r)$, and the components of the matrices C_{rr}^c and $C_{\alpha r}^c$ can contain arbitrarily high-order response sensitivities to model parameters, the expressions presented in Equations (29) and (30) generalize the previous formulas of this type found in data adjustment/assimilation procedures published to date (which contain at most second-order sensitivities). The best-estimate parameter values are the "calibrated model parameters" which can be used for subsequent computations with the "calibrated model."

The second-order moments of the posterior distribution $p_p(r,\alpha)$ comprise the covariances between the best estimated response, which will be denoted as C_{rr}^{bep}, the covariances between the best-estimate parameters, which will be de-

noted as $C_{\alpha\alpha}^{bep}$, and the correlations between the best-estimate parameters and responses, which will be denoted as $C_{\alpha r}^{bep}$. The expression of the "best-estimate probabilistic model" posterior parameter covariance matrix, C_{rr}^{bep}, for the best-estimate responses r^{bep} is derived by using the result given in Equation (29) to obtain the following expression:

$$C_{rr}^{bep} \triangleq \int_{D_p} \left(r - r^{bep}\right)\left(r - r^{bep}\right)^{\dagger} p_p\left(\alpha, r\right) \mathrm{d}\alpha \mathrm{d}r = C_{rr}^{e} - C_{rr}^{e}\left(C_{rr}^{e} + C_{rr}^{c}\right)^{-1} C_{rr}^{e}. \quad (31)$$

The following important result has been used to obtain the expression provided in Equation (31):

$$\begin{aligned}
&\left[E_c\left(r\right) - r^{e}\right]\left[E_c\left(r\right) - r^{e}\right]^{\dagger} \\
&= \left[r - r^{e} - r + E_c\left(r\right)\right]\left[r - r^{e} - r + E_c\left(r\right)\right]^{\dagger} = C_{rr}^{e} + C_{rr}^{c}.
\end{aligned} \quad (32)$$

As indicated in Equation (31), the initial covariance matrix C_{rr}^{e} for the experimentally measured responses is multiplied by the matrix $\left[I - \left(C_{rr}^{e} + C_{rr}^{c}\right)^{-1} C_{rr}^{e}\right]$, which means that the variances contained on the diagonal of the best-estimate matrix C_{rr}^{bep} will be smaller than the experimentally measured variances contained in C_{rr}^{e}. Hence, the incorporation of experimental information reduces the predicted best-estimate response variances in C_{rr}^{bep} by comparison to the measured variances contained a priori in C_{rr}^{e}. Since the components of the matrix C_{rr}^{bep} contain high-order sensitivities, the formula presented in Equation (31) generalizes the previous formulas of this type found in data adjustment/ assimilation procedures published to date.

The expression of the "best-estimate" posterior parameter covariance matrix $C_{\alpha\alpha}^{bep}$ for the best-estimate parameters α^{bep} is derived by using the result given in Equation (30) to obtain:

$$C_{\alpha\alpha}^{bep} \triangleq \int_{D_p} \left(\alpha - \alpha^{bep}\right)\left(\alpha - \alpha^{bep}\right)^{\dagger} p_p\left(\alpha, r\right) \mathrm{d}\alpha \mathrm{d}r = C_{\alpha\alpha} - C_{\alpha r}^{c}\left(C_{rr}^{e} + C_{rr}^{c}\right)^{-1} C_{\alpha r}^{c}. \quad (33)$$

The matrices $C_{\alpha\alpha}$ and $C_{\alpha r}^{c}\left(C_{rr}^{e} + C_{rr}^{c}\right)^{-1} C_{\alpha r}^{c}$ are symmetric and positive definite. Therefore, the subtraction indicated in Equation (33) implies that the components of the main diagonal of $C_{\alpha\alpha}^{bep}$ must have smaller values than the corresponding elements of the main diagonal of $C_{\alpha\alpha}$. In this sense, the combination of computational and experimental information has reduced the best-estimate parameter variances on the diagonal of $C_{\alpha\alpha}^{bep}$. Since the components of the matrices $C_{\alpha\alpha}$, $C_{\alpha r}^{c}$, and C_{rr}^{c} contain high-order response sensitivities, the formula presented in Equation (33) generalizes the previous formulas of this type found in data adjustment/assimilation procedures published to date.

The expressions of the "best-estimate" posterior parameter correlation matrix $C_{\alpha r}^{bep}$ and/or its transpose $C_{r\alpha}^{bep}$, for the best-estimate parameters α^{bep} and best-estimate responses r^{bep}, are derived by using the results given in Equation (29) and Equation (30) to obtain the following expressions:

$$C_{\alpha r}^{bep} \triangleq \int_{D_p} \left(\alpha - \alpha^{bep}\right)\left(r - r^{bep}\right)^{\dagger} p_p\left(\alpha, r\right) \mathrm{d}\alpha \mathrm{d}r = C_{\alpha r}^{e} - C_{\alpha r}^{c}\left(C_{rr}^{e} + C_{rr}^{c}\right)^{-1} C_{rr}^{e}; \quad (34)$$

$$C_{r\alpha}^{bep} \triangleq \int\limits_{D_p} \left(r - r^{bep}\right)\left(\alpha - \alpha^{bep}\right)^{\dagger} p_p\left(\alpha, r\right) \mathrm{d}\alpha \mathrm{d}r \tag{35}$$

$$= C_{r\alpha}^{e} - C_{rr}^{e}\left(C_{rr}^{e} + C_{rr}^{c}\right)^{-1} C_{r\alpha}^{c} = \left(C_{\alpha r}^{bep}\right)^{\dagger}.$$

Since the components of the matrices $C_{\alpha r}^{c}$ and C_{rr}^{c} contain high-order sensitivities, the formulas presented in Equations (34) and (35) generalize the previous formulas of this type found in data adjustment/assimilation procedures published to date.

It is important to note from the results shown in Equations (29)-(35) that the computation of the best estimate parameter and response values, together with their corresponding best-estimate covariance matrices, only involves a single matrix inversion when computing $\left(C_{rr}^{e} + C_{rr}^{c}\right)^{-1}$, which entails the inversion of a matrix of size $TR \times TR$. This is computationally very advantageous, since $TR \ll TP$, *i.e.*, the number of responses is much less than the number of model parameters in the overwhelming majority of practical situations.

Using Equations (24), (29) and (30) in Equation (23) yields the following expression for the minimum value, $Q_{\min} = Q\left(\alpha^{bep}, r^{bep}\right)$, of the quadratic form $Q\left(\alpha, r\right)$:

$$Q_{\min} = \begin{pmatrix} r^{bep} - E_c\left(r\right) \\ \alpha^{bep} - \alpha^{0} \end{pmatrix}^{\dagger} \begin{pmatrix} C_{rr}^{c} & C_{r\alpha}^{c} \\ C_{\alpha r}^{c} & C_{\alpha\alpha}^{c} \end{pmatrix}^{-1} \begin{pmatrix} r^{bep} - E_c\left(r\right) \\ \alpha^{bep} - \alpha^{0} \end{pmatrix}$$

$$+ \left(r^{bep} - r^{e}\right)^{\dagger} \left(C_{rr}^{e}\right)^{-1} \left(r^{bep} - r^{e}\right) \tag{36}$$

$$= \left[E_c\left(r\right) - r^{e}\right]^{\dagger} \left(C_{rr}^{e} + C_{rr}^{c}\right)^{-1} \left[E_c\left(r\right) - r^{e}\right].$$

As the expression obtained in Equation (36) indicates, the quantity $Q_{\min}$ represents the square of the length of the vector $\left[E_c\left(r\right) - r^{e}\right]$, measuring (in the corresponding metric) the deviations between the experimental and nominally computed responses. The quantity $Q_{\min}$ is independent of calibrating (or adjusting) the original data, so it can be evaluated directly from the given data (*i.e.*, model parameters and computed and measured responses, together with their original uncertainties) after having computed the matrix $\left(C_{rr}^{e} + C_{rr}^{c}\right)^{-1}$. As the dimension of the vector $\left[E_c\left(r\right) - r^{e}\right]$ indicates, the number of degrees of freedom characteristic of the calibration under consideration is equal to the number TR of experimental responses.

4. Inter-Comparison: 2nd-BERRU-PMP vs. 2nd-BERRU-PMD

In this section, the results obtained for the best-estimate mean values for the responses and parameters, together with their corresponding best estimate covariances/correlations will be compared to the corresponding results produced by the 2nd-BERRU-PMD methodology, as well as to the corresponding results produced by the 1st-BERRU-PM [10] [11] [12].

4.1. Inter-Comparison of Expressions for the Best-Estimate Predicted Mean Values for Responses

Recall that the 2nd-BERRU-PMD (best-estimated values with deterministically

incorporated computed responses) methodology incorporates deterministically (into the second-order MaxEnt representation of the experimentally measured responses) the following second-order Taylor-expansion of the computed responses:

$$r = r^c\left(\boldsymbol{\alpha}^0\right) + \left[S\left(\boldsymbol{\alpha}^0\right)\right]\left(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0\right) + \boldsymbol{p}\left(\boldsymbol{\alpha};\boldsymbol{\alpha}^0\right);$$

$$\boldsymbol{p}\left(\boldsymbol{\alpha};\boldsymbol{\alpha}^0\right) \triangleq \left[p_1\left(\boldsymbol{\alpha};\boldsymbol{\alpha}^0\right), \cdots, p_{TR}\left(\boldsymbol{\alpha};\boldsymbol{\alpha}^0\right)\right]^\dagger; \tag{37}$$

$$p_k\left(\boldsymbol{\alpha};\boldsymbol{\alpha}^0\right) \triangleq \frac{1}{2}\left(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0\right)^\dagger\left[\boldsymbol{Q}_k\left(\boldsymbol{\alpha}^0\right)\right]\left(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0\right).$$

In Equation (37), the components of the vector $r^c\left(\boldsymbol{\alpha}^0\right) \triangleq \left[r_1^c\left(\boldsymbol{\alpha}^0\right), \cdots, r_{TR}^c\left(\boldsymbol{\alpha}^0\right)\right]^\dagger$ represent the values of the model responses computed at the nominal parameter values $\boldsymbol{\alpha}^0$, and the other vectors and matrices contain first-order and second-order sensitivities of responses with respect to the model parameters, evaluated at the nominal parameter values $\boldsymbol{\alpha}^0$, as defined below:

$$S(\boldsymbol{\alpha}) \triangleq \begin{pmatrix} s_{11} & \cdots & s_{1,TR} \\ \vdots & \ddots & \vdots \\ s_{TR,1} & \cdots & s_{TR,TR} \end{pmatrix} \triangleq \begin{pmatrix} \dfrac{\partial r_1(\boldsymbol{\alpha})}{\partial \alpha_1} & \cdots & \dfrac{\partial r_1(\boldsymbol{\alpha})}{\partial \alpha_{TP}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial r_{TR}(\boldsymbol{\alpha})}{\partial \alpha_1} & \cdots & \dfrac{\partial r_{TR}(\boldsymbol{\alpha})}{\partial \alpha_{TP}} \end{pmatrix};$$

$$\boldsymbol{Q}_k(\boldsymbol{\alpha}) \triangleq \begin{pmatrix} q_{11}^{(k)} & \cdots & q_{1,TP}^{(k)} \\ \vdots & \ddots & \vdots \\ q_{TP,1}^{(k)} & \cdots & q_{TP,TP}^{(k)} \end{pmatrix} \triangleq \begin{pmatrix} \dfrac{\partial^2 r_k^c(\boldsymbol{\alpha})}{\partial \alpha_1 \partial \alpha_1} & \cdots & \dfrac{\partial^2 r_k^c(\boldsymbol{\alpha})}{\partial \alpha_1 \partial \alpha_{TP}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 r_k^c(\boldsymbol{\alpha})}{\partial \alpha_{TP} \partial \alpha_1} & \cdots & \dfrac{\partial^2 r_k^c(\boldsymbol{\alpha})}{\partial \alpha_{TP} \partial \alpha_{TP}} \end{pmatrix}. \tag{38}$$

The expressions of the end-results produced the "best-estimated values with deterministically incorporated computed responses" (2nd-BERRU-PMD) methodology will be designated using the superscript "bed." The expression obtained using the 2nd-BERRU-PMD methodology for the best-estimate responses is as follows [1]:

$$r^{bed} = r^e + \left(C_{rr}^e - C_{r\alpha}^e S^\dagger\right)\boldsymbol{\theta}^{(h,n+1)}, \tag{39}$$

where the vector $\boldsymbol{\theta}^{(h,n+1)} \triangleq \left(\theta_1^{(h,n+1)}, \cdots, \theta_k^{(h,n+1)}, \cdots, \theta_{TR}^{(h,n+1)}\right)^\dagger$ is the solution of the following iterative (Newton's method) equation

$$\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} - \left[A\left(\boldsymbol{\theta}^{(n)}\right)\right]^{-1} \boldsymbol{w}\left(\boldsymbol{\theta}^{(n)}\right); \quad n = 0,1,\cdots. \tag{40}$$

In Equation (40), the components/elements $A_{k,n}(\boldsymbol{\theta})$ of the Jacobian matrix $A(\boldsymbol{\theta})$ have the following expressions:

$$A_{kn}(\boldsymbol{\theta}) = D_{kn} - V_{kn}(\boldsymbol{\theta}), \tag{41}$$

where:

$$D_{kj} \triangleq \left(C_{rr}^e + SC_{\alpha\alpha}S^\dagger - C_{r\alpha}^e S^\dagger - SC_{\alpha r}^e\right)_{kj}, \quad \boldsymbol{D} \triangleq \left[D_{kj}\right]_{TR \times TR}; \tag{42}$$

$$V_{kn}(\boldsymbol{\theta}) \triangleq \sum_{j=1}^{TR} \theta_j \left[M_{nj}^{(k)} + M_{kj}^{(n)} + M_{kn}^{(j)} \right], \; \boldsymbol{V}(\boldsymbol{\theta}) \triangleq \left[V_{kn}(\boldsymbol{\theta}) \right]_{TR \times TR}. \tag{43}$$

$$\boldsymbol{M}^{(k)} \triangleq \left[M_{ij}^{(k)} \right]_{TR \times TR} \triangleq \left(\boldsymbol{C}_{\alpha r}^{e} - \boldsymbol{C}_{\alpha \alpha} \boldsymbol{S}^{\dagger} \right)^{\dagger} \left[\boldsymbol{Q}_k (\boldsymbol{\alpha}^0) \right] \left(\boldsymbol{C}_{\alpha r}^{e} - \boldsymbol{C}_{\alpha \alpha} \boldsymbol{S}^{\dagger} \right) = \left[\boldsymbol{M}^{(k)} \right]^{\dagger}. \tag{44}$$

$$\boldsymbol{w}(\boldsymbol{\theta}) \triangleq \left[w_1(\boldsymbol{\theta}), \cdots, w_k(\boldsymbol{\theta}), \cdots, w_{TR}(\boldsymbol{\theta}) \right]^{\dagger} \triangleq -\boldsymbol{d} + \boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{u}(\boldsymbol{\theta}). \tag{45}$$

$$\boldsymbol{d} \triangleq \boldsymbol{r}^c(\boldsymbol{\alpha}^0) - \boldsymbol{r}^e; \; \boldsymbol{d} \triangleq (d_1, \cdots, d_{TR})^{\dagger}; \; d_1 \triangleq r_i^c(\boldsymbol{\alpha}^0) - r_i^e; \; i = 1, \cdots, TR. \tag{46}$$

$$\boldsymbol{u}(\boldsymbol{\theta}) \triangleq \left[u_1(\boldsymbol{\theta}), \cdots, u_{TR}(\boldsymbol{\theta}) \right]^{\dagger}; \; u_k(\boldsymbol{\theta}) \triangleq -\frac{1}{2} \boldsymbol{\theta}^{\dagger} \boldsymbol{M}^{(k)} \boldsymbol{\theta} - \sum_{i=1}^{TR} \theta_i \sum_{j=1}^{TR} M_{kj}^{(i)} \theta_j. \tag{47}$$

The initial iterate for starting the iterations for obtaining solution of Equation (40) can be chosen [1] as follows:

$$\boldsymbol{\theta}^{(h,0)} \triangleq \left(\theta_1^{(h,0)}, \cdots, \theta_j^{(h,0)}, \cdots, \theta_{TR}^{(h,0)} \right)^{\dagger}; \; \theta_j^{(h,0)} = 2D_{jj} \Big/ 3M_{jj}^{(j)}; \; j = 1, \cdots, TR. \tag{48}$$

$$A_{kn}\left(\boldsymbol{\theta}^{(h,0)} \right) = D_{kn} - \sum_{j=1}^{TR} \theta_j^{(h,0)} \left[M_{nj}^{(k)} + M_{kj}^{(n)} + M_{kn}^{(j)} \right]. \tag{49}$$

$$w_k\left(\boldsymbol{\theta}^{(h,0)} \right) = u_k\left(\boldsymbol{\theta}^{(h,0)} \right) + \sum_{j=1}^{TR} D_{kj} \theta_j^{(h,0)} - d_k, \quad k = 1, \cdots, TR; \tag{50}$$

$$\begin{aligned} u_k\left(\boldsymbol{\theta}^{(h,0)} \right) &= -\frac{1}{2} \sum_{i=1}^{TR} \sum_{j=1}^{TR} \theta_i^{(h,0)} M_{ij}^{(k)} \theta_j^{(h,0)} - \sum_{i=1}^{TR} \theta_i^{(h,0)} \sum_{j=1}^{TR} M_{kj}^{(i)} \theta_j^{(h,0)} \\ &\cong -\frac{4}{9} \sum_{i=1}^{TR} \sum_{j=1}^{TR} \left(\frac{D_{ii}}{M_{ii}^{(i)}} \right) \left(\frac{D_{jj}}{M_{jj}^{(j)}} \right) \left[\frac{1}{2} M_{ij}^{(k)} + M_{kj}^{(i)} \right], \quad k = 1, \cdots, TR. \end{aligned} \tag{51}$$

The first-iterate $\boldsymbol{\theta}^{(h,1)}$ of the solution of Equation (40) has the following expression:

$$\begin{aligned} \boldsymbol{\theta}^{(h,1)} &= \boldsymbol{\theta}^{(h,0)} - \left(\boldsymbol{D} - \boldsymbol{V}^{(h,0)} \right)^{-1} \left(\boldsymbol{u}^{(h,0)} + \boldsymbol{D}\boldsymbol{\theta}^{(h,0)} - \boldsymbol{d} \right); \\ \boldsymbol{V}^{(h,0)} &\triangleq \boldsymbol{V}\left(\boldsymbol{\theta}^{(h,0)} \right); \; \boldsymbol{u}^{(h,0)} \triangleq \boldsymbol{u}\left(\boldsymbol{\theta}^{(h,0)} \right). \end{aligned} \tag{52}$$

Using the first-iterate, $\boldsymbol{\theta}^{(h,1)}$, in Equation (39) yields the following expressions for the "first-iterate best-estimate" values for the responses:

$$\begin{aligned} \boldsymbol{r}^{(1)bed} &= \boldsymbol{r}^e + \boldsymbol{r}^{(0)}; \\ \boldsymbol{r}^{(0)} &\triangleq \left(\boldsymbol{C}_{rr}^e - \boldsymbol{C}_{r\alpha}^e \boldsymbol{S}^{\dagger} \right) \left[\boldsymbol{\theta}^{(h,0)} - \left(\boldsymbol{D} - \boldsymbol{V}^{(h,0)} \right)^{-1} \left(\boldsymbol{u}^{(h,0)} + \boldsymbol{D}\boldsymbol{\theta}^{(h,0)} - \boldsymbol{d} \right) \right]. \end{aligned} \tag{53}$$

It is noteworthy that the "vector of deviations" $\boldsymbol{d}$ can be considered to be a (quasi-) random variable; it has been shown in [1] that the "vector of mean values of deviations" $\bar{\boldsymbol{d}} \triangleq \left[\bar{d}_1, \cdots, \bar{d}_{TR} \right]^{\dagger}$ and the covariance matrix $\boldsymbol{D} \triangleq \left[D_{ij} \right]_{TR \times TR}$ of the vector of deviations $\boldsymbol{d}$ have the following expressions (up to second-order sensitivities):

$$\bar{\boldsymbol{d}} = \left[\boldsymbol{Q}_1(\boldsymbol{\alpha}^0) \boldsymbol{C}_{\alpha\alpha}, \cdots, \boldsymbol{Q}_{TR}(\boldsymbol{\alpha}^0) \boldsymbol{C}_{\alpha\alpha} \right]^{\dagger}; \tag{54}$$

$$\boldsymbol{D} = \boldsymbol{C}_{rr}^e + \boldsymbol{S}\boldsymbol{C}_{\alpha\alpha}\boldsymbol{S}^{\dagger} - \boldsymbol{C}_{r\alpha}^e \boldsymbol{S}^{\dagger} - \boldsymbol{S}\boldsymbol{C}_{\alpha r}^e. \tag{55}$$

Comparing the expression of the first-iterate, cf., Equation (53) produced by

the 2nd-BERRU-PMD methodology with the corresponding expression produced by the 2nd-BERRU-PMP methodology, cf. Equation (29), indicates that both expressions include the second-order sensitivities of the responses with respect to the parameters, but in distinct ways. Furthermore, the expression produced by the 2nd-BERRU-PMP methodology, cf. Equation (29), can include all of the response sensitivities and parameter correlations of order higher than two, if available. Such an inclusion of higher-order sensitivities and parameter correlations is not provided by the result, cf. Equation (53), obtained using the 2nd-BERRU-PMD methodology.

The expression provided in Equation (29) coincides with the expression provided in Equation (53) only if the following relationship holds:

$$
\begin{aligned}
\boldsymbol{r}^{bep} = \boldsymbol{r}^{(1)bed} \ \text{ iff}: \ & \boldsymbol{C}_{rr}^{e}\left(\boldsymbol{C}_{rr}^{e}+\boldsymbol{C}_{rr}^{c}\right)^{-1}\left[\boldsymbol{E}_{c}\left(\boldsymbol{r}\right)-\boldsymbol{r}^{e}\right] \\
& =\left(\boldsymbol{C}_{rr}^{e}-\boldsymbol{C}_{r\alpha}^{e}\boldsymbol{S}^{\dagger}\right)\left[\boldsymbol{\theta}^{(h,0)}-\left(\boldsymbol{D}-\boldsymbol{V}^{(h,0)}\right)^{-1}\left(\boldsymbol{u}^{(h,0)}+\boldsymbol{D}\boldsymbol{\theta}^{(h,0)}-\boldsymbol{d}\right)\right].
\end{aligned}
\tag{56}
$$

The equality in Equation (56) can hold if the following conditions are simultaneously satisfied: $\boldsymbol{C}_{r\alpha}^{e}=\boldsymbol{0}$, $\boldsymbol{\theta}^{(h,0)}=\boldsymbol{0}$, $\boldsymbol{V}^{(h,0)}=\boldsymbol{0}$, $\boldsymbol{u}^{(h,0)}=\boldsymbol{0}$, $\boldsymbol{C}_{rr}^{c}=\boldsymbol{S}\boldsymbol{C}_{\alpha\alpha}\boldsymbol{S}^{\dagger}$, $\boldsymbol{d}=\left[\boldsymbol{E}_{c}\left(\boldsymbol{r}\right)-\boldsymbol{r}^{e}\right]$, in which case we also have the relation $\boldsymbol{D}=\boldsymbol{C}_{rr}^{e}+\boldsymbol{S}\boldsymbol{C}_{\alpha\alpha}\boldsymbol{S}^{\dagger}=\boldsymbol{C}_{rr}^{e}+\boldsymbol{C}_{rr}^{c}$. These conditions are satisfied if all of the sensitivities higher than first-order are ignored, and if the experimental responses are uncorrelated to the model parameters. In this particular case, the results produced by the 2nd-BERRU-PMP and 2nd-BERRU-PMD methodologies also coincide with the results produced by the first-order 1st-BERRU-PM methodology.

4.2. Inter-Comparison of Expressions for the Best-Estimate Predicted Mean Values for Parameters

The expression obtained using the 2nd-BERRU-PMD methodology for the best-estimate parameters is as follows [1]:

$$
\boldsymbol{\alpha}^{bed}=\boldsymbol{\alpha}^{0}+\left(\boldsymbol{C}_{\alpha r}^{e}-\boldsymbol{C}_{\alpha\alpha}\boldsymbol{S}^{\dagger}\right)\boldsymbol{\theta}^{(h,n+1)}.
\tag{57}
$$

Using the first-iterate $\boldsymbol{\theta}^{(h,1)}$ in Equation (57) yields the following first-iterate expression for the best-estimate parameters:

$$
\begin{aligned}
\boldsymbol{\alpha}^{(1)bed} &=\boldsymbol{\alpha}^{0}+\boldsymbol{\beta}^{(0)}; \\
\boldsymbol{\beta}^{(0)} &\triangleq\left(\boldsymbol{C}_{\alpha r}^{e}-\boldsymbol{C}_{\alpha\alpha}\boldsymbol{S}^{\dagger}\right)\left[\boldsymbol{\theta}^{(h,0)}-\left(\boldsymbol{D}-\boldsymbol{V}^{(h,0)}\right)^{-1}\left(\boldsymbol{u}^{(h,0)}+\boldsymbol{D}\boldsymbol{\theta}^{(h,0)}-\boldsymbol{d}\right)\right].
\end{aligned}
\tag{58}
$$

Comparing the expression provided in Equation (58), namely the first-iterate produced by the 2nd-BERRU-PMD methodology, with the corresponding expression produced by the 2nd-BERRU-PMP methodology, cf. Equation (30), indicates that both expressions include the second-order sensitivities of the responses with respect to the parameters, but in distinct ways. Furthermore, the expression produced by the 2nd-BERRU-PMP methodology, cf. Equation (30), can include all of the response sensitivities and parameter correlations of order higher than two, if available. Such an inclusion of higher-order sensitivities and

parameter correlations is not possible with the result obtained using the 2nd-BERRU-PMD methodology, cf. Equation (58).

The expression provided in Equation (30) coincides with the expression provided in Equation (58) only if the following relationship holds:

$$
\begin{aligned}
\boldsymbol{\alpha}^{bep} = \boldsymbol{\alpha}^{(1)bed} \quad \text{iff}: \quad & -\boldsymbol{C}_{\alpha r}^{c}\left(\boldsymbol{C}_{rr}^{e}+\boldsymbol{C}_{rr}^{c}\right)^{-1}\left[\boldsymbol{E}_{c}\left(\boldsymbol{r}\right)-\boldsymbol{r}^{e}\right] \\
= & \left(\boldsymbol{C}_{\alpha r}^{e}-\boldsymbol{C}_{\alpha\alpha}\boldsymbol{S}^{\dagger}\right)\left[\boldsymbol{\theta}^{(h,0)}-\left(\boldsymbol{D}-\boldsymbol{V}^{(h,0)}\right)^{-1}\left(\boldsymbol{u}^{(h,0)}+\boldsymbol{D}\boldsymbol{\theta}^{(h,0)}-\boldsymbol{d}\right)\right].
\end{aligned}
\tag{59}
$$

The equality in Equation (59) can hold if the following conditions are simultaneously satisfied: $\boldsymbol{C}_{\alpha r}^{e}=\boldsymbol{0}$, $\boldsymbol{\theta}^{(h,0)}=\boldsymbol{0}$, $\boldsymbol{V}^{(h,0)}=\boldsymbol{0}$, $\boldsymbol{u}^{(h,0)}=\boldsymbol{0}$, $\boldsymbol{C}_{rr}^{c}=\boldsymbol{S}\boldsymbol{C}_{\alpha\alpha}\boldsymbol{S}^{\dagger}$, $\boldsymbol{d}=\left[\boldsymbol{E}_{c}\left(\boldsymbol{r}\right)-\boldsymbol{r}^{e}\right]$, in which case we also have the relation $\boldsymbol{D}=\boldsymbol{C}_{rr}^{e}+\boldsymbol{S}\boldsymbol{C}_{\alpha\alpha}\boldsymbol{S}^{\dagger}=\boldsymbol{C}_{rr}^{e}+\boldsymbol{C}_{rr}^{c}$. These conditions are satisfied if all of the sensitivities higher than first-order are ignored. In this particular case, the results produced by the 2nd-BERRU-PMP and 2nd-BERRU-PMD methodologies also coincide with the results produced by the first-order 1st-BERRU-PM methodology.

4.3. Inter-Comparison of Expressions for the Best-Estimate Predicted Response Covariances

The expression obtained using the 2nd-BERRU-PMD methodology [1] for the "first-iterate best-estimate" covariance matrix $\boldsymbol{C}_{rr}^{(1)bed}$ for the responses is as follows:

$$
\begin{aligned}
\boldsymbol{C}_{rr}^{(1)bed} = & \boldsymbol{C}_{rr}^{e}-\boldsymbol{C}_{rd}^{(1)}\boldsymbol{Y}_{r}^{\dagger}+\boldsymbol{X}_{r}\boldsymbol{\theta}^{(h,0)}\left(\boldsymbol{\theta}^{(h,0)}\right)^{\dagger}\boldsymbol{X}_{r}^{\dagger}-\boldsymbol{X}_{r}\boldsymbol{\theta}^{(h,0)}\overline{\boldsymbol{d}}^{\dagger}\boldsymbol{Y}_{r}^{\dagger} \\
& +\boldsymbol{X}_{r}\boldsymbol{\theta}^{(h,0)}\left(\boldsymbol{u}^{(h,0)}\right)^{\dagger}\boldsymbol{Y}_{r}^{\dagger}-\boldsymbol{Y}_{r}\left(\boldsymbol{C}_{rd}^{(1)}\right)^{\dagger}-\boldsymbol{Y}_{r}\overline{\boldsymbol{d}}\left(\boldsymbol{\theta}^{(h,0)}\right)^{\dagger}\boldsymbol{X}_{r}^{\dagger} \\
& +\boldsymbol{Y}_{r}\boldsymbol{D}\boldsymbol{Y}_{r}^{\dagger}-\boldsymbol{Y}_{r}\overline{\boldsymbol{d}}\left(\boldsymbol{u}^{(h,0)}\right)^{\dagger}\boldsymbol{Y}_{r}^{\dagger}+\boldsymbol{Y}_{r}\boldsymbol{u}^{(h,0)}\left(\boldsymbol{\theta}^{(h,0)}\right)^{\dagger}\boldsymbol{X}_{r}^{\dagger} \\
& -\boldsymbol{Y}_{r}\boldsymbol{u}^{(h,0)}\overline{\boldsymbol{d}}^{\dagger}\boldsymbol{Y}_{r}^{\dagger}+\boldsymbol{Y}_{r}\boldsymbol{u}^{(h,0)}\left(\boldsymbol{u}^{(h,0)}\right)^{\dagger}\boldsymbol{Y}_{r}^{\dagger}.
\end{aligned}
\tag{60}
$$

where:

$$
\boldsymbol{C}_{rd}^{(1)}=\boldsymbol{C}_{rr}^{e}-\boldsymbol{C}_{r\alpha}^{e}\boldsymbol{S}^{\dagger};\ \boldsymbol{X}_{r}\triangleq\boldsymbol{C}_{rd}^{(1)}\left[\left(\boldsymbol{D}-\boldsymbol{V}^{(h,0)}\right)^{-1}\boldsymbol{D}-\boldsymbol{I}\right];\ \boldsymbol{Y}_{r}\triangleq\boldsymbol{C}_{rd}^{(1)}\left(\boldsymbol{D}-\boldsymbol{V}^{(h,0)}\right)^{-1}. \tag{61}
$$

Comparing the 2nd-BERRU-PMD-expression provided in Equation (60) with the corresponding expression obtained using the 2nd-BERRU-PMP methodology, cf. Equation (31), indicates that both expressions include the second-order sensitivities of the responses with respect to the parameters, but in distinct ways. Furthermore, the expression produced by the 2nd-BERRU-PMP methodology, cf. Equation (31), can include all of the response sensitivities and parameter correlations of order higher than two, if available. Such an inclusion of higher-order sensitivities and parameter correlations is not possible with the result obtained using the 2nd-BERRU-PMD methodology, cf. Equation (60).

The expression provided in Equation (31) coincides with the expression provided in Equation (60) only if the following relationship holds:

$$C_{rr}^{bep} = C_{rr}^{(1)bed} \text{ iff : } -C_{rr}^{e}\left(C_{rr}^{e}+C_{rr}^{c}\right)^{-1}C_{rr}^{e} = -C_{rd}^{(1)}Y_{r}^{\dagger} + X_{r}\boldsymbol{\theta}^{(h,0)}\left(\boldsymbol{\theta}^{(h,0)}\right)^{\dagger}X_{r}^{\dagger}$$

$$-X_{r}\boldsymbol{\theta}^{(h,0)}\overline{\boldsymbol{d}}^{\dagger}Y_{r}^{\dagger} + X_{r}\boldsymbol{\theta}^{(h,0)}\left(\boldsymbol{u}^{(h,0)}\right)^{\dagger}Y_{r}^{\dagger} - Y_{r}\left(C_{rd}^{(1)}\right)^{\dagger} - Y_{r}\overline{\boldsymbol{d}}\left(\boldsymbol{\theta}^{(h,0)}\right)^{\dagger}X_{r}^{\dagger} + Y_{r}DY_{r}^{\dagger} \quad (62)$$

$$-Y_{r}\overline{\boldsymbol{d}}\left(\boldsymbol{u}^{(h,0)}\right)^{\dagger}Y_{r}^{\dagger} + Y_{r}\boldsymbol{u}^{(h,0)}\left(\boldsymbol{\theta}^{(h,0)}\right)^{\dagger}X_{r}^{\dagger} - Y_{r}\boldsymbol{u}^{(h,0)}\overline{\boldsymbol{d}}^{\dagger}Y_{r}^{\dagger} + Y_{r}\boldsymbol{u}^{(h,0)}\left(\boldsymbol{u}^{(h,0)}\right)^{\dagger}Y_{r}^{\dagger}.$$

The equality shown in Equation (62) can hold only if all of the second-order sensitivities are neglected, and if the experimental responses are uncorrelated to the model parameters, in which case the following simplifications occur: $X_{r} = \mathbf{0}$, $Y_{r} = C_{rd}^{(1)}D^{-1}$, and $C_{rd}^{(1)} = C_{rr}^{e}$, in which case we also have the relation $D = C_{rr}^{e} + SC_{\alpha\alpha}S^{\dagger} = C_{rr}^{e} + C_{rr}^{c}$. In this particular case, the results produced by the 2nd-BERRU-PMP and 2nd-BERRU-PMD methodologies also coincide with the results produced by the first-order 1st-BERRU-PM methodology.

4.4. Inter-Comparison of Expressions for the Best-Estimate Predicted Parameter Covariances

The expression obtained using the 2nd-BERRU-PMD methodology [1] for the "first-iterate best-estimate" covariance matrix $C_{\alpha\alpha}^{(1)be}$ for the model parameters is as follows:

$$C_{\alpha\alpha}^{(1)bed} = C_{\alpha\alpha}^{e} - C_{\alpha d}^{(1)}Y_{\alpha}^{\dagger} + X_{\alpha}\boldsymbol{\theta}^{(h,0)}\left(\boldsymbol{\theta}^{(h,0)}\right)^{\dagger}X_{\alpha}^{\dagger} - X_{\alpha}\boldsymbol{\theta}^{(h,0)}\overline{\boldsymbol{d}}^{\dagger}Y_{\alpha}^{\dagger}$$

$$+ X_{\alpha}\boldsymbol{\theta}^{(h,0)}\left(\boldsymbol{u}^{(h,0)}\right)^{\dagger}Y_{\alpha}^{\dagger} - Y_{\alpha}\left(C_{\alpha d}^{(1)}\right)^{\dagger} - Y_{\alpha}\overline{\boldsymbol{d}}\left(\boldsymbol{\theta}^{(h,0)}\right)^{\dagger}X_{\alpha}^{\dagger}$$

$$+ Y_{\alpha}DY_{\alpha}^{\dagger} - Y_{\alpha}\overline{\boldsymbol{d}}\left(\boldsymbol{u}^{(h,0)}\right)^{\dagger}Y_{\alpha}^{\dagger} + Y_{\alpha}\boldsymbol{u}^{(h,0)}\left(\boldsymbol{\theta}^{(h,0)}\right)^{\dagger}X_{\alpha}^{\dagger} \quad (63)$$

$$- Y_{\alpha}\boldsymbol{u}^{(h,0)}\overline{\boldsymbol{d}}^{\dagger}Y_{\alpha}^{\dagger} + Y_{\alpha}\boldsymbol{u}^{(h,0)}\left(\boldsymbol{u}^{(h,0)}\right)^{\dagger}Y_{\alpha}^{\dagger}.$$

were:

$$C_{\alpha d}^{(1)} = C_{\alpha r}^{e} - C_{\alpha\alpha}^{e}S^{\dagger}; \quad X_{\alpha} \triangleq C_{\alpha d}^{(1)}\left[\left(D - V^{(h,0)}\right)^{-1}D - I\right]; \quad Y_{\alpha} \triangleq C_{\alpha d}^{(1)}\left(D - V^{(h,0)}\right)^{-1}. (64)$$

Comparing the 2nd-BERRU-PMD-expression provided in Equation (63) with the corresponding expression obtained using the 2nd-BERRU-PMP methodology, cf. Equation (33), indicates that both expressions include the second-order sensitivities of the responses with respect to the parameters, but in distinct ways. Furthermore, the expression produced by the 2nd-BERRU-PMP methodology, cf. Equation (33), can include all of the response sensitivities and parameter correlations of order higher than two, if available. Such an inclusion of higher-order sensitivities and parameter correlations is not possible with the result obtained using the 2nd-BERRU-PMD methodology, cf. Equation (63).

The expression provided in Equation (33) coincides with the expression provided in Equation (63) only if the following relationship holds:

$$C_{\alpha\alpha}^{bep} = C_{\alpha\alpha}^{(1)bed} \text{ iff : } -C_{\alpha r}^{c}\left(C_{rr}^{e}+C_{rr}^{c}\right)^{-1}C_{\alpha r}^{c} = -C_{\alpha d}^{(1)}Y_{\alpha}^{\dagger} + X_{\alpha}\boldsymbol{\theta}^{(h,0)}\left(\boldsymbol{\theta}^{(h,0)}\right)^{\dagger}X_{\alpha}^{\dagger}$$

$$-X_{\alpha}\boldsymbol{\theta}^{(h,0)}\overline{\boldsymbol{d}}^{\dagger}Y_{\alpha}^{\dagger} + X_{\alpha}\boldsymbol{\theta}^{(h,0)}\left(\boldsymbol{u}^{(h,0)}\right)^{\dagger}Y_{\alpha}^{\dagger} - Y_{\alpha}\left(C_{\alpha d}^{(1)}\right)^{\dagger} - Y_{\alpha}\overline{\boldsymbol{d}}\left(\boldsymbol{\theta}^{(h,0)}\right)^{\dagger}X_{\alpha}^{\dagger} + Y_{\alpha}DY_{\alpha}^{\dagger} \quad (65)$$

$$-Y_{\alpha}\overline{\boldsymbol{d}}\left(\boldsymbol{u}^{(h,0)}\right)^{\dagger}Y_{\alpha}^{\dagger} + Y_{\alpha}\boldsymbol{u}^{(h,0)}\left(\boldsymbol{\theta}^{(h,0)}\right)^{\dagger}X_{\alpha}^{\dagger} - Y_{\alpha}\boldsymbol{u}^{(h,0)}\overline{\boldsymbol{d}}^{\dagger}Y_{\alpha}^{\dagger} + Y_{\alpha}\boldsymbol{u}^{(h,0)}\left(\boldsymbol{u}^{(h,0)}\right)^{\dagger}Y_{\alpha}^{\dagger}.$$

The equality shown in Equation (65) can hold only if all of the second-order sensitivities are neglected, and if the experimental responses are uncorrelated to the model parameters, in which case the following simplifications occur: $X_\alpha = 0$, $Y_\alpha = C_{\alpha d}^{(1)} D^{-1}$, in which case we also have the relation $D = C_{rr}^e + S C_{\alpha\alpha} S^\dagger = C_{rr}^e + C_{rr}^c$. In this particular case, the results produced by the 2nd-BERRU-PMP and 2nd-BERRU-PMD methodologies also coincide with the results produced by the first-order 1st-BERRU-PM methodology.

4.5. Inter-Comparison of Expressions for the Best-Estimate Predicted Correlations between Parameters and Responses

The expression obtained using the 2nd-BERRU-PMD methodology [1] for the "first-iterate best-estimate" correlation matrix $C_{\alpha r}^{(1)bed}$ for the parameters and responses is as follows:

$$
\begin{aligned}
C_{\alpha r}^{(1)bed} = {} & C_{\alpha r}^e - C_{\alpha d}^{(1)} Y_r^\dagger + X_\alpha \theta^{(h,0)} \left(\theta^{(h,0)}\right)^\dagger X_r^\dagger - X_\alpha \theta^{(h,0)} \overline{d}^\dagger Y_r^\dagger \\
& + X_\alpha \theta^{(h,0)} \left(u^{(h,0)}\right)^\dagger Y_r^\dagger - Y_\alpha \left(C_{rd}^{(1)}\right)^\dagger - Y_\alpha \overline{d} \left(\theta^{(h,0)}\right)^\dagger X_r^\dagger \\
& + Y_\alpha D Y_r^\dagger - Y_\alpha \overline{d} \left(u^{(h,0)}\right)^\dagger Y_r^\dagger + Y_\alpha u^{(h,0)} \left(\theta^{(h,0)}\right)^\dagger X_r^\dagger \\
& - Y_\alpha u^{(h,0)} \overline{d}^\dagger Y_r^\dagger + Y_\alpha u^{(h,0)} \left(u^{(h,0)}\right)^\dagger Y_r^\dagger .
\end{aligned}
\tag{66}
$$

Comparing the 2nd-BERRU-PMD-expression provided in Equation (66) with the corresponding expression obtained using the 2nd-BERRU-PMP methodology, cf. Equation (34), indicates that both expressions include the second-order sensitivities of the responses with respect to the parameters, but in distinct ways. Furthermore, the expression produced by the 2nd-BERRU-PMP methodology, cf. Equation (34), can include all of the response sensitivities and parameter correlations of order higher than two, if available. Such an inclusion of higher-order sensitivities and parameter correlations is not possible with the result obtained using the 2nd-BERRU-PMD methodology, cf. Equation (66)

The expression provided in Equation (34) coincides with the expression provided in Equation (66) only if the following relationship holds:

$$
\begin{aligned}
C_{\alpha r}^{bep} = C_{\alpha r}^{(1)bed} \ \text{iff}: \ & -C_{\alpha r}^c \left(C_{rr}^e + C_{rr}^c\right)^{-1} C_{rr}^e = -C_{\alpha d}^{(1)} Y_r^\dagger + X_\alpha \theta^{(h,0)} \left(\theta^{(h,0)}\right)^\dagger X_r^\dagger \\
& - X_\alpha \theta^{(h,0)} \overline{d}^\dagger Y_r^\dagger + X_\alpha \theta^{(h,0)} \left(u^{(h,0)}\right)^\dagger Y_r^\dagger - Y_\alpha \left(C_{rd}^{(1)}\right)^\dagger - Y_\alpha \overline{d} \left(\theta^{(h,0)}\right)^\dagger X_r^\dagger + Y_\alpha D Y_r^\dagger \\
& - Y_\alpha \overline{d} \left(u^{(h,0)}\right)^\dagger Y_r^\dagger + Y_\alpha u^{(h,0)} \left(\theta^{(h,0)}\right)^\dagger X_r^\dagger - Y_\alpha u^{(h,0)} \overline{d}^\dagger Y_r^\dagger + Y_\alpha u^{(h,0)} \left(u^{(h,0)}\right)^\dagger Y_r^\dagger .
\end{aligned}
\tag{67}
$$

The equality shown in Equation (67) can hold only if all of the second-order sensitivities are neglected, and if the experimental responses are uncorrelated to the model parameters. In this particular case, the results produced by the 2nd-BERRU-PMP and 2nd-BERRU-PMD methodologies also coincide with the results produced by the first-order 1st-BERRU-PM methodology.

5. Fundamental Advantages of 2nd-BERRU-PMP over Second-Order Data Assimilation

This section summarizes the decisive advantages of the 2nd-BERRU-PM metho-

dology over the 2nd-Order Data Assimilation methodology [3] [4]. The 2nd-Order Data Assimilation methodology [3] [4] relies on using a 2nd-order procedure to minimize a user-defined functional which is meant to represent, in a chosen norm (usually the energy-norm), the discrepancies between computed and experimental results ("responses"). Mathematically, the 2nd-Order Data Assimilation method involves the following steps [3] [4]:

1) Consider that the physical system is represented, as in [1], by the operator equations $N^m[v(x),\alpha] = Q^m(x,\alpha)$. Consider that the vector of measured responses ("observations"), denoted by the vector $z \triangleq (z_1,\cdots,z_{TR})^\dagger$, is a known function of the vector of state-variables $v(x) \triangleq [v_1(x),\cdots,v_{TD}(x)]^\dagger$ and of the vector of errors $w \triangleq (w_1,\cdots,w_{TR})^\dagger$, having the following expression: $z = h(v) + w$, where $h(v) \triangleq [h_1(v),\cdots,h_{TR}(v)]^\dagger$ denotes a known vector-function of v. The error term, w, is considered here to include "representative errors" stemming from sampling and grid interpolation; the mean value of w corresponds to $r_e \triangleq (r_1^e,\cdots,r_{TR}^e)^\dagger$ and the covariance matrixof w corresponds to C_{rr}^e. As described in [3] [4], w is often considered to have the characteristics of "white noise," in which case $z \sim N[h(v),C_{rr}^e]$ is a normal distribution with mean $h(v)$ and covariance C_e^{rr}. In addition, it is assumed that the prior "background" information is also known, being represented by a multivariate normal distribution with a known mean, denoted as v_b, and a known covariance matrix denoted as B, i.e., $v \sim N[v_b,B]$. The posterior distribution, $p(z|v)$, is obtained by applying Bayes' Theorem to the above information, which yields the result $p(z|v) \sim \exp[-J(v)]$, where:

$$J(v) \triangleq \frac{1}{2}\left\{[z - h(v)]^\dagger (C_{rr}^e)^{-1}[z - h(v)] + (v - v_b)^\dagger B^{-1}(v - v_b)\right\}. \tag{68}$$

The maximum posterior estimate is obtained by determining the minimum of the functional $J(v)$, which occurs at the root (s) of the following equation:

$$0 = \nabla J(v) = B^{-1}(v - v_b) - [D_h(v)]^\dagger (C_{rr}^e)^{-1}[z - h(v)]. \tag{69}$$

where $D_h(v)$ denotes the Jacobian matrix of $h(v)$ with respect to the components of v.

The "first-order data assimilation" procedure solves Equation (69) by using a "partial quadratic approximation" to $J(v)$, while the "second-order data assimilation" procedure solves Equation (69) by using a "full quadratic approximation" to $J(v)$, as detailed in [3] [4], to obtain the "optimal data assimilation solution," which is here denoted as v_{DA}, as the solution of Equation (69). The following fundamental differences become apparent by comparing the "Data Assimilation" result represented by Equation (69) and the 2nd-BERRU-PMP results.

2) Data assimilation (DA) is formulated conceptually [3] [4] either just in the phase-space of measured responses ("observation-space formulation") or just in the phase-space of the model's dependent variables ("state-space formulation"). Hence, DA can calibrate initial conditions as "direct results" but cannot directly

calibrate any other model parameters. In contradistinction, the 2nd-BERRU-PMP methodology is formulated conceptually in the most inclusive "joint-phase-space of parameters, computed and measured responses." Consequently, the 2nd-BERRU-PMP methodology simultaneously calibrates responses and parameters, thus simultaneously providing results for forward and inverse problems.

3) If experiments are perfectly well known, *i.e.*, if $C_{rr}^e = \mathbf{0}$, Equation (69) indicates that the DA methodology fails fundamentally. In contradistinction, Equations (33)-(35) indicate that the 2nd-BERRU-PMP methodology does not fail when $C_{rr}^e = \mathbf{0}$ because, in any situation, $C_{rr}^c \neq \mathbf{0}$.

4) The DA methodology also fails fundamentally when the response measurements happen to coincide with the computed value of the response, *i.e.*, when $z = \boldsymbol{h}(\boldsymbol{v})$ at some point in the state-space. In such a case, the DA's Equation (69) yields the trivial result $\boldsymbol{v}_{DA} = \boldsymbol{v}_b$. In contradistinction, the 2nd-BERRU-PMP methodology does not yield such a trivial result when the response measurements happen to coincide with the computed value of the response, *i.e.*, when $\boldsymbol{r}_e = \boldsymbol{r}_k(\boldsymbol{\alpha}^0)$, because the difference $\boldsymbol{E}_c(\boldsymbol{r}) - \boldsymbol{r}_e$, which appears on the rights sides of Equations (29) and (30), remains non-zero due to the contributions of the second- and higher-order sensitivities of the responses with respect to the model parameters, as shown in Equation (4), *i.e.*,:

$$\left\{ E_c(r_k) - r_k^e \right\}_{r_k^e = r_k(\boldsymbol{\alpha}^0)} = \sum_{i=1}^{TP} \sum_{j=1}^{TP} \left\{ \partial^2 r_k(\boldsymbol{\alpha}) / \partial \alpha_i \partial \alpha_j \right\}_{\boldsymbol{\alpha}^0} \mathrm{cov}(\alpha_i, \alpha_j) / 2 + \cdots \neq 0,$$

for $k = 1, \cdots, TR$. This situation clearly underscores the need for computing and retaining (at least) the second-order response sensitivities to the model parameters. Although a situation when $\boldsymbol{r}_e = \boldsymbol{r}_k(\boldsymbol{\alpha}^0)$ is not expected to occur frequently in practice, there are no negative consequences (should such a situation occur) if the 2nd-BERRU-PMP methodology is used, in contradistinction to using the DA methodology.

5) The 2nd-BERRU-PMP methodology only requires the inversion of the matrix $\left(C_{rr}^e + C_{rr}^c \right)$ of size $TR \times TR$. In contradistinction, the solution of the "1st-order DA" requires the inversion of the Jacobian $\boldsymbol{D}_h(\boldsymbol{v})$ of $\boldsymbol{h}(\boldsymbol{v})$, while the solution of the "2nd-order DA" also requires the inversion of a matrix-vector product involving the Hessian matrix of $\boldsymbol{h}(\boldsymbol{v})$; these matrices are significantly larger [3] [4] than the matrix $\left(C_{rr}^e + C_{rr}^c \right)$. Hence, the 2nd-BERRU-PM methodology is significantly more efficient computationally than DAE.

6) The DA methodology is practically non-extendable beyond "second-order." A "3rd-order DA" would be computationally impractical because of the massive sizes of the matrices that would need to be inverted. In contradistinction, the 2nd-BERRU-PM methodology presented herein already comprises 4th-order sensitivities of responses to parameters and can be readily extended/generalized to include even higher-order sensitivities and parameter correlations.

All of the above advantages of the 2nd-BERRU-PM methodology over the DA methodology stem from the fact that the 2nd-BERRU-PM methodology is fundamentally anchored in physics-based principles (thermodynamics & informa-

tion theory) formulated in the most inclusive possible phase-space (namely the combined phase-space of computed and measured parameters and responses), whereas the DA methodology is fundamentally based on the minimization of a subjective user-chosen functional.

6. Discussion and Conclusions

This work has presented the "second order predictive modeling methodology with probabilistically included computed responses" (2nd-BERRU-PMP) methodology, as underscored by the letter "P" in the acronym (which indicates the "probabilistic" inclusion of the computed model responses). This methodology is a companion to the alternative 2nd-BERRU-PMD methodology, in which the computational model is included deterministically, as underscored by the letter "D" in this acronym. Both methodologies are fundamentally based on the Max-Ent principle. Although both the 2nd-BERRU-PMP and the 2nd-BERRU-PMD methodologies yield expressions that include second (and higher) order sensitivities of responses to model parameters, it is shown in this work that the respective expressions for the predicted responses, calibrated predicted parameters and their predicted uncertainties (covariances) are not identical to each other. Nevertheless, these second-order methodologies encompass, as particular cases, the results produced by the extant data assimilation [3] [4], data adjustment procedures [5] [6] [7] [8] [9], and the first-order BERRU-PM methodology [10] [11] [12].

Notably, the 2nd-BERRU-PMP methodology enables the use of sensitivities and parameter correlations beyond second-order, if available; this opportunity is not available within the 2nd-BERRU-PMD methodology. If it is imperative to combine experimental information for both the responses and parameters with the computational information produced by the model, then the complete Gaussian expression of the 2nd-BERRU-PMP should be used, if it is possible to accommodate the massive computational requirements underlying the inversion of the matrices that arise within this methodology. On the other hand, the simplest to implement computationally is the simplified 2nd-BERRU-PMP, which should be used when experimental information about the model parameters in addition to the information used in the computational model is not important. The 2nd-BERRU-PMD methodology offers an intermediary alternative, being easier to implement computationally than the complete version of the 2nd-BERRU-PMP methodology but involving more computations than the simplified 2nd-BERRU-PMP methodology.

Illustrative applications of the 2nd-BERRU-PMP and the 2nd-BERRU-PMD methodologies to various paradigm forward and inverse problems are currently in progress.

Conflicts of Interest

The author declares no conflict of interest regarding the publication of this paper.

References

[1]   Cacuci, D.G. (2023) Second-Order MaxEnt Predictive Modeling Methodology. I: Deterministically Incorporated Computational Model Responses (2nd-BERRU-PMD). *American Journal of Computational Mathematics.*

[2]   Jaynes, E.T. (1957) Information Theory and Statistical Mechanics. *Physical Review Journals Archive*, **106**, 620-630. https://doi.org/10.1103/PhysRev.106.620

[3]   Lewis, J.M., Lakshmivarahan, S. and Dhall, S.K. (2006) Dynamic Data Assimilation: A Least Square Approach. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511526480

[4]   Cacuci, D.G., Navon, M.I. and Ionescu-Bujor, M. (2014) Computational Methods for Data Evaluation and Assimilation. Chapman & Hall/CRC, Boca Raton.

[5]   Rowlands, J., *et al.* (1973) The Production and Performance of the Adjusted Cross-Section Set FGL5. *Proceeding of International Symposium on Physics of Fast Reactors*, Tokio. http://www.cern.ch/root

[6]   Gandini, A. and Petilli, M. (1973) A Code Using the Lagrange's Multipliers Method for Nuclear Data Adjustment. CNEN-RT/FI(73)39, Comitato Nazionale Energia Nucleare, Rome.

[7]   Kuroi, H. and Mitani, H. (1975) Adjustment to Cross-Section Data to Fit Integral Experiments by Least Squares Method. *Journal of Nuclear Science and Technology*, **12**, 663-680. https://doi.org/10.1080/18811248.1975.9733172

[8]   Dragt, J.B., Dekker, J.W.M., Gruppelaar, H. and Janssen, A.J. (1977) Methods of Adjustment and Error Evaluation of Neutron Capture Cross Sections. *Nuclear Science and Engineering*, **62**, 117-129. https://doi.org/10.13182/NSE77-3

[9]   Barhen, J., Cacuci, D.G., Wagschal, J.J., Bjerke, M.A. and Mullins, C.B. (1982) Uncertainty Analysis of Time-Dependent Nonlinear Systems: Theory and Application to Transient Thermal Hydraulics. *Nuclear Science and Engineering*, **81**, 23-44. https://doi.org/10.13182/NSE82-3

[10] Cacuci, D.G. (2014) Predictive Modeling of Coupled Multi-Physics Systems: I. Theory. *Annals of Nuclear Energy*, **70**, 266-278. https://doi.org/10.1016/j.anucene.2013.11.027.

[11] Cacuci, D.G. (2019) BERRU Predictive Modeling: Best-Estimate Results with Reduced Uncertainties. Springer, Berlin. https://doi.org/10.1007/978-3-662-58395-1

[12] Cacuci, D.G. (2023) Perspective on Predictive Modeling: Current Status, New High-Order Methodology and Outlook for Energy Systems. *Energies*, **16**, Article 933. https://doi.org/10.3390/en16020933

[13] Bellman, R.E. (1957) Dynamic Programming. Rand Corporation, Princeton University Press, Princeton.

[14] Cacuci, D.G. (2022) The nth-Order Comprehensive Adjoint Sensitivity Analysis Methodology for Nonlinear Systems (n^{th}-CASAM-N): Mathematical Framework. *Journal of Nuclear Engineering*, **3**, 163-190. https://doi.org/10.3390/jne3030010

[15] Cacuci, D.G. (2022) The nth-Order Comprehensive Adjoint Sensitivity Analysis Methodology: Overcoming the Curse of Dimensionality. Volume I: Linear Systems. Springer Nature Switzerland, Cham. https://doi.org/10.1007/978-3-030-96364-4

[16] Tukey, J.W. (1957) The Propagation of Errors, Fluctuations and Tolerances. Technical Reports No. 10-12, Princeton University, Princeton.

[17] Shun, Z. and McCullagh. P. (1995) Laplace Approximation of High Dimensional Integrals. *Journal of the Royal Statistical Society: Series B (Methodological)*, **57**, 749-760. https://doi.org/10.1111/j.2517-6161.1995.tb02060.x

[18] Tang, Y. and Reid, N. (2021) Laplace and Saddle Point Approximations in High Dimensions. arXiv: 2107.1088.

[19] Shannon, C.E. (1948) A Mathematical Theory of Communication. *The Bell System Technical Journal*, **27**, 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Appendix: Construction of the Second-Order Maximum Entropy Distribution for Responses and Parameters

When an unknown distribution $p(\boldsymbol{\alpha}, \boldsymbol{r})$, defined on a domain $D \triangleq D_\alpha \cup D_r$ (where D_α denotes the domain of definition of the parameters and D_r denotes the domain of definition of the responses) needs to be reconstructed from a finite number of its known moments, the principle of maximum entropy (MaxEnt) originally formulated by Jaynes [2] provides the optimal compatibility with the available information, while simultaneously ensuring minimal spurious information content. In particular, when only the first-order and second-order moments of the joint distribution of model parameters and responses $p(\boldsymbol{\alpha}, \boldsymbol{r})$ are known, the corresponding second-order MaxEnt distribution $p(\boldsymbol{\alpha}, \boldsymbol{r})$ is constructed by following the procedure outlined below.

1) Known means and covariances for responses:

$$r_i^0 \triangleq \int_D r_i p(\boldsymbol{\alpha}, \boldsymbol{r}) \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{\alpha}, \quad i = 1, \cdots, TR; \tag{70}$$

$$\mathrm{cov}(r_i, r_j) \triangleq \int_{D_e} (r_i - r_i^0)(r_j - r_j^0) p(\boldsymbol{\alpha}, \boldsymbol{r}) \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{\alpha}; \, i, j = 1, \cdots, TR. \tag{71}$$

2) Known means and covariances for parameters:

$$\alpha_j^0 \triangleq \int_{D_e} \alpha_j p(\boldsymbol{\alpha}, \boldsymbol{r}) \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{\alpha}, \quad j = 1, \cdots, TP; \tag{72}$$

$$\mathrm{cov}(\alpha_i, \alpha_j) \triangleq \int_{D_e} (\alpha_i - \alpha_i^0)(\alpha_j - \alpha_j^0) p(\boldsymbol{\alpha}, \boldsymbol{r}) \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{\alpha}; \, i, j = 1, \cdots, TP; \tag{73}$$

$$\mathrm{cov}(\alpha_i, r_j) \triangleq \int_{D_e} (\alpha_i - \alpha_i^0)(r_j - r_j^0) p(\boldsymbol{\alpha}, \boldsymbol{r}) \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{\alpha}; \, i = 1, \cdots, TP; \, j = 1, \cdots, TR. \tag{74}$$

According to the MaxEnt principle, the probability density $p(\boldsymbol{\alpha}, \boldsymbol{r})$ would satisfy the "available information" provided in Equations (70)-(74), without implying any spurious information or hidden assumptions, if:

1) $p(\boldsymbol{\alpha}, \boldsymbol{r})$ maximizes the Shannon [19] information entropy, S, as defined below:

$$S = -\int_D p(\boldsymbol{\alpha}, \boldsymbol{r}) \ln[p(\boldsymbol{\alpha}, \boldsymbol{r})] \mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{r}, \tag{75}$$

2) $p(\boldsymbol{\alpha}, \boldsymbol{r})$ satisfies the "moments constraints" defined by Equations (70)-(74);

3) $p(\boldsymbol{\alpha}, \boldsymbol{r})$ satisfies the normalization condition:

$$\int_D p(\boldsymbol{\alpha}, \boldsymbol{r}) \mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{r} = 1. \tag{76}$$

The MaxEnt distribution $p(\boldsymbol{\alpha}, \boldsymbol{r})$ is obtained as the solution of the constrained variational problem $\partial H(p)/\partial p = 0$, where the entropy (Lagrangian functional) $H(p)$ is defined as follows:

$$H(p) = -\int_D p(\boldsymbol{\alpha}, \boldsymbol{r}) \ln[p(\boldsymbol{\alpha}, \boldsymbol{r})] \mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{r} - \lambda_0 \left[\int_D p(\boldsymbol{\alpha}, \boldsymbol{r}) \mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{r} - 1 \right]$$

$$- \sum_{k=1}^{TR} \lambda_k^{(1)} \left[\int_D r_k p(\boldsymbol{\alpha}, \boldsymbol{r}) \mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{r} - r_k^0 \right] - \sum_{i=1}^{TP} \lambda_i^{(2)} \left[\int_D \alpha_i p(\boldsymbol{\alpha}, \boldsymbol{r}) \mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{r} - \alpha_i^0 \right]$$

$$-\frac{1}{2}\sum_{k=1}^{TR}\sum_{\ell=1}^{TR}\lambda_{k\ell}^{(11)}\left[\int_D r_k r_\ell p(\boldsymbol{\alpha},\boldsymbol{r})\,\mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{r}-cov(r_k,r_\ell)-r_k^0 r_\ell^0\right]$$

$$-\sum_{k=1}^{TR}\sum_{i=1}^{TP}\lambda_{ki}^{(12)}\left[\int_D r_k \alpha_i p(\boldsymbol{\alpha},\boldsymbol{r})\,\mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{r}-cov(r_k,\alpha_i)-r_k^0 \alpha_i^0\right] \qquad (77)$$

$$-\frac{1}{2}\sum_{i=1}^{TP}\sum_{j=1}^{TP}\lambda_{ij}^{(22)}\left[\int_D \alpha_i \alpha_j p(\boldsymbol{\alpha},\boldsymbol{r})\,\mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{r}-\mathrm{cov}(\alpha_i,\alpha_j)-\alpha_i^0 \alpha_j^0\right].$$

In Equation (77), the quantities $\lambda_k^{(1)}$, $\lambda_k^{(2)}$, $\lambda_{k\ell}^{(11)}$, $\lambda_{k\ell}^{(12)}$, and $\lambda_{k\ell}^{(22)}$ denote the respective Lagrange multipliers, and the factors 1/2 have been introduced for subsequent computational convenience.

Solving the equation $\partial H(p)/\partial p = 0$ yields the following expression for the resulting MaxEnt distribution $p(z)$:

$$p_c(z)=\frac{1}{Z(\boldsymbol{b},\boldsymbol{\Lambda})}\exp\left(-\boldsymbol{b}^\dagger z-\frac{1}{2}z^\dagger \boldsymbol{\Lambda} z\right), \qquad (78)$$

where the various vectors and matrices are defined as follows:

$$z=\begin{pmatrix}\boldsymbol{r}\\\boldsymbol{\alpha}\end{pmatrix};\quad \boldsymbol{b}\triangleq\begin{pmatrix}\boldsymbol{\lambda}^{(1)}\\\boldsymbol{\lambda}^{(2)}\end{pmatrix};\quad \boldsymbol{\lambda}^{(1)}\triangleq\begin{pmatrix}\lambda_1^{(1)}\\\bullet\\\lambda_{TR}^{(1)}\end{pmatrix};\quad \boldsymbol{\lambda}^{(2)}\triangleq\begin{pmatrix}\lambda_1^{(2)}\\\bullet\\\lambda_{TP}^{(1)}\end{pmatrix}; \qquad (79)$$

$$\boldsymbol{\Lambda}\triangleq\begin{pmatrix}\boldsymbol{\Lambda}^{(11)} & \boldsymbol{\Lambda}^{(12)}\\ \left[\boldsymbol{\Lambda}^{(12)}\right]^\dagger & \boldsymbol{\Lambda}^{(22)}\end{pmatrix};\quad \boldsymbol{\Lambda}^{(11)}\triangleq\begin{pmatrix}\lambda_{11}^{(11)} & \bullet & \lambda_{1,TR}^{(11)}\\ \bullet & \lambda_{k\ell}^{(11)} & \bullet\\ \lambda_{TR,1}^{(11)} & \bullet & \lambda_{TR,TR}^{(11)}\end{pmatrix};$$

$$\boldsymbol{\Lambda}^{(12)}\triangleq\begin{pmatrix}\lambda_{11}^{(12)} & \bullet & \lambda_{1,TP}^{(12)}\\ \bullet & \lambda_{k\ell}^{(12)} & \bullet\\ \lambda_{TR,1}^{(12)} & \bullet & \lambda_{TR,TP}^{(12)}\end{pmatrix};\quad \boldsymbol{\Lambda}^{(22)}\triangleq\begin{pmatrix}\lambda_{11}^{(22)} & \bullet & \lambda_{1,TP}^{(22)}\\ \bullet & \lambda_{k\ell}^{(22)} & \bullet\\ \lambda_{TP,1}^{(22)} & \bullet & \lambda_{TP,TP}^{(22)}\end{pmatrix}. \qquad (80)$$

The normalization constant $Z(\boldsymbol{b},\boldsymbol{\Lambda})$ in Equation (104) is defined as follows:

$$Z(\boldsymbol{b},\boldsymbol{\Lambda})=\int_D \exp\left(-\boldsymbol{b}^\dagger z-\frac{1}{2}z^\dagger \boldsymbol{\Lambda} z\right)\mathrm{d}z;\quad \mathrm{d}z\equiv\mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{r}. \qquad (81)$$

In statistical mechanics, the normalization constant Z is called the partition function (or sum over states) and carries all of the information available about the possible states of the system, while the MaxEnt distribution $p(z)$ is called the canonical Boltzmann-Gibbs distribution. The integral in Equation (107) can be evaluated explicitly by conservatively extending the computational domain D to the entire multidimensional real vector space $\mathbb{R}^N$, where $N\triangleq TR+TP$, to obtain the following expression:

$$Z_c(\boldsymbol{b},\boldsymbol{\Lambda}_c)=\int_{\mathbb{R}^N}\exp\left(-\boldsymbol{b}^\dagger z-\frac{1}{2}z^\dagger \boldsymbol{\Lambda} z\right)\mathrm{d}z=\frac{(2\pi)^{N/2}}{\sqrt{Det(\boldsymbol{\Lambda})}}\mathrm{e}^{\left(\frac{1}{2}\boldsymbol{b}^\dagger \boldsymbol{\Lambda}^{-1}\boldsymbol{b}\right)}. \qquad (82)$$

The Lagrange multipliers are determined in terms of the known information (means and covariances of parameters and responses) by differentiating the "free energy" $F(\boldsymbol{b},\boldsymbol{C})\triangleq-\ln Z(\boldsymbol{b},\boldsymbol{C})$ with respect to the components of the vector

$b \triangleq \left(\lambda^{(1)}, \lambda^{(2)} \right)^{\dagger}$ to obtain the following expressions:

$$\frac{\partial F(b,C)}{\partial \lambda_k^{(1)}} = \frac{1}{Z} \int_D r_k e^{\left(-b^{\dagger}z - \frac{1}{2}z^{\dagger}\Lambda z \right)} dz = \int_{D_c} r_k p(\alpha, r) dz = r_k^0; \quad k = 1, \cdots, TR; \quad (83)$$

$$\frac{\partial F(b,C)}{\partial \lambda_i^{(2)}} = \frac{1}{Z} \int_D \alpha_i e^{\left(-b^{\dagger}z - \frac{1}{2}z^{\dagger}\Lambda z \right)} dz = \int_D \alpha_i p(\alpha, r) d\alpha dr = \alpha_i^0; \quad i = 1, \cdots, TP. \quad (84)$$

The results obtained in Equations (83) and (84) can be collectively written in vector-matrix form as follows:

$$\frac{\partial F(b,C)}{\partial b} = z^0; \quad z^0 \triangleq \left(r^0, \alpha^0 \right)^{\dagger}; \quad r^0 \triangleq \left(r_1^0, \cdots, r_{TR}^0 \right). \quad (85)$$

On the other hand, it follows from Equation (82) that:

$$\frac{\partial F(b,C)}{\partial b} = -\frac{\partial \left[\ln Z(b,C) \right]}{\partial b} = -\Lambda^{-1} b. \quad (86)$$

The relations obtained in Equations (85) and (86) imply the following relation:

$$b = -\Lambda z^0. \quad (87)$$

Differentiating a second time the relation provided in Equation (83) or (84) yields the following relations:

$$\frac{\partial^2 F(b,C)}{\partial \lambda_j^{(1)} \partial \lambda_k^{(1)}} = -\frac{1}{Z^2} \frac{\partial Z}{\partial \lambda_j^{(1)}} r_k^0 + \frac{1}{Z} \int_D (-r_j) r_k e^{\left(-b^{\dagger}z - \frac{1}{2}z^{\dagger}\Lambda z \right)} dz$$
$$= -\text{cov}(r_j, r_k); \quad j, k = 1, \cdots, TR; \quad (88)$$

$$\frac{\partial^2 F(b,C)}{\partial \lambda_i^{(2)} \partial \lambda_k^{(1)}} = -\text{cov}(\alpha_i, r_k); \quad i = 1, \cdots, TP; \quad k = 1, \cdots, TR; \quad (89)$$

$$\frac{\partial^2 F(b,C)}{\partial \lambda_i^{(2)} \partial \lambda_j^{(2)}} = -\text{cov}(\alpha_i, \alpha_j); \quad i, j = 1, \cdots, TP. \quad (90)$$

The results obtained in Equations (88)-(90) can be collectively written in vector-matrix form as follows:

$$\frac{\partial^2 F(b,C_c)}{\partial b \partial b} = -C; \quad C \triangleq \begin{pmatrix} C^{rr} & C^{r\alpha} \\ C^{\alpha r} & C^{\alpha\alpha} \end{pmatrix}; \quad C^{rr} \triangleq \left[\text{cov}(r_j, r_k) \right]_{TR \times TR};$$
$$C^{r\alpha} \triangleq \left[\text{cov}(r_k, \alpha_i) \right]_{TR \times TP} = \left(C^{\alpha r} \right)^{\dagger}; \quad C^{\alpha\alpha} \triangleq \left[\text{cov}(\alpha_i, \alpha_j) \right]_{TP \times TP}. \quad (91)$$

On the other hand, it follows from Equation (86) that:

$$\frac{\partial^2 F(b,C)}{\partial b \partial b} = -\Lambda^{-1}. \quad (92)$$

The relations obtained in Equations (91) and (92) imply the following relation:

$$\Lambda^{-1} = C. \quad (93)$$

Introducing the results obtained in Equations (87) and (93) into Equation (78)

and (82) yields the following expression for the MaxEnt distribution $p(z|z,C)$:

$$p(z|z,C) = \frac{(2\pi)^{-N/2}}{\sqrt{Det(C)}} \exp\left[-\frac{1}{2}(z-z^0)^\dagger C^{-1}(z-z^0)\right].$$ (94)