

Second-Order MaxEnt Predictive Modelling Methodology. I: Deterministically Incorporated Computational Model (2nd-BERRU-PMD)

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

How to cite this paper: Cacuci, D.G. (2023) Second-Order MaxEnt Predictive Modelling Methodology. I: Deterministically Incorporated Computational Model (2nd-BERRU-PMD). *American Journal of Computational Mathematics*, 13, 236-266. <https://doi.org/10.4236/ajcm.2023.132013>

Received: April 18, 2023

Accepted: June 17, 2023

Published: June 20, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This work presents a comprehensive second-order predictive modeling (PM) methodology 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. The 2nd-BERRU-PMD methodology is fundamentally based on the maximum entropy (MaxEnt) principle. This principle is in contradistinction to the fundamental principle that underlies the extant data assimilation and/or adjustment procedures which minimize in a least-square sense a subjective user-defined functional which is meant to represent the discrepancies between measured and computed model responses. It is shown that the 2nd-BERRU-PMD methodology generalizes and extends current data assimilation and/or data adjustment procedures while overcoming the fundamental limitations of these procedures. In the accompanying work (Part II), the alternative framework for developing the “second-order MaxEnt predictive modelling methodology” is presented by incorporating probabilistically (as opposed to “deterministically”) the computed model responses.

Keywords

Second-Order Predictive Modeling, Data Assimilation, Data Adjustment, Uncertainty Quantification, Reduced Predicted Uncertainties

1. Introduction

The results of measurements and computations are never perfectly accurate. On the one hand, results of measurements inevitably reflect the influence of experimental errors, imperfect instruments, or imperfectly known calibration standards. Around any reported experimental value, therefore, there always exists a range of values that may also be plausibly representative of the true but unknown value of the measured quantity. On the other hand, computations are afflicted by errors stemming from numerical procedures, uncertain model parameters, boundary and initial conditions, and/or imperfectly known physical processes or problem geometry. Therefore, knowing just the nominal values experimentally measured or computed quantities is insufficient for applications. The quantitative uncertainties accompanying measurements and computations are also needed, along with the respective nominal values. The discrepancies between experimental and computational results provide the basic motivation for performing quantitative model verification (meaning: “is the mathematical model solved correctly?”) and model validation (meaning: “does the model represent reality?”), which are essential components of “predictive modeling.”

Predictive modeling commences by identifying and characterizing the uncertainties involved in every step in the sequence of the numerical simulation processes that ultimately lead to a prediction. Predictive modeling comprises three key elements, namely model calibration, model extrapolation, and estimation of the validation domain. Model calibration addresses the combination of experimental and computational data and their uncertainties for the purpose of obtaining “best estimate” values for model parameters (to be used for updating the model’s parameters) and predicted model results, along with “best estimate” uncertainties (covariance/correlation matrices) for these “best-estimate” predicted parameters and results requires. Such a combination of computational and experimental information requires reasoning from incomplete, error-afflicted, and occasionally discrepant information, which includes: 1) errors and uncertainties in the data used in the simulation (e.g., input data, model parameters, initial conditions, boundary conditions, sources and forcing functions); 2) numerical discretization errors; and 3) uncertainties in (e.g., lack of knowledge of) the processes being modeled.

Under ideal circumstances, the result of predictive modeling is a probabilistic description of possible future outcomes based on all recognized errors and uncertainties. This probabilistic description enables the subsequent activity of “model extrapolation,” which aims at quantifying the uncertainties in predictions under new environments or conditions, including both untested regions of the parameter space and higher levels of system complexity in the validation hierarchy. The quantification of the validation domain underlying the models of interest requires estimation of contours of constant uncertainty in the high-dimensional space that characterizes the application of interest, including the identification of areas where the predictive estimation of uncertainty meets spe-

cified requirements for the performance, reliability, or safety of the system of interest.

The earliest activities aimed at extracting best-estimate values for model parameters by combining computational and experimental information using variational methods were initiated in the 1960s in the atmospheric and oceanographic sciences [1] [2] [3] and, in parallel, in the nuclear energy field [4]. In the earth, atmospheric and oceanographic sciences these activities were carried out under the name of “data assimilation” (DA) as described in well-known books [5] [6] [7] [8], while in the nuclear sciences these activities reached conceptual maturity under the name of “nuclear data or cross section adjustment” [9] [10] [11] [12]. The “data adjustment” and “data assimilation” methodologies can thus be considered to have been the earliest systematic methodologies that embody the principles of “predictive modeling.”

The fundamental criterion used in data adjustment and data assimilation methods is the least squares criterion, which is employed in a variety of deterministic (variational) and/or statistical forms (including Bayesian minimization, maximum likelihood, and minimum variance methods). The fundamental tenet of data adjustment and data assimilation is the minimization of a user-defined “cost functional” which describes the squared departures between computational results and the observations/experiments of the respective results. In contradistinction to the least-squares tenet underlying data adjustment and/or assimilation, the “BERRU-PM” methodology developed by Cacuci [13] [14] employs the “maximum entropy” (MaxEnt) principle [15] to combine computational and experimental information for obtaining best-estimate predicted mean values for model responses and parameters, together with reduced predicted uncertainties for these best-estimate values, thereby eliminating the need for minimizing the user-chosen “quadratic cost functional representing the weighted errors between measured and computed responses.” BERRU-PM is an acronym for “Best-Estimate Results with Reduced Uncertainties -Predictive Modeling” and is a MaxEnt methodology that incorporates first-order sensitivities of model responses with respect to the model parameters. Such sensitivities are most efficiently computed by using the adjoint sensitivity analysis method for nonlinear systems originally developed by Cacuci [16] [17]. The first-order BERRU-PM methodology was recently extended by Cacuci [18] to include second-order sensitivities. The main differences between DA [5] [6] [7] [8] results and the BERRU-PM [16] [17] [18] results are as follows:

- 1) DA [5] [6] [7] [8] is formulated conceptually 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 BERRU-PM methodology is formulated conceptually in the most inclusive “joint-phase-space of parameters, computed and measured responses.” Consequently, the BERRU-PM methodology simultaneously calibrates responses and parameters, thus simulta-

neously providing results for forward and inverse problems.

2) If experiments are perfectly well known, the DA methodology fails fundamentally, but the BERRU-PM methodology does not. The DA methodology also fails fundamentally when the response measurements happen to coincide with the computed value of the response, in which DA yields the trivial result. In contradistinction, the BERRU-PM methodology yields a non-trivial result.

3) The BERRU-PM methodology is significantly more efficient computationally than DA since it only requires the inversion a matrix of size having the dimensions of the number of model responses, whereas DA requires inversion of much larger matrices in the phase-space of dependent variables.

In this work, the methodology presented in [18] will be extended by including the full second-order representation of the computed responses, thereby creating the 2nd-BERRU-PMD methodology, where 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, and where the acronym BERRU stands, as before, for “Best-Estimate Results with Reduced Uncertainties;” the last letter (“D”) in the acronym indicates “deterministic,” referring to the deterministic inclusion of the computational model responses. The 2nd-BERRU-PMD methodology is fundamentally based on the MaxEnt principle.

This work is structured as follows: Section 2 presents the mathematical modeling of the physical system under consideration. Section 3 presents the development of the 2nd-BERRU-PMD methodology. It is shown that the results produced by the 2nd-BERRU-PMD methodology include and extend the results produced by the first-order BERRU-PM methodology and the extant data assimilation and/or data adjustment procedures while overcoming the fundamental limitations of these procedures. The concluding discussion presented in Section 4 lays the ground for the subsequent presentation [19] of the alternative framework for developing the “second-order MaxEnt predictive modelling methodology” which probabilistically (as opposed to “deterministically”) incorporates the computed model responses. This alternative methodology [19] will be designated by the acronym 2nd-BERRU-PMP, where the last letter (“P”) in the acronym indicates “probabilistic” inclusion of the computational model responses.

2. Mathematical Modeling of the Physical System

In general terms, the modeling of a physical system and/or the result of an indirect experimental measurement requires consideration of the following modeling 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;
- 2) Inequality and/or equality constraints that delimit the ranges of the system’s parameters;

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

4) Experimentally measured responses, with their respective nominal (mean) values and uncertainties (variances, covariances, skewness, kurtosis, etc.).

A nonlinear physical system can be generally modeled by means of coupled equations which can be represented in operator form as follows:

$$N^m[v(x), \alpha] = Q^m(x, \alpha), \quad x \in \Omega_x(\alpha). \quad (1)$$

The superscript “m” in Equation (1) indicates “model.” Boundary and/or initial conditions must also be provided if differential operators appear in Equation (1). In operator form, these boundaries and/or initial conditions are represented as follows:

$$B^m[v(x); \alpha; x] - B^s(x, \alpha) = \mathbf{0}, \quad x \in \partial\Omega_x(\alpha), \quad (2)$$

where the vector $B^s(x, \alpha)$ symbolically indicates “source terms on the boundary.” 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 “ \triangleq ” will be used to denote “is defined as” or “is by definition equal to.” Transposition will be indicated by a dagger (\dagger) superscript. The equalities in this work are considered to hold in the weak (“distributional”) sense. The right-sides of Equations (1) and (2), as well as of other various equations to be derived in this work, may contain “generalized functions/functionals”, particularly Dirac-distributions and derivatives thereof. The quantities which appear in Equations (1) and (2) are defined below.

1) The quantity $\alpha \triangleq (\alpha_1, \dots, \alpha_{TP})^\dagger$ is a TP -dimensional vector, having components $\alpha_1, \dots, \alpha_{TP}$, which denote the model’s imprecisely known parameters. The quantity “ TP ” denotes the “total number of model parameters.” Without loss of generality, the model parameters $\alpha \triangleq (\alpha_1, \dots, \alpha_{TP})^\dagger \in D_\alpha \in \mathbb{R}^{TP}$ can be considered to be real scalars defined over a domain D_α , which is included in a TP -dimensional subset of the \mathbb{R}^{TP} . These model parameters usually stem from processes that are external to the physical system under consideration and their precise value is seldom, if ever, known. The known characteristics of the model parameters may include their nominal (expected/mean) values and, possibly, higher-order moments or cumulants (*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. The components of the TP -dimensional column vector $\alpha \in \mathbb{R}^{TP}$ are considered to also include imprecisely known geometrical parameters that characterize the physical system’s boundaries in the phase-space of the model’s independent variables. Mathematically, the model parameters can be considered to be quasi-random scalar-valued quantities which follow an unknown multivariate distribution de-

noted as $p_\alpha(\boldsymbol{\alpha})$. The mean values (which in the context of computational modeling are called “nominal” values) of the model parameters will be denoted as α_i^0 ; the superscript “0” will be used throughout this monograph to denote “nominal values.” These nominal values are formally defined as follows:

$$\alpha_i^0 \triangleq \int_{D_\alpha} \alpha_i p_\alpha(\boldsymbol{\alpha}) d\boldsymbol{\alpha}, \quad i = 1, \dots, TP. \tag{3}$$

The expected values of the measured parameters will be considered to constitute the components of a vector denoted as $\boldsymbol{\alpha}^0 \triangleq [\alpha_1^0, \dots, \alpha_i^0, \dots, \alpha_{TP}^0]^\dagger$. The covariance, $\text{cov}(\alpha_i, \alpha_j)$, of two model parameters, α_i and α_j , is defined as follows:

$$\text{cov}(\alpha_i, \alpha_j) \triangleq \int_{D_\alpha} (\alpha_i - \alpha_i^0)(\alpha_j - \alpha_j^0) p_\alpha(\boldsymbol{\alpha}) d\boldsymbol{\alpha}; \quad i, j = 1, \dots, TP. \tag{4}$$

The covariances $\text{cov}(\alpha_i, \alpha_j)$ are considered to constitute the components of the covariance matrix for the model parameters and will be denoted as $C_{\alpha\alpha} \triangleq [\text{cov}(\alpha_i, \alpha_j)]_{TP \times TP}$.

2) The generic nonlinear model is considered to comprise TI independent variables which will be denoted as $x_i, i = 1, \dots, TI$, and which are considered to be components of a TI -dimensional column vector denoted as $\mathbf{x} \triangleq (x_1, \dots, x_{TI})^\dagger \in \mathbb{R}^{TI}$, where the sub/superscript “ TI ” denotes the “total number of independent variables.” The vector $\mathbf{x} \in \mathbb{R}^{TI}$ of independent variables is considered to be defined on a phase-space domain which will be denoted as $\Omega(\boldsymbol{\alpha})$ and which is defined as follows: $\Omega(\boldsymbol{\alpha}) \triangleq \{-\infty \leq \lambda_i(\boldsymbol{\alpha}) \leq x_i \leq \omega_i(\boldsymbol{\alpha}) \leq \infty; i = 1, \dots, TI\}$. The lower boundary-point of an independent variable is denoted as $\lambda_i(\boldsymbol{\alpha})$ and the corresponding upper boundary-point is denoted as $\omega_i(\boldsymbol{\alpha})$. The boundary $\Omega(\boldsymbol{\alpha})$ is also considered to be imprecisely known since it may depend on both geometrical parameters and material properties. A typical example of boundaries that depend on both geometrical parameters and material properties are the “boundaries facing vacuum” in models based on diffusion theory, where conditions are imposed on the “extrapolated boundary” of the respective spatial domain. The “extrapolated boundary” depends both on the imprecisely known physical dimensions of the problem’s domain and also on the medium’s properties, such as atomic number densities and microscopic transport cross sections. The boundary of $\Omega(\boldsymbol{\alpha})$, denoted as $\partial\Omega(\boldsymbol{\alpha}) \triangleq \{\lambda_i(\boldsymbol{\alpha}) \cup \omega_i(\boldsymbol{\alpha}), i = 1, \dots, TI\}$, includes the set of the endpoints $\lambda_i(\boldsymbol{\alpha}), \omega_i(\boldsymbol{\alpha}), i = 1, \dots, TI$ of the respective intervals on which the components of \mathbf{x} are defined.

3) The vector $\mathbf{v}(\mathbf{x}) \triangleq [v_1(\mathbf{x}), \dots, v_{TD}(\mathbf{x})]^\dagger$ is a TD -dimensional column vector, having components $v_i(\mathbf{x}), i = 1, \dots, TD$, which represent the model’s dependent variables (also called the model’s “state functions”). The abbreviation “ TD ” denotes “total number of dependent variables.” Without loss of generality, we can consider that $\mathbf{v}(\mathbf{x}) \in E_v$, where E_v is a normed linear space over the scalar field F of real numbers.

4) The quantity $N^m[\mathbf{v}(\mathbf{x}); \boldsymbol{\alpha}] \triangleq [N_1^m(\mathbf{v}; \boldsymbol{\alpha}), \dots, N_{TD}^m(\mathbf{v}; \boldsymbol{\alpha})]^\dagger$ is a TD -dimen-

sional column vector comprising the left-sides of the equations underlying the mathematical/computational model; the superscript “m” indicates “model.” The components $N_i^m(\mathbf{v};\boldsymbol{\alpha}), i=1,\dots,TD$ are operators (including differential, difference, integral, distributions, and/or finite or infinite matrices) acting nonlinearly on the dependent variables $\mathbf{v}(\mathbf{x})$, the independent variables \mathbf{x} and the model parameters $\boldsymbol{\alpha}$. The mapping $N^m(\mathbf{v};\boldsymbol{\alpha})$ is defined on the combined domains of the model’s parameters and state functions, *i.e.*, $N : D \subset E \rightarrow E_Q$, where $D = D_u \oplus D_\alpha$, $D_u \subset E_u$, $D_\alpha \subset E_\alpha$, $E = E_u \oplus E_\alpha$.

5) The vector $\mathbf{Q}^m(\mathbf{x},\boldsymbol{\alpha}) \triangleq [q_1^m(\mathbf{x};\boldsymbol{\alpha}), \dots, q_{TD}^m(\mathbf{x};\boldsymbol{\alpha})]^\dagger \in E_Q$ is a TD -dimensional column vector which represents inhomogeneous source terms (*i.e.*, the right-sides of the equations underlying the model), which usually depend nonlinearly on the uncertain parameters $\boldsymbol{\alpha}$. The vector $\mathbf{Q}^m(\mathbf{x},\boldsymbol{\alpha})$ is defined on a normed linear space denoted as E_Q .

6) In Equation (2), the vector $\mathbf{B}^m(\mathbf{v};\boldsymbol{\alpha}) \triangleq [B_1^m(\mathbf{v};\boldsymbol{\alpha}), \dots, B_{TB}^m(\mathbf{v};\boldsymbol{\alpha})]^\dagger$, where the subscript “TB” denotes “total number of boundary conditions,” comprises components $B_i^m(\mathbf{v};\boldsymbol{\alpha}), i=1,\dots,TB$, which are nonlinear operators in $\mathbf{v}(\mathbf{x})$ and $\boldsymbol{\alpha}$, which are defined on the boundary $\partial\Omega_x(\boldsymbol{\alpha})$ of the model’s domain $\Omega_x(\boldsymbol{\alpha})$. The vector $\mathbf{B}^s(\mathbf{x},\boldsymbol{\alpha}) \triangleq [B_1^s(\mathbf{x};\boldsymbol{\alpha}), \dots, B_{TB}^s(\mathbf{x};\boldsymbol{\alpha})]^\dagger$ comprises inhomogeneous boundary sources (indicated by the superscript “s”) which are nonlinear functions of $\boldsymbol{\alpha}$. The column vector $\mathbf{0}$ has TB components, all of which are identically zero.

Solving Equations (1) and (2) at the nominal parameter values, $\boldsymbol{\alpha}^0$, provides the “nominal solution” $\mathbf{v}^0(\mathbf{x})$, *i.e.*, the vectors $\mathbf{v}^0(\mathbf{x})$ and $\boldsymbol{\alpha}^0$ satisfy the following equations:

$$N^m[\mathbf{v}^0(\mathbf{x});\boldsymbol{\alpha}^0] = \mathbf{Q}^m(\mathbf{x},\boldsymbol{\alpha}^0), \quad \mathbf{x} \in \Omega_x, \tag{5}$$

$$\mathbf{B}^m[\mathbf{v}^0(\mathbf{x});\boldsymbol{\alpha}^0;\mathbf{x}] - \mathbf{B}^s(\mathbf{x},\boldsymbol{\alpha}^0) = \mathbf{0}, \quad \mathbf{x} \in \partial\Omega_x(\boldsymbol{\alpha}^0). \tag{6}$$

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” $\mathbf{r} \triangleq (r_1, \dots, r_k, \dots, r_{TR})^\dagger$. Each of these model responses is formally a function (implicit and/or explicit) of the dependent variables and model parameters $\boldsymbol{\alpha}$, which can be represented formally as follows:

$$r_k = r_k[\mathbf{v}(\mathbf{x});\boldsymbol{\alpha}]; \quad k = 1, \dots, TR. \tag{7}$$

In particular, a measurement of a physical quantity that depends on the model’s state functions and parameters can be considered to be a response denoted as $R_p[\mathbf{v}(\mathbf{x});\boldsymbol{\alpha}]$, which is to be evaluated at $\mathbf{x} = \mathbf{x}^p(\boldsymbol{\alpha})$, where $\mathbf{x}^p(\boldsymbol{\alpha}) \triangleq [x_1^p(\boldsymbol{\alpha}), \dots, x_k^p(\boldsymbol{\alpha}), \dots, x_{TR}^p(\boldsymbol{\alpha})]^\dagger$ denotes the location in phase-space of the specific “measurement point.” Such a measurement (or measurement-like) response can be represented mathematically as follows:

$$R_p[\mathbf{v}(\mathbf{x});\boldsymbol{\alpha}] \triangleq \int_{\lambda_1(\boldsymbol{\alpha})}^{\omega_1(\boldsymbol{\alpha})} \cdots \int_{\lambda_{TI}(\boldsymbol{\alpha})}^{\omega_{TI}(\boldsymbol{\alpha})} F[\mathbf{v}(\mathbf{x});\boldsymbol{\alpha};\mathbf{x}] \delta[x_1 - x_1^p(\boldsymbol{\alpha})] \times \cdots \times \delta[x_{TI} - x_{TI}^p(\boldsymbol{\alpha})] dx_1 \cdots dx_{TI}. \quad (8)$$

where the function $F[\mathbf{v}(\mathbf{x});\boldsymbol{\alpha};\mathbf{x}]$ comprises the mathematical dependence of the measurement device on the model's dependent variable(s), and where the quantity $\delta[x_i - x_i^p(\boldsymbol{\alpha})]$ denotes the Dirac-delta functional. The measurement's location in phase-space, $\mathbf{x}^p(\boldsymbol{\alpha})$, may itself be afflicted by measurement (experimental) uncertainties. Hence, the components of $\mathbf{x}^p(\boldsymbol{\alpha})$ must be included among the components of the vector $\boldsymbol{\alpha}$ of model parameters, even though the quantity $\mathbf{x}^p(\boldsymbol{\alpha})$ appears only in the definition of the response but does not appear in Equations (1) and (2), which mathematically define the physical model. Thus, the physical "system" is to be understood as being defined to comprise both the system's computational model and the system's responses. In most cases, the coordinates $x_k^p(\boldsymbol{\alpha})$, $k=1, \dots, TI$, will simply be independent (albeit uncertain) model parameters included among the components of the vector $\boldsymbol{\alpha}$, in which case $\partial x_k^p(\boldsymbol{\alpha})/\partial \alpha_n = 1$, if $\alpha_n \equiv x_k^p$ and $\partial x_k^p(\boldsymbol{\alpha})/\partial \alpha_n = 0$, if $\alpha_n \neq x_k^p$.

The expression on the right-side of Equation (8) can evidently be computed using the model, thus producing the "computed response" value, which will be denoted as $r_k^c[\mathbf{v}(\mathbf{x});\boldsymbol{\alpha};\mathbf{x}]$, and which can be compared to the corresponding "experimentally measured" value; the superscript "c" indicates "computed. The computed response $r_k^c[\mathbf{v}(\mathbf{x});\boldsymbol{\alpha};\mathbf{x}]$ depends (implicitly and/or explicitly) on the model's parameters and dependent variables, which also depend, in turn, on the model's parameters. Therefore, 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 simply as $r_k^c(\boldsymbol{\alpha})$. The nominal value of the response will be denoted $r_k(\boldsymbol{\alpha}^0)$, indicating that this value is obtained by computing the model response using the expected/nominal parameter values $\boldsymbol{\alpha}^0 \triangleq (\alpha_1^0, \dots, \alpha_{TP}^0)^\dagger$.

3. 2nd-BERRU-PMD: Second Order MaxEnt Predictive Modeling Methodology with Deterministically Included Computed Responses

This Section presents the mathematical and physical considerations leading to the development of the "second-order MaxEnt predictive modeling methodology for obtaining best-estimate results with reduced uncertainties" by deterministically incorporating the computational model. This methodology is designated using the acronym 2nd-BERRU-PMD, where the last letter ("D") indicates "deterministic," referring to the inclusion of the computed model responses. The posterior first-order moments (best-estimate predicted mean values) and posterior second-order moments (best-estimate predicted correlations) produced by the 2nd-BERRU-PMD methodology are compared with the results produced by the 1st-BERRU-PM methodology [14], as well as to the results produced by data

assimilation procedures, indicating the specific ways in which the 2nd-BERRU-PMD overcomes limitations of these latter procedures.

3.1. MaxEnt Representation of Experimental Information for Responses and Parameters

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 , $i = 1, \dots, TR$, and the second-order moments (variances/covariances), which will be denoted as $\text{cov}(r_i, r_j)_e$, $i, j = 1, \dots, TR$, for the measured responses. The letter “e” will be used either as a superscript or a superscript to indicate experimentally measured quantities. The expected values of the experimentally measured responses will be considered to constitute the components of a vector denoted as $\mathbf{r}^e \triangleq (r_1^e, \dots, r_{TR}^e)^\dagger$. 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

$\mathbf{C}_{rr}^e \triangleq [\text{cov}(r_i, r_j)_e]_{TR \times TR}$. 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 will be denoted as $\text{cor}(\alpha_i, r_j)_e$, $i = 1, \dots, TP; j = 1, \dots, TR$, and they can formally be considered to be elements of a rectangular correlation matrix which will be denoted as $\mathbf{C}_{\alpha r}^e \triangleq [\text{cor}(\alpha_i, r_j)_e]_{TP \times TR}$. As discussed in Section 2, cf. Equations (3) and (4), the model parameters are characterized by the vector of mean values $\boldsymbol{\alpha}^0 \triangleq [\alpha_1^0, \dots, \alpha_i^0, \dots, \alpha_{TP}^0]^\dagger$ and the covariance matrix

$$\mathbf{C}_{\alpha\alpha} \triangleq [\text{cov}(\alpha_i, \alpha_j)]_{TP \times TP}.$$

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

$$p_e(\mathbf{z} | \mathbf{z}_e, \mathbf{C}_e) = \frac{(2\pi)^{-(TR+TP)/2}}{\sqrt{\text{Det}(\mathbf{C}_e)}} \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{z}_e)^\dagger \mathbf{C}_e^{-1} (\mathbf{z} - \mathbf{z}_e)\right], \tag{9}$$

where $\text{Det}(\mathbf{C}_e)$ denotes the determinant of the matrix \mathbf{C}_e and where:

$$\mathbf{C}_e \triangleq \begin{pmatrix} \mathbf{C}_{rr}^e & \mathbf{C}_{r\alpha}^e \\ \mathbf{C}_{\alpha r}^e & \mathbf{C}_{\alpha\alpha} \end{pmatrix}; \mathbf{z} = \begin{pmatrix} \mathbf{r} \\ \boldsymbol{\alpha} \end{pmatrix}; \mathbf{z}_e = \begin{pmatrix} \mathbf{r}^e \\ \boldsymbol{\alpha}^0 \end{pmatrix}. \tag{10}$$

3.2. Construction of the Joint Posterior MaxEnt Probability Distribution 2nd-BERRU-PMD

Consistent with the consideration that only mean values and covariances (*i.e.*, first- and second-order distributional moments) are available, the 2nd-BERRU-PMD methodology considers that only first- and second-order sensitivities of the computed responses with respect to the model parameters are available. In this case, the vector of model responses can be represented by the following

second-order Taylor-series written in “vector/matrix” form:

$$\begin{aligned}
 \mathbf{r} &= \mathbf{r}^c(\boldsymbol{\alpha}^0) + [\mathbf{S}(\boldsymbol{\alpha}^0)](\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) + \mathbf{p}(\boldsymbol{\alpha}; \boldsymbol{\alpha}^0); \\
 \mathbf{p}(\boldsymbol{\alpha}; \boldsymbol{\alpha}^0) &\triangleq [p_1(\boldsymbol{\alpha}; \boldsymbol{\alpha}^0), \dots, p_{TR}(\boldsymbol{\alpha}; \boldsymbol{\alpha}^0)]^\dagger; \\
 p_k(\boldsymbol{\alpha}; \boldsymbol{\alpha}^0) &\triangleq \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0)^\dagger [\mathbf{Q}_k(\boldsymbol{\alpha}^0)](\boldsymbol{\alpha} - \boldsymbol{\alpha}^0),
 \end{aligned}
 \tag{11}$$

where the components of the vector $\mathbf{r}^c(\boldsymbol{\alpha}^0) \triangleq [r_1^c(\boldsymbol{\alpha}^0), \dots, r_{TR}^c(\boldsymbol{\alpha}^0)]^\dagger$ represent the values of the model responses computed at the nominal parameter values $\boldsymbol{\alpha}^0$, while 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 $\boldsymbol{\alpha}^0$, and are defined below.

$$\begin{aligned}
 \mathbf{S}(\boldsymbol{\alpha}) &\triangleq \begin{pmatrix} s_{11} & \dots & s_{1,TR} \\ \vdots & \ddots & \vdots \\ s_{TR,1} & \dots & s_{TR,TR} \end{pmatrix} \triangleq \begin{pmatrix} \frac{\partial r_1(\boldsymbol{\alpha})}{\partial \alpha_1} & \dots & \frac{\partial r_1(\boldsymbol{\alpha})}{\partial \alpha_{TP}} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_{TR}(\boldsymbol{\alpha})}{\partial \alpha_1} & \dots & \frac{\partial r_{TR}(\boldsymbol{\alpha})}{\partial \alpha_{TP}} \end{pmatrix}; \\
 \mathbf{Q}_k(\boldsymbol{\alpha}) &\triangleq \begin{pmatrix} q_{11}^{(k)} & \dots & q_{1,TP}^{(k)} \\ \vdots & \ddots & \vdots \\ q_{TP,1}^{(k)} & \dots & q_{TP,TP}^{(k)} \end{pmatrix} \triangleq \begin{pmatrix} \frac{\partial^2 r_k^c(\boldsymbol{\alpha})}{\partial \alpha_1 \partial \alpha_1} & \dots & \frac{\partial^2 r_k^c(\boldsymbol{\alpha})}{\partial \alpha_1 \partial \alpha_{TP}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 r_k^c(\boldsymbol{\alpha})}{\partial \alpha_{TP} \partial \alpha_1} & \dots & \frac{\partial^2 r_k^c(\boldsymbol{\alpha})}{\partial \alpha_{TP} \partial \alpha_{TP}} \end{pmatrix}.
 \end{aligned}
 \tag{12}$$

The merging of the experimental information with the computational information using the mathematical model is accomplished by seeking a mathematical representation of a “true” probability distribution $p(\mathbf{z})$, $\mathbf{z} \triangleq (\mathbf{r}, \boldsymbol{\alpha})^\dagger$, which incorporates all of the second-order information provided in the foregoing about the model parameters, measured and computed responses. This “true” probability distribution must satisfy the following characteristics: 1) $p(\mathbf{z})$ is normalized to unity; 2) $p(\mathbf{z})$ is approximated as closely as possible by the experimental distribution $p_e(\mathbf{z}|\mathbf{z}^e, \mathbf{C}_e)$ and 3) $p(\mathbf{z})$ satisfies the following integral forms of the deterministic relationships among responses and parameters represented mathematically by Equation (11):

$$\begin{aligned}
 \int_{D_z} g_k(\mathbf{z}) p(\mathbf{z}) d\mathbf{z} &= 0; \\
 g_k(\mathbf{z}) &\triangleq r_k - r_k^c(\boldsymbol{\alpha}^0) - \sum_{i=1}^{TP} s_{ki} \delta \alpha_i - \frac{1}{2} \sum_{i=1}^{TP} \sum_{j=1}^{TP} q_{ij}^{(k)} \delta \alpha_i \delta \alpha_j = 0; \quad k = 1, \dots, TR.
 \end{aligned}
 \tag{13}$$

The “true” normalized distribution $p(\mathbf{z})$ which minimizes the discrepancy between it and the distribution $p_e(\mathbf{z}|\mathbf{z}^e, \mathbf{C}_e)$ while satisfying the constraints expressed by Equation (13) is obtained by applying the steps generally outlined in Appendix B, which yields the following expressions:

$$p(\mathbf{z}) = \frac{p_e(\mathbf{z}|\mathbf{z}^e, \mathbf{C}_e)}{Z} \exp \left\{ - \sum_{k=1}^{TR} \theta_k g_k(\mathbf{z}) \right\};
 \tag{14}$$

$$Z \triangleq \int_{D_z} p_e(z|z^e, C_e) \exp\left\{-\sum_{k=1}^{TR} \theta_k g_k(z)\right\} dz. \tag{15}$$

In Equations (14) and (15), the quantities $\theta_1, \dots, \theta_k, \dots, \theta_{TR}$ are Lagrange multipliers which remain to be determined. In statistical mechanics, the normalization integral Z defined in Equation (15) is called the partition function (or sum over states) and carries all of the information available about the possible states of the system. The quantity $(-\log Z)$ is called the “free energy” and plays an important role in determining the expressions of the Lagrange multipliers and hence of the probability distribution $p(z)$. In the present case, it follows from Equations (15) and (13) that:

$$\frac{\partial[-\ln Z]}{\partial \theta_k} = \frac{1}{Z} \int_{D_z} g_k(z) p_e(z|z^e, C_e) \exp\left\{-\sum_{k=1}^{TR} \theta_k g_k(z)\right\} dz = 0, \quad k = 1, \dots, TR. \tag{16}$$

On the other hand, alternative expressions for the derivatives $\partial[-\ln Z]/\partial \theta_k$ of the free energy with respect to the Lagrange multipliers can be obtained by evaluating the integral which defines the partition function Z in Equation (15). Equating these alternative expressions to zero, in view of Equation (16), will provide equations which are to be solved to obtain the explicit expressions of the Lagrange multipliers.

Integrals such as represented by Equation (15) can be quantified, to any desired order of accuracy—even exactly—using the well-known saddle-point (Laplace expansion) method, see e.g., [20]. Examining the exponent of the function entering the normalization integral Z defined in Equation (15) readily indicates that if the term involving the second-order sensitivities of the response with respect to the parameters is retained in the exponent, the expression of the saddle-point will involve the vector of Lagrange multipliers, $\theta \triangleq (\theta_1, \dots, \theta_k, \dots, \theta_{TR})^\dagger$, in a nonlinear equation which cannot be solved analytically to obtain a closed-form expression for these Lagrange multipliers. Furthermore, solving numerically this nonlinear equation in θ would require the inversion of the (very large) parameter-covariance matrix $C_{\alpha\alpha}$, which is impractical. On the other hand, as will be shown below, a closed-form expression for the saddle-point and, consequently, for the Lagrange multipliers can be obtained if the terms containing the second-order response sensitivities are considered by expanding the respective exponential term in a Taylor-series in powers of $\theta \triangleq (\theta_1, \dots, \theta_k, \dots, \theta_{TR})^\dagger$ or, equivalently, around the nominal parameter values. Expanding the second-order terms transforms the normalization integral Z defined in Equation (15) into the following form:

$$Z \triangleq \int_{D_z} f(r, \alpha) e^{-h(r, \alpha; \theta)} dr d\alpha, \tag{17}$$

where $\theta \triangleq (\theta_1, \dots, \theta_k, \dots, \theta_{TR})^\dagger$ and where:

$$h(r, \alpha; \theta) \triangleq \frac{1}{2} \begin{pmatrix} r - r^e \\ \alpha - \alpha^0 \end{pmatrix}^\dagger \begin{pmatrix} C_{rr}^e & C_{ra}^e \\ C_{ar}^e & C_{\alpha\alpha}^e \end{pmatrix}^{-1} \begin{pmatrix} r - r^e \\ \alpha - \alpha^0 \end{pmatrix} + \theta^\dagger \left[-r + r^c(\alpha^0) + S(\alpha - \alpha^0) \right], \tag{18}$$

$$f(\mathbf{r}, \boldsymbol{\alpha}) \triangleq 1 - \frac{1}{2} \sum_{k=1}^{TR} \theta_k (\boldsymbol{\alpha} - \boldsymbol{\alpha}^0)^\dagger [\mathbf{Q}_k(\boldsymbol{\alpha}^0)] (\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) + \text{hots}, \quad (19)$$

where the acronym “hots” stands for “higher-order terms in the expansion of the exponential $\exp\left\{-\sum_{k=1}^{TR} \theta_k g_k(\mathbf{z})\right\}$.” Neglecting these higher-order terms, it follows from Equations (17)-(19) that

$$Z = I_1 - \frac{1}{2} \sum_{k=1}^{TR} \theta_k J_k, \quad (20)$$

where

$$I_1 \triangleq \int_{D_z} e^{-h(\mathbf{r}, \boldsymbol{\alpha}; \boldsymbol{\theta})} d\mathbf{r} d\boldsymbol{\alpha}, \quad (21)$$

$$J_k \triangleq \int_{D_z} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^0)^\dagger [\mathbf{Q}_k(\boldsymbol{\alpha}^0)] (\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) e^{-h(\mathbf{r}, \boldsymbol{\alpha}; \boldsymbol{\theta})} d\mathbf{r} d\boldsymbol{\alpha}. \quad (22)$$

The saddle-point, denoted as $\mathbf{z}^{sp} \triangleq (\mathbf{r}^{sp}, \boldsymbol{\alpha}^{sp})^\dagger$, of the function $h(\mathbf{r}, \boldsymbol{\alpha}; \boldsymbol{\theta})$ is defined as the point where the following partial gradients vanish:

$$\nabla_{\mathbf{r}} h(\mathbf{r}, \boldsymbol{\alpha}; \boldsymbol{\theta}) = \mathbf{0}, \quad \nabla_{\boldsymbol{\alpha}} h(\mathbf{r}, \boldsymbol{\alpha}; \boldsymbol{\theta}) = \mathbf{0}, \quad \text{at } \mathbf{z} = \mathbf{z}^{sp}. \quad (23)$$

The conditions represented by Equation (23) yield the following equation:

$$\begin{pmatrix} \mathbf{C}_{rr}^e & \mathbf{C}_{r\alpha}^e \\ \mathbf{C}_{\alpha r}^e & \mathbf{C}_{\alpha\alpha}^e \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{r}^{sp} - \mathbf{r}^e \\ \boldsymbol{\alpha}^{sp} - \boldsymbol{\alpha}^0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\theta} \\ -\mathbf{S}^\dagger \boldsymbol{\theta} \end{pmatrix}. \quad (24)$$

Solving Equation (24) yields the following expressions for the components of the saddle point $\mathbf{z}^{sp} \triangleq (\mathbf{r}^{sp}, \boldsymbol{\alpha}^{sp})^\dagger$ of the function $h(\mathbf{r}, \boldsymbol{\alpha}; \boldsymbol{\theta})$:

$$\mathbf{r}^{sp} = \mathbf{r}^e + (\mathbf{C}_{rr}^e - \mathbf{C}_{r\alpha}^e \mathbf{S}^\dagger) \boldsymbol{\theta}, \quad (25)$$

$$\boldsymbol{\alpha}^{sp} = \boldsymbol{\alpha}^0 + (\mathbf{C}_{\alpha r}^e - \mathbf{C}_{\alpha\alpha}^e \mathbf{S}^\dagger) \boldsymbol{\theta}. \quad (26)$$

Using the saddle point method (which is exact for the integral I_1) or “completing the square in the exponent of I_1 ” yields the following exact expression for the integral I_1 :

$$I_1 = (2\pi)^{(TR+TP)/2} |C_e|^{1/2} \exp\left[-\boldsymbol{\theta}^\dagger \mathbf{d} + \frac{1}{2} \begin{pmatrix} \boldsymbol{\theta} \\ -\mathbf{S}^\dagger \boldsymbol{\theta} \end{pmatrix}^\dagger C_e \begin{pmatrix} \boldsymbol{\theta} \\ -\mathbf{S}^\dagger \boldsymbol{\theta} \end{pmatrix}\right], \quad (27)$$

where $|C_e|$ denotes the determinant of the matrix C_e and where \mathbf{d} denotes the “vector of deviations between the nominal values of computed and measured responses” and is defined below:

$$\mathbf{d} \triangleq \mathbf{r}^c(\boldsymbol{\alpha}^0) - \mathbf{r}^e; \quad \mathbf{d} \triangleq (d_1, \dots, d_{TR})^\dagger; \quad d_i \triangleq I_i^c(\boldsymbol{\alpha}^0) - r_i^e; \quad i = 1, \dots, TR. \quad (28)$$

The expression of the response in terms of the computational model, *i.e.*, Equation (11), can be used to obtain the following representation for \mathbf{d} in terms of the computed responses:

$$\mathbf{d} \triangleq \mathbf{r} - \mathbf{r}^e - \mathbf{S}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) - \mathbf{p}(\boldsymbol{\alpha}; \boldsymbol{\alpha}^0). \quad (29)$$

Thus, the “vector of deviations” \mathbf{d} can be considered to be a (quasi-) ran-

dom variable. The expressions of the “vector of mean values of deviations” $\bar{\mathbf{d}} \triangleq [\bar{d}_1, \dots, \bar{d}_{TR}]^\dagger$ and the covariance matrix $\mathbf{D} \triangleq [D_{ij}]_{TR \times TR}$ of the vector of deviations \mathbf{d} are derived below:

$$\bar{\mathbf{d}} \triangleq \int_{D_z} [\mathbf{r} - \mathbf{r}^e - \mathbf{S}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) - \mathbf{p}(\boldsymbol{\alpha}; \boldsymbol{\alpha}^0)] p_e(\mathbf{z} | \mathbf{z}^e, \mathbf{C}_e) d\mathbf{z} = [\bar{d}_1, \dots, \bar{d}_{TR}]^\dagger, \quad (30)$$

$$\bar{d}_k \triangleq \sum_{i=1}^{TP} \sum_{j=1}^{TP} q_{ij}^{(k)} \text{cov}(\alpha_i, \alpha_j); \quad k = 1, \dots, TR; \quad (31)$$

$$\begin{aligned} \mathbf{D} &\triangleq \int_{D_z} (\mathbf{d} \mathbf{d}^\dagger) p_e(\mathbf{z} | \mathbf{z}^e, \mathbf{C}_e) d\mathbf{z} \\ &= \int_{D_z} p_e(\mathbf{z} | \mathbf{z}^e, \mathbf{C}_e) [\mathbf{r} - \mathbf{r}^e - \mathbf{S}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) - \mathbf{p}(\boldsymbol{\alpha}; \boldsymbol{\alpha}^0)] \\ &\quad \times [\mathbf{r} - \mathbf{r}^e - \mathbf{S}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) - \mathbf{p}(\boldsymbol{\alpha}; \boldsymbol{\alpha}^0)]^\dagger d\mathbf{z} \\ &\cong \mathbf{C}_{rr}^e + \mathbf{S} \mathbf{C}_{\alpha\alpha} \mathbf{S}^\dagger - \mathbf{C}_{ra}^e \mathbf{S}^\dagger - \mathbf{S} \mathbf{C}_{ar}^e. \end{aligned} \quad (32)$$

The third- and fourth-order correlations were neglected in the final expression obtained for the covariance matrix $\mathbf{D} \triangleq [D_{ij}]_{TR \times TR}$ in Equation (32). Notably, the result obtained in Equation (30) indicates that if the second-order sensitivities of the response with respect to the model parameters are neglected, then $\bar{d}_k = 0, k = 1, \dots, TR$, in which case \mathbf{d} would be a random vector with zero-mean and covariance \mathbf{D} . Therefore, neglecting the second-order sensitivities would imply that the computed model responses would be consistent with the measurements only if $\bar{\mathbf{d}} = \mathbf{0}$, i.e., $\mathbf{r}^e(\boldsymbol{\alpha}^0) \cong \mathbf{r}^e$. Although such a situation would not be impossible to conceive, it would probably not occur often in practice. These considerations further underscore the significant impact which the second-order response sensitivities have on the correct interpretation of actual, physical systems.

Applying the saddle-point method to the expression provided in Equation (22) yields the following expression:

$$J_k = I_1(\boldsymbol{\alpha}^{sp} - \boldsymbol{\alpha}^0)^\dagger [\mathbf{Q}_k(\boldsymbol{\alpha}^0)] (\boldsymbol{\alpha}^{sp} - \boldsymbol{\alpha}^0) [1 + O(C_{\alpha\alpha})], \quad (33)$$

where the quantity $O(C_{\alpha\alpha})$ denotes terms of the order of the components of the parameter covariance matrix $C_{\alpha\alpha}$, and can therefore be neglected by comparison to the leading term of order unity.

Replacing the expression obtained in Equations (27) and (33) into Equation (20) and neglecting the higher-order terms yields the following expression for the partition function:

$$Z = I_1 \left\{ 1 - \frac{1}{2} \sum_{k=1}^{TR} \theta_k (\boldsymbol{\alpha}^{sp} - \boldsymbol{\alpha}^0)^\dagger [\mathbf{Q}_k(\boldsymbol{\alpha}^0)] (\boldsymbol{\alpha}^{sp} - \boldsymbol{\alpha}^0) \right\}. \quad (34)$$

Taking the logarithm of Equation (34) yields the following expression for the “free energy”:

$$-\ln Z = \frac{TR + TP}{2} \ln(2\pi) + \frac{1}{2} \ln |C_m| - \boldsymbol{\theta}^\dagger \mathbf{d} + h_1(\boldsymbol{\theta}) + h_2(\boldsymbol{\theta}), \quad (35)$$

$$h_1(\boldsymbol{\theta}) \triangleq \frac{1}{2} \left[\boldsymbol{\theta}^\dagger \mathbf{C}_{rr}^e \boldsymbol{\theta} - \boldsymbol{\theta}^\dagger (\mathbf{C}_{ra}^e \mathbf{S}^\dagger + \mathbf{S} \mathbf{C}_{ar}^e) \boldsymbol{\theta} + \boldsymbol{\theta}^\dagger \mathbf{S} \mathbf{C}_{aa}^e \mathbf{S}^\dagger \boldsymbol{\theta} \right], \quad (36)$$

$$h_2(\boldsymbol{\theta}) \triangleq \ln \left\{ 1 - \frac{1}{2} \sum_{k=1}^{TR} \theta_k (\boldsymbol{\alpha}^{sp} - \boldsymbol{\alpha}^0)^\dagger \left[\mathbf{Q}_k(\boldsymbol{\alpha}^0) \right] (\boldsymbol{\alpha}^{sp} - \boldsymbol{\alpha}^0) \right\} \\ = \ln \left\{ 1 - \frac{1}{2} \sum_{k=1}^{TR} \theta_k \left[\boldsymbol{\theta}^\dagger \mathbf{M}^{(k)} \boldsymbol{\theta} \right] \right\}, \quad (37)$$

$$\mathbf{M}^{(k)} \triangleq \left[M_{ij}^{(k)} \right]_{TR \times TR} \triangleq (\mathbf{C}_{ar}^e - \mathbf{C}_{aa}^e \mathbf{S}^\dagger)^\dagger \left[\mathbf{Q}_k(\boldsymbol{\alpha}^0) \right] (\mathbf{C}_{ar}^e - \mathbf{C}_{aa}^e \mathbf{S}^\dagger) = \left[\mathbf{M}^{(k)} \right]^\dagger. \quad (38)$$

The expression for $h_2(\boldsymbol{\theta})$ presented in Equation (37) has been obtained by using Equation (26). As indicated in Equations (36) and (37), respectively, the quantity $h_1(\boldsymbol{\theta})$ contains only first-order response sensitivities (with respect to the model's parameters) and second-order bilinear terms in the components of $\boldsymbol{\theta}$, while the quantity $h_2(\boldsymbol{\theta})$ comprises all second-order response sensitivities and terms higher than second-order in the components of $\boldsymbol{\theta}$.

In view of Equations (35)-(37), the derivatives of the free energy with respect to the Lagrange multipliers θ_k are provided by the expression below:

$$\frac{\partial [-\ln Z]}{\partial \theta_k} = -d_k + \frac{\partial h_1(\boldsymbol{\theta})}{\partial \theta_k} + \frac{\partial h_2(\boldsymbol{\theta})}{\partial \theta_k}, \quad k = 1, \dots, TR. \quad (39)$$

It follows from Equations (39) and (16) that the vector of Lagrange multipliers must be the solution of the following system of coupled nonlinear equations:

$$w_k(\boldsymbol{\theta}) \triangleq \frac{\partial h_1(\boldsymbol{\theta})}{\partial \theta_k} + \frac{\partial h_2(\boldsymbol{\theta})}{\partial \theta_k} - d_k = 0, \quad k = 1, \dots, TR, \quad (40)$$

where:

$$\frac{\partial h_1(\boldsymbol{\theta})}{\partial \theta_k} = \sum_{j=1}^{TR} D_{kj} \theta_j; \quad \left[D_{kj} \right]_{TR \times TR} \triangleq \mathbf{D} \triangleq \mathbf{C}_{rr}^e + \mathbf{S} \mathbf{C}_{aa}^e \mathbf{S}^\dagger - \mathbf{C}_{ra}^e \mathbf{S}^\dagger - \mathbf{S} \mathbf{C}_{ar}^e; \quad (41)$$

$$\frac{\partial h_2(\boldsymbol{\theta})}{\partial \theta_k} = -\frac{1}{2} \frac{\frac{\partial}{\partial \theta_k} \sum_{i=1}^{TR} \theta_i \left[\boldsymbol{\theta}^\dagger \mathbf{M}^{(i)} \boldsymbol{\theta} \right]}{1 - \frac{1}{2} \sum_{i=1}^{TR} \theta_i \left[\boldsymbol{\theta}^\dagger \mathbf{M}^{(i)} \boldsymbol{\theta} \right]} \cong -\frac{1}{2} \boldsymbol{\theta}^\dagger \mathbf{M}^{(k)} \boldsymbol{\theta} - \sum_{i=1}^{TR} \theta_i \sum_{j=1}^{TR} M_{kj}^{(i)} \theta_j \triangleq u_k(\boldsymbol{\theta}). \quad (42)$$

The approximation $1 - \frac{1}{2} \sum_{j=1}^{TR} \theta_j \left[\boldsymbol{\theta}^\dagger \mathbf{M}^{(j)} \boldsymbol{\theta} \right] \cong 1$ has been used to obtain the approximate expression on the right-side of Equation (42), which thus comprises just the bilinear terms in $\boldsymbol{\theta}$, neglecting the high-order terms involving the components of $\boldsymbol{\theta}$, which are expected to be comparatively small.

Introducing the vectors $\mathbf{w}(\boldsymbol{\theta}) \triangleq [w_1(\boldsymbol{\theta}), \dots, w_k(\boldsymbol{\theta}), \dots, w_{TR}(\boldsymbol{\theta})]^\dagger$ and $\mathbf{u}(\boldsymbol{\theta}) \triangleq [u_1(\boldsymbol{\theta}), \dots, u_k(\boldsymbol{\theta}), \dots, u_{TR}(\boldsymbol{\theta})]^\dagger$ enables the compact writing, in matrix-vector form, of Equation (40), as follows:

$$\mathbf{w}(\boldsymbol{\theta}) \triangleq -\mathbf{d} + \mathbf{D}\boldsymbol{\theta} + \mathbf{u}(\boldsymbol{\theta}) = \mathbf{0}. \quad (43)$$

If only the first-order sensitivities of the response with respect to the model parameters are considered (and thus the second- and higher-order sensitivities are neglected or are unavailable), the term $\mathbf{u}(\boldsymbol{\theta})$ is neglected and the remain-

ing first-order term in Equation (43) yields the following first-order solution, denoted as $\boldsymbol{\theta}^{(1)}$, for the vector of Lagrange multipliers:

$$\boldsymbol{\theta}^{(1)} = \mathbf{D}^{-1}\mathbf{d}. \quad (44)$$

Notably, the matrix \mathbf{D} has dimensions $TR \times TR$. Since $TR \ll TP$ in practical situations, it follows that the computation of the coordinates of the saddle-point $\mathbf{z}^{sp} \triangleq (\mathbf{r}^{sp}, \boldsymbol{\alpha}^{sp})^\dagger$ involves the inversion of just the smallest-possible matrix (*i.e.*, \mathbf{D}) that arises in the “predictive modeling” methodology. Replacing the result obtained for $\boldsymbol{\theta}^{(1)}$ in Equations (25) and (26) provides the corresponding expressions for the coordinates \mathbf{r}^{sp} and $\boldsymbol{\alpha}^{sp}$ of the saddle-point $\mathbf{z}^{sp} \triangleq (\mathbf{r}^{sp}, \boldsymbol{\alpha}^{sp})^\dagger$.

Using the results obtained for $\boldsymbol{\theta}^{(1)}$ in Equation 44) leads to the expressions already presented in the 1st-BERRU-PM methodology [14] for the saddle point $\mathbf{z}^{sp} \triangleq (\mathbf{r}^{sp}, \boldsymbol{\alpha}^{sp})^\dagger$ and, subsequently for the best-estimate (calibrated) mean values for the parameters and responses, as well as for the best-estimate covariance matrices for the corresponding best-estimate parameters and responses. These expressions will not be reproduced here, since they will also be obtained as particular cases of the 2nd-BERRU-PM methodology to be presented in the remainder of this Section.

The goal of predictive modeling is to reduce, as much possible, the discrepancies between the measured response values and the corresponding values predicted by the computational model. But if the nominal values of the measured responses happen—by nature, design or accident—to coincide with the corresponding computed values, the vector of deviations vanishes, *i.e.*, $\mathbf{d} = \mathbf{0}$, in which case Equations (44), (25), (26) and (14) indicate that:

$$\mathbf{d} = \mathbf{0}; \Rightarrow \boldsymbol{\theta}^{(1)} = \mathbf{0}; \Rightarrow \mathbf{r}^{sp} = \mathbf{r}^e, \quad \boldsymbol{\alpha}^{sp} = \boldsymbol{\alpha}^0; \Rightarrow p(\mathbf{z}) = p_e(\mathbf{z} | \mathbf{z}^e, \mathbf{C}_e). \quad (45)$$

The results shown in Equation (45) indicate that if only first-order sensitivities are used and if the nominal values of the measured responses happen to coincide with the corresponding computed values, the model would have no impact on the initial (measured) distribution of responses and parameters, so there would be no calibration of either the predicted responses or parameters. Evidently this conclusion reveals a fundamental limitation of using just the first-order sensitivities for predictive modeling (and data assimilation and/or data adjustment methods). But this conclusion also provides the fundamental motivation for using second-order sensitivities, in addition to the natural expectation that using the second-order sensitivities will produce more accurate results, in general.

It is also important to note that if the response measurements were perfect and if they were uncorrelated to the model parameters, *i.e.*, if $\mathbf{C}_{rr}^e = \mathbf{0}$ and $\mathbf{C}_{ra}^e = \mathbf{C}_{ar}^e = \mathbf{0}$, the matrix $\mathbf{D} = \mathbf{S}\mathbf{C}_{\alpha\alpha}\mathbf{S}^\dagger$ would nevertheless be non-singular since the matrix $\mathbf{S}\mathbf{C}_{\alpha\alpha}\mathbf{S}^\dagger$ is nonsingular. This, the solution shown in Equation (44) would still exist; this is in contradistinction with the extant data assimilation schemes, which fail if the measurements are perfect. Notably, all extant data assimilation schemes consider that the response measurements and model para-

meters are uncorrelated.

If the second-order sensitivities, which are included in the term $\mathbf{u}(\boldsymbol{\theta})$, are retained in Equation (40), this equation becomes nonlinear in the components of $\boldsymbol{\theta} \triangleq (\theta_1, \dots, \theta_k, \dots, \theta_{TR})^\dagger$ and therefore requires an iterative method for solving. The structure of $\mathbf{u}(\boldsymbol{\theta})$ indicates that the simplest (and probably the most efficient) method for solving Equation (43) would be Newton's method, as can be surmised by the similarity between this equation and a simple quadratic algebraic equation. In this vein, consider the application of Newton's method to the quadratic equation, as described below:

$$\begin{aligned} q(x) &\triangleq ax^2 + bx + d = 0; \quad q'(x) = 2ax + b; \\ x^{(n+1)} &= x^{(n)} - [q'(x^{(n)})]^{-1} q(x^{(n)}); \quad n = 0, 1, \dots \end{aligned} \tag{46}$$

If $d \neq 0$, the usual initial-guess for starting the Newton iteration is $x^{(0)} = -d/b$, which yields the first-iterate $x^{(1)} = -[1 + ad/(b^2 - 2ad)]d/b$. The correction term included in this first-iterate solution is of first order in the product (ad) , so the subsequent iterates become increasingly more accurate when this product is small; even the first-iterate becomes exact in the limit as $a \rightarrow 0$. On the other hand, if $d = 0$, the initial guess $x^{(0)} = -d/b = 0$ would yield solely the trivial solution $x = 0$. Therefore, if $d \cong 0$, the initial-guess for starting the Newton iteration would be $x^{(0)} = -b/a$, which would yield the non-zero first-iterate solution $x^{(1)} = -b/a + d/b$, and so on.

Applying Newton's method to solve Equation (43) yields the following iterative equation for finding its solution(s) $\boldsymbol{\theta} \triangleq (\theta_1, \dots, \theta_k, \dots, \theta_{TR})^\dagger$:

$$\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} - [\mathbf{A}(\boldsymbol{\theta}^{(n)})]^{-1} \mathbf{w}(\boldsymbol{\theta}^{(n)}); \quad n = 0, 1, \dots, \tag{47}$$

where the Jacobian matrix $\mathbf{A}(\boldsymbol{\theta})$ is defined as usual, namely:

$$\begin{aligned} \mathbf{A}(\boldsymbol{\theta}) &\triangleq \frac{\partial(w_1, \dots, w_k, \dots, w_{TR})}{\partial(\theta_1, \dots, \theta_n, \dots, \theta_{TR})} \triangleq \begin{bmatrix} \partial w_1 / \partial \theta_1 & \cdot & \partial w_1 / \partial \theta_{TR} \\ \cdot & \cdot & \cdot \\ \partial w_{TR} / \partial \theta_1 & \cdot & \partial w_{TR} / \partial \theta_{TR} \end{bmatrix} \\ &\triangleq \begin{bmatrix} A_{1,1}(\boldsymbol{\theta}) & \cdot & A_{1,TR}(\boldsymbol{\theta}) \\ \cdot & \cdot & \cdot \\ A_{TR,1}(\boldsymbol{\theta}) & \cdot & A_{TR,TR}(\boldsymbol{\theta}) \end{bmatrix}. \end{aligned} \tag{48}$$

Using Equations (42) and (43) yields the following expression for the components/elements $A_{k,n}(\boldsymbol{\theta})$ of the Jacobian matrix $\mathbf{A}(\boldsymbol{\theta})$:

$$\begin{aligned} A_{kn}(\boldsymbol{\theta}) &= \frac{\partial w_k}{\partial \theta_n} = \frac{\partial}{\partial \theta_n} \left[\sum_{j=1}^{TR} D_{kj} \theta_j - \frac{1}{2} \boldsymbol{\theta}^\dagger \mathbf{M}^{(k)} \boldsymbol{\theta} - \sum_{i=1}^{TR} \theta_i \sum_{j=1}^{TR} M_{kj}^{(i)} \theta_j \right] \\ &= D_{kn} - \sum_{j=1}^{TR} M_{nj}^{(k)} \theta_j - \sum_{j=1}^{TR} M_{kj}^{(n)} \theta_j - \sum_{i=1}^{TR} \theta_i M_{kn}^{(i)} = D_{kn} - V_{kn}(\boldsymbol{\theta}), \end{aligned} \tag{49}$$

where:

$$V_{kn}(\boldsymbol{\theta}) \triangleq \sum_{j=1}^{TR} \theta_j [M_{nj}^{(k)} + M_{kj}^{(n)} + M_{kn}^{(j)}], \quad \mathbf{V}(\boldsymbol{\theta}) \triangleq [V_{kn}(\boldsymbol{\theta})]_{TR \times TR}. \tag{50}$$

As indicated in Equation (49), if the initial guess $\theta^{(0)} = \mathbf{0}$ is used, the Jacobian matrix would take on the value $A(\theta = \mathbf{0}) = D$. Consequently, the resulting solution of Equation (47) would be obtained in a single iteration and would be the same as the first-order solution obtained in Equation (44), *i.e.*, $\theta^{(1)} = D^{-1}d$. Similarly, if the second-order response sensitivities to the model parameters are neglected, Equation (50) indicates that $A(M^{(k)} = \mathbf{0}) = D$.

To obtain the nontrivial solution of Equation (47) when d is very small or actually vanishes ($d = \mathbf{0}$), a non-zero initial guess ($\theta^{(0)} \neq \mathbf{0}$) must be used for starting the Newton iterations, as was discussed in the foregoing by analogy with the solution of a quadratic algebraic equation using Newton's method. The nontrivial solution of Equation (47) which would be valid even when $d = \mathbf{0}$ will be denoted as $\theta^{(h,n+1)}$, where the superscript "h" indicates that the expression of $\theta^{(h,n+1)}$ will include higher-order terms, which will involve the second-order sensitivities of computed responses with respect to model parameters. The corresponding Newton iteration indicated by Equation (47) will thus take on the following form for finding the nontrivial root of Equation (43) when d is very small or actually vanishes:

$$\theta^{(h,n+1)} = \theta^{(h,n)} - [D - V(\theta^{(h,n)})]^{-1} w(\theta^{(h,n)}); \quad n = 0, 1, \dots \quad (51)$$

As has been discussed in the foregoing in conjunction with the algebraic quadratic equation, it would be advantageous to start the Newton iteration shown in Equation (51) with an initial guess $\theta^{(h,0)}$ that would have components analogous to the initial guess $x^{(0)} = -b/a$ for the quadratic equations. It would be further advantageous if the computation of the initial guess $\theta^{(h,0)}$ would not require inversion of matrices. An expression for $\theta^{(h,0)}$ which satisfies these considerations can be derived by setting $d = \mathbf{0}$ in Equation (43) and by considering just the diagonal terms of the matrices involved in this equation, thus obtaining:

$$\theta^{(h,0)} \triangleq (\theta_1^{(h,0)}, \dots, \theta_j^{(h,0)}, \dots, \theta_{TR}^{(h,0)})^\dagger; \quad \theta_j^{(h,0)} = 2D_{jj} / 3M_{jj}^{(j)}; \quad j = 1, \dots, TR. \quad (52)$$

Using the initial guess provided in Equation (52) in Equation (50) yields the following starting expressions for components of the Jacobian matrix $A(\theta)$:

$$A_{kn}(\theta^{(h,0)}) = D_{kn} - \sum_{j=1}^{TR} \theta_j^{(h,0)} [M_{nj}^{(k)} + M_{kj}^{(n)} + M_{kn}^{(j)}]. \quad (53)$$

Furthermore, using the initial guess provided in Equation (52) in Equation (43) yields the following starting expressions for components $w_k(\theta)$ of the vector $w(\theta) \triangleq [w_1(\theta), \dots, w_k(\theta), \dots, w_{TR}(\theta)]^\dagger$:

$$w_k(\theta^{(h,0)}) = u_k(\theta^{(h,0)}) + \sum_{j=1}^{TR} D_{kj} \theta_j^{(h,0)} - d_k, \quad k = 1, \dots, TR; \quad (54)$$

$$\begin{aligned} u_k(\theta^{(h,0)}) &= -\frac{1}{2} \sum_{i=1}^{TR} \sum_{j=1}^{TR} \theta_i^{(h,0)} M_{ij}^{(k)} \theta_j^{(h,0)} - \sum_{i=1}^{TR} \theta_i^{(h,0)} \sum_{j=1}^{TR} M_{kj}^{(i)} \theta_j^{(h,0)} \\ &\cong -\frac{4}{9} \sum_{i=1}^{TR} \sum_{j=1}^{TR} \left(\frac{D_{ii}}{M_{ii}^{(i)}} \right) \left(\frac{D_{jj}}{M_{jj}^{(j)}} \right) \left[\frac{1}{2} M_{ij}^{(k)} + M_{kj}^{(i)} \right], \quad k = 1, \dots, TR. \end{aligned} \quad (55)$$

The impact of the second-order sensitivities through the components $M_{ij}^{(k)}$ of the matrices $\mathbf{M}^{(k)}$ is evident in Equations (53)-(55). The first-iterate $\boldsymbol{\theta}^{(h,1)}$ of the solution of Equation (51) has the following expression:

$$\begin{aligned} \boldsymbol{\theta}^{(h,1)} &= \boldsymbol{\theta}^{(h,0)} - \left(\mathbf{D} - \mathbf{V}^{(h,0)}\right)^{-1} \left(\mathbf{u}^{(h,0)} + \mathbf{D}\boldsymbol{\theta}^{(h,0)} - \mathbf{d}\right); \\ \mathbf{V}^{(h,0)} &\triangleq \mathbf{V}\left(\boldsymbol{\theta}^{(h,0)}\right); \quad \mathbf{u}^{(h,0)} \triangleq \mathbf{u}\left(\boldsymbol{\theta}^{(h,0)}\right). \end{aligned} \tag{56}$$

It is evident from the expressions provided in Equations (53)-(55) that even if $\mathbf{d} = \mathbf{0}$ (i.e., if the vector of deviations between the nominally measured and computed values of the responses vanishes), the starting-values of the components $w_k\left(\boldsymbol{\theta}^{(h,0)}\right) \neq 0$ are non-zero because of the impact of the second-order response sensitivities. In addition, the initial-Jacobian matrix $\mathbf{A}\left(\boldsymbol{\theta}^{(h,0)}\right)$ is non-singular (also because of the impact of the second-order response sensitivities), so the first-iterate $\boldsymbol{\theta}^{(h,1)}$ has a finite non-zero value. All subsequent iterates will also have non-zero values. Thus, using the second-order response sensitivities overcomes the limitation of the first-order predictive modeling methodology 1st-BERRU-PM, as well as the limitation of extant data adjustment and data assimilation procedures, which all fail when the measured response values coincide with the computed response values.

Replacing the solution of Equation (51) into Equations (25) and (26) yields the following expressions for the coordinates of the saddle point $\mathbf{z}^{sp} \triangleq \left(\mathbf{r}^{sp}, \boldsymbol{\alpha}^{sp}\right)^\dagger$ of the function $h(\mathbf{r}, \boldsymbol{\alpha}; \boldsymbol{\theta})$:

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

$$\boldsymbol{\alpha}^{sp} = \boldsymbol{\alpha}^0 + \left(\mathbf{C}_{\alpha r}^e - \mathbf{C}_{\alpha\alpha}^e \mathbf{S}^\dagger\right) \boldsymbol{\theta}^{(h,n+1)}. \tag{58}$$

The best-estimate (optimal) mean values (first-order moments) and correlations (second-order moments) of the posterior distribution of the combined computational model with measured responses and parameters are obtained by using the saddle-point method to evaluate the integrals that define these first- and second-order moments. These best-estimate values will be denoted using the superscript “bed” to indicate that these “best-estimate” values arise from deterministically incorporated/assimilated computed responses.

Thus, the predicted “best-estimated values with deterministically incorporated computed responses” for the predicted mean values of responses and parameters, respectively, have the following expressions:

$$\mathbf{r}^{bed} \triangleq \int_{D_z} \mathbf{r} p(\mathbf{z}) d\mathbf{z} = \mathbf{r}^{sp} = \mathbf{r}^e + \left(\mathbf{C}_{rr}^e - \mathbf{C}_{r\alpha}^e \mathbf{S}^\dagger\right) \boldsymbol{\theta}^{(h,n+1)}; \tag{59}$$

$$\boldsymbol{\alpha}^{bed} \triangleq \int_{D_z} \boldsymbol{\alpha} p(\mathbf{z}) d\mathbf{z} = \boldsymbol{\alpha}^{sp} = \boldsymbol{\alpha}^0 + \left(\mathbf{C}_{\alpha r}^e - \mathbf{C}_{\alpha\alpha}^e \mathbf{S}^\dagger\right) \boldsymbol{\theta}^{(h,n+1)}. \tag{60}$$

The expression of the “best-estimate with deterministically incorporated computed responses” covariance matrix for the responses, denoted as \mathbf{C}_{rr}^{bed} , is obtained from its definition in conjunction with Equation (59), which yields:

$$\begin{aligned}
 C_{rr}^{bed} &\triangleq \int_{D_z} (\mathbf{r} - \mathbf{r}^{bed})(\mathbf{r} - \mathbf{r}^{bed})^\dagger p(\mathbf{z}) d\mathbf{z} \\
 &= \int_{D_z} \left[\mathbf{r} - \mathbf{r}^e - (\mathbf{C}_{rr}^e - \mathbf{C}_{r\alpha}^e \mathbf{S}^\dagger) \boldsymbol{\theta}^{(h,n+1)} \right] \left[\mathbf{r} - \mathbf{r}^e - (\mathbf{C}_{rr}^e - \mathbf{C}_{r\alpha}^e \mathbf{S}^\dagger) \boldsymbol{\theta}^{(h,n+1)} \right]^\dagger p(\mathbf{z}) d\mathbf{z}.
 \end{aligned} \tag{61}$$

The expression of the “best-estimate with deterministically incorporated computed responses” covariance matrix for the parameters, denoted as $C_{\alpha\alpha}^{bed}$, is obtained from its definition in conjunction with Equation (60), which yields:

$$\begin{aligned}
 C_{\alpha\alpha}^{bed} &\triangleq \int_{D_z} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^{bed})(\boldsymbol{\alpha} - \boldsymbol{\alpha}^{bed})^\dagger p(\mathbf{z}) d\mathbf{z} \\
 &= \int_{D_z} \left[\boldsymbol{\alpha} - \boldsymbol{\alpha}^0 - (\mathbf{C}_{\alpha r}^e - \mathbf{C}_{\alpha\alpha}^e \mathbf{S}^\dagger) \boldsymbol{\theta}^{(h,n+1)} \right] \left[\boldsymbol{\alpha} - \boldsymbol{\alpha}^0 - (\mathbf{C}_{\alpha r}^e - \mathbf{C}_{\alpha\alpha}^e \mathbf{S}^\dagger) \boldsymbol{\theta}^{(h,n+1)} \right]^\dagger p(\mathbf{z}) d\mathbf{z}.
 \end{aligned} \tag{62}$$

The expression of the “best-estimate with deterministically incorporated computed responses” correlation matrix among parameters and responses, denoted as $C_{\alpha r}^{bed}$, is obtained by using its definition in conjunction with Equations (59) and (60), which yields:

$$\begin{aligned}
 C_{\alpha r}^{bed} &\triangleq \int_{D_z} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^{bed})(\mathbf{r} - \mathbf{r}^{bed})^\dagger p(\mathbf{z}) d\mathbf{z} \\
 &= \int_{D_z} \left[\boldsymbol{\alpha} - \boldsymbol{\alpha}^0 - (\mathbf{C}_{\alpha r}^e - \mathbf{C}_{\alpha\alpha}^e \mathbf{S}^\dagger) \boldsymbol{\theta}^{(h,n+1)} \right] \left[\mathbf{r} - \mathbf{r}^e - (\mathbf{C}_{rr}^e - \mathbf{C}_{r\alpha}^e \mathbf{S}^\dagger) \boldsymbol{\theta}^{(h,n+1)} \right]^\dagger p(\mathbf{z}) d\mathbf{z}.
 \end{aligned} \tag{63}$$

The impact of the second-order response sensitivities on the mean values and correlation/covariance matrices can be assessed already by using the first iterate solution, $\boldsymbol{\theta}^{(h,1)}$, in Equations (59)-(63). Using the first-iterate, $\boldsymbol{\theta}^{(h,1)}$, in Equations (59) and (60), respectively, yields the following expressions for the “first-iterate best-estimate” values for the responses and parameters, respectively:

$$\mathbf{r}^{(1)bed} = \mathbf{r}^e + (\mathbf{C}_{rr}^e - \mathbf{C}_{r\alpha}^e \mathbf{S}^\dagger) \left[\boldsymbol{\theta}^{(h,0)} - (\mathbf{D} - \mathbf{V}^{(h,0)})^{-1} (\mathbf{u}^{(h,0)} + \mathbf{D}\boldsymbol{\theta}^{(h,0)} - \mathbf{d}) \right]; \tag{64}$$

$$\boldsymbol{\alpha}^{(1)bed} = \boldsymbol{\alpha}^0 + (\mathbf{C}_{\alpha r}^e - \mathbf{C}_{\alpha\alpha}^e \mathbf{S}^\dagger) \left[\boldsymbol{\theta}^{(h,0)} - (\mathbf{D} - \mathbf{V}^{(h,0)})^{-1} (\mathbf{u}^{(h,0)} + \mathbf{D}\boldsymbol{\theta}^{(h,0)} - \mathbf{d}) \right]. \tag{65}$$

As the expressions obtained in Equations (64) and (65) indicate, combining experimental and computational information modifies the correlations among the parameters through couplings introduced by the sensitivities of the participating responses. The expressions obtained in Equations (64) and (65) possess the following new features by comparisons to their counterparts provided by extant data assimilation and data adjustment procedures:

1) If the second-order sensitivities are neglected, then the expressions in (64) and (65) take on the following simplified forms:

$$\text{no 2}^{nd}\text{-order sensitivities : } \mathbf{r}^{(1)bed} = \mathbf{r}^e + (\mathbf{C}_{rr}^e - \mathbf{C}_{r\alpha}^e \mathbf{S}^\dagger) \mathbf{D}^{-1} \mathbf{d}; \tag{66}$$

$$\text{no 2}^{nd}\text{-order sensitivities : } \boldsymbol{\alpha}^{(1)bed} = \boldsymbol{\alpha}^0 + (\mathbf{C}_{\alpha r}^e - \mathbf{C}_{\alpha\alpha}^e \mathbf{S}^\dagger) \mathbf{D}^{-1} \mathbf{d}. \tag{67}$$

As expected, the expressions in Equations (66) and (67) are identical to the

expressions provided by the 1st-BERRU-PM [14] methodology and therefore possess the same strengths (easy to implement) and weaknesses (trivial results, no calibration if $\mathbf{d} = \mathbf{0}$) as the results produced by the 1st-BERRU-PM methodology.

2) If the vector of deviations between the nominally measured and computed values of the responses vanishes, *i.e.*, if $\mathbf{d} = \mathbf{0}$, the expressions provided in Equations (64) and (65) take on the following simplified forms:

if $\mathbf{d} = \mathbf{0}$:

$$\mathbf{r}^{(1)bed}(\mathbf{d} = \mathbf{0}) = \mathbf{r}^e + (\mathbf{C}_{rr}^e - \mathbf{C}_{r\alpha}^e \mathbf{S}^\dagger) \left[\boldsymbol{\theta}^{(h,0)} - (\mathbf{D} - \mathbf{V}^{(h,0)})^{-1} (\mathbf{u}^{(h,0)} + \mathbf{D}\boldsymbol{\theta}^{(h,0)}) \right] \neq \mathbf{0} \quad (68)$$

if $\mathbf{d} = \mathbf{0}$:

$$\boldsymbol{\alpha}^{(1)bed}(\mathbf{d} = \mathbf{0}) = \boldsymbol{\alpha}^0 + (\mathbf{C}_{\alpha r}^e - \mathbf{C}_{\alpha\alpha}^e \mathbf{S}^\dagger) \left[\boldsymbol{\theta}^{(h,0)} - (\mathbf{D} - \mathbf{V}^{(h,0)})^{-1} (\mathbf{u}^{(h,0)} + \mathbf{D}\boldsymbol{\theta}^{(h,0)}) \right] \neq \mathbf{0} \quad (69)$$

As evidenced by comparing Equations (68) and (69) with Equations (66) and (67), respectively, the contributions of the second-order sensitivities provide the fundamental difference when $\mathbf{d} = \mathbf{0}$ between non-calibration [*i.e.*,

$\mathbf{r}^{(1)bed}(\mathbf{d} = \mathbf{0}) = \mathbf{r}^e$ and $\boldsymbol{\alpha}^{(1)bed}(\mathbf{d} = \mathbf{0}) = \boldsymbol{\alpha}^0$] in the case for data adjustment, data assimilation and 1st-BERRU-PM methodologies, cf. Equations (66) and (67), respectively, versus the calibration $\mathbf{r}^{(1)bed}(\mathbf{d} = \mathbf{0}) \neq \mathbf{r}^e$ and $\boldsymbol{\alpha}^{(1)bed}(\mathbf{d} = \mathbf{0}) \neq \boldsymbol{\alpha}^0$ provided by the 2nd-BERRU-PM expressions obtained in Equations (68) and (69).

3) The components of the matrix $\mathbf{V}^{(h,0)}$ are usually smaller (since they contain higher-order quantities) than the corresponding components of the matrix \mathbf{D} , so the contributions provided by the term $\mathbf{V}^{(h,0)}$ can be considered to be “corrections” to the components of the matrix \mathbf{D} . If the contributions of the matrix $\mathbf{V}^{(h,0)}$ are neglected by comparison to the corresponding components of the matrix \mathbf{D} , the expressions provided in Equations (64) and (65) reduce to the following expressions:

$$\text{if } \mathbf{V}^{(h,0)} = \mathbf{0}: \mathbf{r}^{(1)bed}(\mathbf{V}^{(h,0)} = \mathbf{0}) = \mathbf{r}^e + (\mathbf{C}_{rr}^e - \mathbf{C}_{r\alpha}^e \mathbf{S}^\dagger) \mathbf{D}^{-1} (\mathbf{d} - \mathbf{u}^{(h,0)}); \quad (70)$$

$$\text{if } \mathbf{V}^{(h,0)} = \mathbf{0}: \boldsymbol{\alpha}^{(1)bed}(\mathbf{V}^{(h,0)} = \mathbf{0}) = \boldsymbol{\alpha}^0 + (\mathbf{C}_{\alpha r}^e - \mathbf{C}_{\alpha\alpha}^e \mathbf{S}^\dagger) \mathbf{D}^{-1} (\mathbf{d} - \mathbf{u}^{(h,0)}). \quad (71)$$

The above expressions underscore, through the term $\mathbf{u}^{(h,0)}$, the fundamental importance of including second-order sensitivities, since they provide the basic contribution to maintaining the effects of calibration for both responses and parameters, even when $\mathbf{d} = \mathbf{0}$, *i.e.*, when the values of the measured responses coincide with the corresponding values of the computed responses. This essential contribution stemming from the second-order responses sensitivities is clearly highlighted by comparing the expressions obtained in Equations (70) and (71) with the corresponding first-order (1st-BERRU-PM) expressions provided in Equations (66) and (67), respectively.

Using the first-iterate, $\boldsymbol{\theta}^{(h,1)}$, in Equation (61) yields the following expression for the “first-iterate best-estimate” covariance matrix $\mathbf{C}_{rr}^{(1)bed}$ for the responses:

$$\begin{aligned}
 \mathbf{C}_{rr}^{(1)bed} &= \int_{D_z} \left\{ \mathbf{r} - \mathbf{r}^e - (\mathbf{C}_{rr}^e - \mathbf{C}_{r\alpha}^e \mathbf{S}^\dagger) \left[\boldsymbol{\theta}^{(h,0)} - (\mathbf{D} - \mathbf{V}^{(h,0)})^{-1} (\mathbf{u}^{(h,0)} + \mathbf{D}\boldsymbol{\theta}^{(h,0)} - \mathbf{d}) \right] \right\} \\
 &\times \left\{ \mathbf{r} - \mathbf{r}^e - (\mathbf{C}_{rr}^e - \mathbf{C}_{r\alpha}^e \mathbf{S}^\dagger) \left[\boldsymbol{\theta}^{(h,0)} - (\mathbf{D} - \mathbf{V}^{(h,0)})^{-1} (\mathbf{u}^{(h,0)} + \mathbf{D}\boldsymbol{\theta}^{(h,0)} - \mathbf{d}) \right] \right\}^\dagger p(\mathbf{z}) d\mathbf{z} \\
 &= \left\langle \left[\mathbf{r} - \mathbf{r}^e + \mathbf{X}_r \boldsymbol{\theta}^{(h,0)} - \mathbf{Y}_r \mathbf{d} + \mathbf{Y}_r \mathbf{u}^{(h,0)} \right] \left[\mathbf{r} - \mathbf{r}^e + \mathbf{X}_r \boldsymbol{\theta}^{(h,0)} - \mathbf{Y}_r \mathbf{d} + \mathbf{Y}_r \mathbf{u}^{(h,0)} \right]^\dagger \right\rangle.
 \end{aligned} \tag{72}$$

The following definitions were used in Equation (72):

$$\langle \bullet \rangle \triangleq \int_{D_z} (\bullet) p(\mathbf{z}) d\mathbf{z}, \tag{73}$$

$$\begin{aligned}
 \mathbf{C}_{rd}^{(1)} &\triangleq \left\langle (\mathbf{r} - \mathbf{r}^e) \mathbf{d}^\dagger \right\rangle = \left\langle (\mathbf{r} - \mathbf{r}^e) \left[\mathbf{r} - \mathbf{r}^e - \mathbf{S}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) - p(\boldsymbol{\alpha}; \boldsymbol{\alpha}^0) \right]^\dagger \right\rangle \\
 &\cong \mathbf{C}_{rr}^e - \mathbf{C}_{r\alpha}^e \mathbf{S}^\dagger;
 \end{aligned} \tag{74}$$

$$\mathbf{X}_r \triangleq \mathbf{C}_{rd}^{(1)} \left[(\mathbf{D} - \mathbf{V}^{(h,0)})^{-1} \mathbf{D} - \mathbf{I} \right]; \tag{75}$$

$$\mathbf{Y}_r \triangleq \mathbf{C}_{rd}^{(1)} (\mathbf{D} - \mathbf{V}^{(h,0)})^{-1}. \tag{76}$$

Expanding the integrand in the second equality in Equation (72) yields the following expression for the “first-iterate best-estimate” response covariance matrix $\mathbf{C}_{rr}^{(1)bed}$:

$$\begin{aligned}
 \mathbf{C}_{rr}^{(1)bed} &= \mathbf{C}_{rr}^e - \mathbf{C}_{rd}^{(1)} \mathbf{Y}_r^\dagger + \mathbf{X}_r \boldsymbol{\theta}^{(h,0)} \left(\boldsymbol{\theta}^{(h,0)} \right)^\dagger \mathbf{X}_r^\dagger - \mathbf{X}_r \boldsymbol{\theta}^{(h,0)} \bar{\mathbf{d}}^\dagger \mathbf{Y}_r^\dagger + \mathbf{X}_r \boldsymbol{\theta}^{(h,0)} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_r^\dagger \\
 &\quad - \mathbf{Y}_r \left(\mathbf{C}_{rd}^{(1)} \right)^\dagger - \mathbf{Y}_r \bar{\mathbf{d}} \left(\boldsymbol{\theta}^{(h,0)} \right)^\dagger \mathbf{X}_r^\dagger + \mathbf{Y}_r \mathbf{D} \mathbf{Y}_r^\dagger - \mathbf{Y}_r \bar{\mathbf{d}} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_r^\dagger \\
 &\quad + \mathbf{Y}_r \mathbf{u}^{(h,0)} \left(\boldsymbol{\theta}^{(h,0)} \right)^\dagger \mathbf{X}_r^\dagger - \mathbf{Y}_r \mathbf{u}^{(h,0)} \bar{\mathbf{d}}^\dagger \mathbf{Y}_r^\dagger + \mathbf{Y}_r \mathbf{u}^{(h,0)} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_r^\dagger.
 \end{aligned} \tag{77}$$

Using the first-iterate, $\boldsymbol{\theta}^{(h,1)}$, in Equation (62) and evaluating the resulting integrals yields the following expression for the “first-iterate best-estimate” covariance matrix $\mathbf{C}_{\alpha\alpha}^{(1)be}$ for the model parameters:

$$\begin{aligned}
 \mathbf{C}_{\alpha\alpha}^{(1)bed} &= \left\langle \left[\boldsymbol{\alpha} - \boldsymbol{\alpha}^0 + \mathbf{X}_\alpha \boldsymbol{\theta}^{(h,0)} - \mathbf{Y}_\alpha \mathbf{d} + \mathbf{Y}_\alpha \mathbf{u}^{(h,0)} \right] \left[\boldsymbol{\alpha} - \boldsymbol{\alpha}^0 + \mathbf{X}_\alpha \boldsymbol{\theta}^{(h,0)} - \mathbf{Y}_\alpha \mathbf{d} + \mathbf{Y}_\alpha \mathbf{u}^{(h,0)} \right]^\dagger \right\rangle \\
 &= \mathbf{C}_{\alpha\alpha}^e - \mathbf{C}_{\alpha d}^{(1)} \mathbf{Y}_\alpha^\dagger + \mathbf{X}_\alpha \boldsymbol{\theta}^{(h,0)} \left(\boldsymbol{\theta}^{(h,0)} \right)^\dagger \mathbf{X}_\alpha^\dagger - \mathbf{X}_\alpha \boldsymbol{\theta}^{(h,0)} \bar{\mathbf{d}}^\dagger \mathbf{Y}_\alpha^\dagger + \mathbf{X}_\alpha \boldsymbol{\theta}^{(h,0)} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_\alpha^\dagger \\
 &\quad - \mathbf{Y}_\alpha \left(\mathbf{C}_{\alpha d}^{(1)} \right)^\dagger - \mathbf{Y}_\alpha \bar{\mathbf{d}} \left(\boldsymbol{\theta}^{(h,0)} \right)^\dagger \mathbf{X}_\alpha^\dagger + \mathbf{Y}_\alpha \mathbf{D} \mathbf{Y}_\alpha^\dagger - \mathbf{Y}_\alpha \bar{\mathbf{d}} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_\alpha^\dagger \\
 &\quad + \mathbf{Y}_\alpha \mathbf{u}^{(h,0)} \left(\boldsymbol{\theta}^{(h,0)} \right)^\dagger \mathbf{X}_\alpha^\dagger - \mathbf{Y}_\alpha \mathbf{u}^{(h,0)} \bar{\mathbf{d}}^\dagger \mathbf{Y}_\alpha^\dagger + \mathbf{Y}_\alpha \mathbf{u}^{(h,0)} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_\alpha^\dagger.
 \end{aligned} \tag{78}$$

The following definitions were used in Equation (78):

$$\begin{aligned}
 \mathbf{C}_{\alpha d}^{(1)} &\triangleq \left\langle (\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) \mathbf{d}^\dagger \right\rangle = \left\langle (\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) \left[\mathbf{r} - \mathbf{r}^e - \mathbf{S}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0) - p(\boldsymbol{\alpha}; \boldsymbol{\alpha}^0) \right]^\dagger \right\rangle \\
 &\cong \mathbf{C}_{\alpha r}^e - \mathbf{C}_{\alpha\alpha}^e \mathbf{S}^\dagger;
 \end{aligned} \tag{79}$$

$$\mathbf{X}_\alpha \triangleq \mathbf{C}_{\alpha d}^{(1)} \left[(\mathbf{D} - \mathbf{V}^{(h,0)})^{-1} \mathbf{D} - \mathbf{I} \right]; \tag{80}$$

$$Y_\alpha \triangleq C_{\alpha d}^{(1)}(D - V^{(h,0)})^{-1}. \tag{81}$$

Using the first-iterate, $\theta^{(h,1)}$, in Equation (63) and evaluating the resulting integrals yields the following expression for the “first-iterate best-estimate” correlation matrix $C_{\alpha r}^{(1)bed}$ for the parameters and responses:

$$\begin{aligned} C_{\alpha r}^{(1)bed} &= \left\langle \left[\alpha - \alpha^0 + X_\alpha \theta^{(h,0)} - Y_\alpha d + Y_\alpha u^{(h,0)} \right] \left[r - r^e + X_r \theta^{(h,0)} - Y_r d + Y_r u^{(h,0)} \right]^\dagger \right\rangle \\ &= C_{\alpha r}^e - C_{\alpha d}^{(1)} Y_r^\dagger + X_\alpha \theta^{(h,0)} \left(\theta^{(h,0)} \right)^\dagger X_r^\dagger - X_\alpha \theta^{(h,0)} \bar{d}^\dagger Y_r^\dagger + X_\alpha \theta^{(h,0)} \left(u^{(h,0)} \right)^\dagger Y_r^\dagger \\ &\quad - Y_\alpha \left(C_{rd}^{(1)} \right)^\dagger - Y_\alpha \bar{d} \left(\theta^{(h,0)} \right)^\dagger X_r^\dagger + Y_\alpha D Y_r^\dagger - Y_\alpha \bar{d} \left(u^{(h,0)} \right)^\dagger Y_r^\dagger \\ &\quad + Y_\alpha u^{(h,0)} \left(\theta^{(h,0)} \right)^\dagger X_r^\dagger - Y_\alpha u^{(h,0)} \bar{d}^\dagger Y_r^\dagger + Y_\alpha u^{(h,0)} \left(u^{(h,0)} \right)^\dagger Y_r^\dagger. \end{aligned} \tag{82}$$

As the expressions obtained in Equations (77), (78) and (82) indicate, combining experimental and computational information modifies the correlations among the parameters through couplings introduced by the sensitivities of the participating responses.

The expression obtained in Equations (77), (78) and (82) for $C_{rr}^{(1)bed}$, $C_{\alpha\alpha}^{(1)bed}$ and $C_{\alpha r}^{(1)bed}$, respectively, will now be analyzed below by considering the same simplifying conditions as were considered for the best-estimate response and parameter mean-values $r^{(1)bed}$ and $\alpha^{(1)bed}$, respectively, as follows:

1) If all of the second-order sensitivities are neglected, then the following simplifications occur: $X_r = X_\alpha = \mathbf{0}$, $Y_r = C_{rd}^{(1)} D^{-1}$ and $Y_\alpha = C_{\alpha d}^{(1)} D^{-1}$. Therefore, the expressions in Equations (77), (78) and (82) take on the following simplified forms:

$$\text{no 2}^{nd}\text{-order sensitivities: } C_{rr}^{(1)bed} = C_{rr}^e - C_{rd}^{(1)} D^{-1} C_{rd}^{(1)\dagger}. \tag{83}$$

$$\text{no 2}^{nd}\text{-order sensitivities: } C_{\alpha\alpha}^{(1)bed} = C_{\alpha\alpha}^e - C_{\alpha d}^{(1)} D^{-1} C_{\alpha d}^{(1)\dagger}. \tag{84}$$

$$\text{no 2}^{nd}\text{-order sensitivities: } C_{\alpha r}^{(1)bed} = C_{\alpha r}^e - C_{\alpha d}^{(1)} D^{-1} C_{rd}^{(1)\dagger}. \tag{85}$$

As expected, the expressions obtained in Equation (83)-(85) are identical to the expressions provided by the first-order predictive modeling methodology 1st-BERRU-PM [14] and therefore possess the same strengths (easy to implement) and weaknesses (*i.e.*, trivial results and no calibration if $d = \mathbf{0}$) as 1st-BERRU-PM.

2) If the vector of deviations between the nominally measured and computed values of the responses vanishes, *i.e.*, if $d = \mathbf{0}$, the expressions obtained in Equations (77) for $C_{rr}^{(1)bed}$, $C_{\alpha\alpha}^{(1)bed}$ and $C_{\alpha r}^{(1)bed}$, respectively, will take on the following simplified forms:

$$\begin{aligned} \text{if } d = \mathbf{0}: \quad C_{rr}^{(1)bed} &= C_{rr}^e - C_{rd}^{(1)} Y_r^\dagger + X_r \theta^{(h,0)} \left(\theta^{(h,0)} \right)^\dagger X_r^\dagger + X_r \theta^{(h,0)} \left(u^{(h,0)} \right)^\dagger Y_r^\dagger \\ &\quad - Y_r \left(C_{rd}^{(1)} \right)^\dagger + Y_r D Y_r^\dagger + Y_r u^{(h,0)} \left(\theta^{(h,0)} \right)^\dagger X_r^\dagger + Y_r u^{(h,0)} \left(u^{(h,0)} \right)^\dagger Y_r^\dagger; \end{aligned} \tag{86}$$

$$\begin{aligned} \text{if } \mathbf{d} = \mathbf{0}: \mathbf{C}_{\alpha\alpha}^{(1)bed} &= \mathbf{C}_{\alpha\alpha}^e - \mathbf{C}_{\alpha d}^{(1)} \mathbf{Y}_\alpha^\dagger + \mathbf{X}_\alpha \boldsymbol{\theta}^{(h,0)} \left(\boldsymbol{\theta}^{(h,0)} \right)^\dagger \mathbf{X}_\alpha^\dagger + \mathbf{X}_\alpha \boldsymbol{\theta}^{(h,0)} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_\alpha^\dagger \\ &- \mathbf{Y}_\alpha \left(\mathbf{C}_{\alpha d}^{(1)} \right)^\dagger + \mathbf{Y}_\alpha \mathbf{D} \mathbf{Y}_\alpha^\dagger + \mathbf{Y}_\alpha \mathbf{u}^{(h,0)} \left(\boldsymbol{\theta}^{(h,0)} \right)^\dagger \mathbf{X}_\alpha^\dagger + \mathbf{Y}_\alpha \mathbf{u}^{(h,0)} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_\alpha^\dagger; \end{aligned} \tag{87}$$

$$\begin{aligned} \text{if } \mathbf{d} = \mathbf{0}: \mathbf{C}_{\alpha r}^{(1)bed} &= \mathbf{C}_{\alpha r}^e - \mathbf{C}_{\alpha d}^{(1)} \mathbf{Y}_r^\dagger + \mathbf{X}_\alpha \boldsymbol{\theta}^{(h,0)} \left(\boldsymbol{\theta}^{(h,0)} \right)^\dagger \mathbf{X}_r^\dagger + \mathbf{X}_\alpha \boldsymbol{\theta}^{(h,0)} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_r^\dagger \\ &- \mathbf{Y}_\alpha \left(\mathbf{C}_{rd}^{(1)} \right)^\dagger + \mathbf{Y}_\alpha \mathbf{D} \mathbf{Y}_r^\dagger + \mathbf{Y}_\alpha \mathbf{u}^{(h,0)} \left(\boldsymbol{\theta}^{(h,0)} \right)^\dagger \mathbf{X}_r^\dagger + \mathbf{Y}_\alpha \mathbf{u}^{(h,0)} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_r^\dagger. \end{aligned} \tag{88}$$

If the contributions of the second-order sensitivities are neglected in the expressions obtained in Equations (86)-(88), these expressions will reduce to the first-order expressions shown in Equations (83)-(85).

3) If the contributions of the matrix $\mathbf{V}^{(h,0)}$ are neglected by comparison to the corresponding components of the matrix \mathbf{D} , then the following simplifications occur: $\mathbf{X}_r = \mathbf{X}_\alpha = \mathbf{0}$, $\mathbf{Y}_r = \mathbf{C}_{rd}^{(1)} \mathbf{D}^{-1}$ and $\mathbf{Y}_\alpha = \mathbf{C}_{\alpha d}^{(1)} \mathbf{D}^{-1}$. Consequently, the expressions provided in Equations (77), (78) and (82) will reduce to the following forms:

$$\begin{aligned} \text{if } \mathbf{V}^{(h,0)} = \mathbf{0}: \mathbf{C}_{rr}^{(1)bed} \left(\mathbf{V}^{(h,0)} = \mathbf{0} \right) &= \mathbf{C}_{rr}^e - \mathbf{C}_{rd}^{(1)} \mathbf{D}^{-1} \left(\mathbf{C}_{rd}^{(1)} \right)^\dagger \\ &- \mathbf{Y}_r \bar{\mathbf{d}} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_r^\dagger - \mathbf{Y}_r \mathbf{u}^{(h,0)} \bar{\mathbf{d}}^\dagger \mathbf{Y}_r^\dagger + \mathbf{Y}_r \mathbf{u}^{(h,0)} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_r^\dagger; \end{aligned} \tag{89}$$

$$\begin{aligned} \text{if } \mathbf{V}^{(h,0)} = \mathbf{0}: \mathbf{C}_{\alpha\alpha}^{(1)bed} \left(\mathbf{V}^{(h,0)} = \mathbf{0} \right) &= \mathbf{C}_{\alpha\alpha}^e - \mathbf{C}_{\alpha d}^{(1)} \mathbf{D}^{-1} \left(\mathbf{C}_{\alpha d}^{(1)} \right)^\dagger \\ &- \mathbf{Y}_\alpha \bar{\mathbf{d}} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_\alpha^\dagger - \mathbf{Y}_\alpha \mathbf{u}^{(h,0)} \bar{\mathbf{d}}^\dagger \mathbf{Y}_\alpha^\dagger + \mathbf{Y}_\alpha \mathbf{u}^{(h,0)} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_\alpha^\dagger; \end{aligned} \tag{90}$$

$$\begin{aligned} \text{if } \mathbf{V}^{(h,0)} = \mathbf{0}: \mathbf{C}_{\alpha r}^{(1)bed} \left(\mathbf{V}^{(h,0)} = \mathbf{0} \right) &= \mathbf{C}_{\alpha r}^e - \mathbf{C}_{\alpha d}^{(1)} \mathbf{D}^{-1} \left(\mathbf{C}_{rd}^{(1)} \right)^\dagger \\ &- \mathbf{Y}_\alpha \bar{\mathbf{d}} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_r^\dagger - \mathbf{Y}_\alpha \mathbf{u}^{(h,0)} \bar{\mathbf{d}}^\dagger \mathbf{Y}_r^\dagger + \mathbf{Y}_\alpha \mathbf{u}^{(h,0)} \left(\mathbf{u}^{(h,0)} \right)^\dagger \mathbf{Y}_r^\dagger. \end{aligned} \tag{91}$$

The above expressions underscore, through the terms involving the vector $\mathbf{u}^{(h,0)}$, the fundamental importance of including second-order sensitivities, which provide non-zero contributions even when $\bar{\mathbf{d}} = \mathbf{0}$.

It is important to emphasize that the posterior distribution of responses and parameters is non-Gaussian since the “best estimate” triple correlations among the best-estimate responses and parameters are nonzero even if the experimentally measured responses and parameters were normally distributed initially. This fact can readily be shown by considering Equations (64) and (65), which can be written as follows:

$$\begin{aligned} \mathbf{r}^{(1)bed} &= \mathbf{r}^e + \mathbf{r}^{(0)}; \\ \mathbf{r}^{(0)} &\triangleq \left(\mathbf{C}_{rr}^e - \mathbf{C}_{r\alpha}^e \mathbf{S}^\dagger \right) \left[\boldsymbol{\theta}^{(h,0)} - \left(\mathbf{D} - \mathbf{V}^{(h,0)} \right)^{-1} \left(\mathbf{u}^{(h,0)} + \mathbf{D} \boldsymbol{\theta}^{(h,0)} - \mathbf{d} \right) \right]; \end{aligned} \tag{92}$$

$$\begin{aligned} \boldsymbol{\alpha}^{(1)bed} &= \boldsymbol{\alpha}^0 + \boldsymbol{\beta}^{(0)}; \\ \boldsymbol{\beta}^{(0)} &\triangleq \left(\mathbf{C}_{\alpha r}^e - \mathbf{C}_{\alpha\alpha}^e \mathbf{S}^\dagger \right) \left[\boldsymbol{\theta}^{(h,0)} - \left(\mathbf{D} - \mathbf{V}^{(h,0)} \right)^{-1} \left(\mathbf{u}^{(h,0)} + \mathbf{D} \boldsymbol{\theta}^{(h,0)} - \mathbf{d} \right) \right]. \end{aligned} \tag{93}$$

Using Equations (92) and (93) in conjunction with the definition of triple correlations for multivariate random quantities yields the following “lowest-

order estimate” for the best-estimate triple correlations among responses and parameters, which denoted below as $T_{rr,ijk}^{(1)bed}$ and $T_{\alpha\alpha,ijk}^{(1)bed}$, respectively:

$$\begin{aligned} T_{rr,ijk}^{(1)bed} &\triangleq \left\langle \left(r_i - r_i^e - r_i^{(0)} \right) \left(r_j - r_j^e - r_j^{(0)} \right) \left(r_k - r_k^e - r_k^{(0)} \right) \right\rangle \\ &= E \left[\left(r_i - r_i^e \right) \left(r_j - r_j^e \right) \left(r_k - r_k^e \right) \right] - r_i^{(0)} \operatorname{cov} \left(r_j, r_k \right) - r_j^{(0)} \operatorname{cov} \left(r_i, r_k \right) \\ &\quad - r_k^{(0)} \operatorname{cov} \left(r_i, r_j \right) - r_i^{(0)} r_j^{(0)} r_k^{(0)}; \quad i, j, k = 1, \dots, TR; \end{aligned} \quad (94)$$

$$\begin{aligned} T_{\alpha\alpha,ijk}^{(1)bed} &\triangleq \left\langle \left(\alpha_i - \alpha_i^0 - \beta_i^{(0)} \right) \left(\alpha_j - \alpha_j^0 - \beta_j^{(0)} \right) \left(\alpha_k - \alpha_k^0 - \beta_k^{(0)} \right) \right\rangle \\ &= E \left[\left(\alpha_i - \alpha_i^0 \right) \left(\alpha_j - \alpha_j^0 \right) \left(\alpha_k - \alpha_k^0 \right) \right] - \beta_i^{(0)} \operatorname{cov} \left(\alpha_j, \alpha_k \right) \\ &\quad - \beta_j^{(0)} \operatorname{cov} \left(\alpha_i, \alpha_k \right) - \beta_k^{(0)} \operatorname{cov} \left(\alpha_i, \alpha_j \right) - \beta_i^{(0)} \beta_j^{(0)} \beta_k^{(0)}; \quad i, j, k = 1, \dots, TP. \end{aligned} \quad (95)$$

If the responses are normally distributed initially, then the initial triple correlations are all zero, *i.e.*, $E \left[\left(\alpha_i - \alpha_i^0 \right) \left(\alpha_j - \alpha_j^0 \right) \left(\alpha_k - \alpha_k^0 \right) \right] = 0$ for all $i, j, k = 1, \dots, TR$. Nevertheless, as indicated in Equation (94), the “best-estimate response triple correlations” $T_{rr,ijk}^{(1)bed} \neq 0$ are non-zero even in such a case, unless there is “no calibration,” *i.e.*, if $r_i^{(0)} = r_j^{(0)} = r_k^{(0)} = 0$ for all $i, j, k = 1, \dots, TR$. Similarly, if the parameters are normally distributed initially, then the initial triple correlations are all zero, *i.e.*, $E \left[\left(r_i - r_i^e \right) \left(r_j - r_j^e \right) \left(r_k - r_k^e \right) \right] = 0$ for all $i, j, k = 1, \dots, TP$. Nevertheless, as indicated in Equation (95), the “best-estimate parameter triple correlations” $T_{\alpha\alpha,ijk}^{(1)bed} \neq 0$ are non-zero even in such a case, unless there is “no calibration,” *i.e.*, if $\alpha_i^{(0)} = \alpha_j^{(0)} = \alpha_k^{(0)} = 0$ for all $i, j, k = 1, \dots, TP$.

4. Discussion and Conclusions

This work has presented the 2nd-BERRU-PMD methodology, where: 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 letter “D” indicates “deterministic,” referring to the deterministic inclusion of the computational model responses. The 2nd-BERRU-PMD methodology is fundamentally based on the maximum entropy (MaxEnt) principle. This principle is in contradistinction to the fundamental principle that underlies the extant data assimilation and/or adjustment procedures which minimize in a least-square sense a subjective user-defined functional which is meant to represent the discrepancies between measured and computed model responses. It has shown that the 2nd-BERRU-PMD methodology includes all of the previous results produced by the first-order BERRU-PM methodology, generalizing and extending current data assimilation and/or data adjustment procedures while overcoming the fundamental limitations of these procedures. Since the framework of the 2nd-BERRU-PMD methodology comprises the combined phase-space of parameters and responses, the expressions obtained for the predicted best-estimate responses and calibrated parameters can be used for solving both forward/direct and inverse problems.

The accompanying work [19] will present the alternative framework for de-

veloping the “second-order MaxEnt predictive modelling methodology” by incorporating the computed model responses probabilistically (as opposed to “deterministically”). This alternative methodology [19] will be designated by the acronym 2nd-BERRU-PMP, where the last letter (“P”) in the acronym indicates “probabilistic” inclusion of the computational model responses. It will be shown that although both the 2nd-BERRU-PMD and the 2nd-BERRU-PMP frameworks predict analogous 2nd-order expressions for the best-estimate posterior first- and second-order moments, the respective expressions for the predicted responses, calibrated predicted parameters and their predicted uncertainties (covariances) are not identical to each other. The additional considerations that make the corresponding expressions predicted by the 2nd-BERRU-PMD and the 2nd-BERRU-PMP become identical will also be analyzed in the accompanying Part II [19].

Conflicts of Interest

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

References

- [1] Sasaki, Y.K. (1955) A Fundamental Study of the Numerical Prediction Based on the Variational Principle. *Journal of the Meteorological Society of Japan*, **33**, 262-275. https://doi.org/10.2151/jmsj1923.33.6_262
- [2] Sasaki, Y.K. (1958) An Objective Analysis Based on the Variational Method. *Journal of the Meteorological Society of Japan*, **36**, 77-88. https://doi.org/10.2151/jmsj1923.36.3_77
- [3] Eckart, C. (1960) Hydrodynamics of Ocean and Atmosphere. Pergamon Press, Oxford.
- [4] Humi, I., Wagschal, J.J. and Yeivin, Y. (1964) Multi-Group Constants from Integral Data. *Proceedings of the 3rd International Conference on the Peaceful Uses of Atomic Energy*, Geneva, 31 August-9 September 1964, 398.
- [5] 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>
- [6] Lahoz, W., Khattatov, B. and Ménard, R. (2010) Data Assimilation: Making Sense of Observations. Springer Verlag, Heidelberg. <https://doi.org/10.1007/978-3-540-74703-1>
- [7] Faragó, I., Havasi, Á. and Zlatev, Z. (2013) Advanced Numerical Methods for Complex Environmental Models: Needs and Availability. Bentham Science Publishers, Bussum. <https://doi.org/10.2174/97816080577881130101>
- [8] Cacuci, D.G., Navon, M.I. and Ionescu-Bujor, M. (2014) Computational Methods for Data Evaluation and Assimilation. Chapman & Hall/CRC, Boca Raton.
- [9] Rowlands, J., *et al.* (1973) The Production and Performance of the Adjusted Cross-Section Set FGL5. *Proceedings International Symposium on Physics of Fast Reactors*, Tokio. <http://www.cern.ch/root>
- [10] 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.

-
- [11] 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>
- [12] 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>
- [13] 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>
- [14] 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>
- [15] Jaynes, E.T. (1957) Information Theory and Statistical Mechanics. *Physical Review*, **106**, 620-630. <https://doi.org/10.1103/PhysRev.106.620>
- [16] Cacuci, D.G. (1981) Sensitivity Theory for Nonlinear Systems: I. Nonlinear Functional Analysis Approach. *Journal of Mathematical Physics*, **22**, 2794-2812. <https://doi.org/10.1063/1.525186>
- [17] Cacuci, D.G. (2016) The Second-Order Adjoint Sensitivity Analysis Methodology for Nonlinear Systems—I: Theory. *Nuclear Science and Engineering*, **184**, 16-30. <https://doi.org/10.13182/NSE16-16>
- [18] 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>
- [19] Cacuci, D.G. (2023) Second-Order MaxEnt Predictive Modeling Methodology. II: Probabilistically Incorporated Computational Model Responses (2nd-BERRU-PMP). *American Journal of Computational Mathematics*, **13**, No. 2.
- [20] 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>
- [21] 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>
- [22] Bernardo, J.M. and Smith, A.F.M. (2006) Bayesian Theory. John Wiley & Sons, Chichester.

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

When an unknown distribution $p(\boldsymbol{\alpha}, \mathbf{r})$, 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 [15] 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 $p(\boldsymbol{\alpha}, \mathbf{r})$ are known, the corresponding second-order MaxEnt distribution $p(\boldsymbol{\alpha}, \mathbf{r})$ is constructed by following the procedure outlined below.

1) Known means and covariances for responses:

$$r_i^0 \triangleq \int_D r_i p(\boldsymbol{\alpha}, \mathbf{r}) d\boldsymbol{\alpha} d\mathbf{r}, \quad i = 1, \dots, TR; \tag{96}$$

$$\text{cov}(r_i, r_j) \triangleq \int_{D_e} (r_i - r_i^0)(r_j - r_j^0) p(\boldsymbol{\alpha}, \mathbf{r}) d\boldsymbol{\alpha} d\mathbf{r}; \quad i, j = 1, \dots, TR. \tag{97}$$

2) Known means and covariances for parameters:

$$\alpha_j^0 \triangleq \int_{D_e} \alpha_j p(\boldsymbol{\alpha}, \mathbf{r}) d\boldsymbol{\alpha} d\mathbf{r}, \quad j = 1, \dots, TP; \tag{98}$$

$$\text{cov}(\alpha_i, \alpha_j) \triangleq \int_{D_e} (\alpha_i - \alpha_i^0)(\alpha_j - \alpha_j^0) p(\boldsymbol{\alpha}, \mathbf{r}) d\boldsymbol{\alpha} d\mathbf{r}; \quad i, j = 1, \dots, TP; \tag{99}$$

$$\text{cov}(\alpha_i, r_j) \triangleq \int_{D_e} (\alpha_i - \alpha_i^0)(r_j - r_j^0) p(\boldsymbol{\alpha}, \mathbf{r}) d\boldsymbol{\alpha} d\mathbf{r}; \quad i = 1, \dots, TP; \quad j = 1, \dots, TR. \tag{100}$$

According to the MaxEnt principle, the probability density $p(\boldsymbol{\alpha}, \mathbf{r})$ would satisfy the “available information” provided in Equations (96)-(100), without implying any spurious information or hidden assumptions, if:

1) $p(\boldsymbol{\alpha}, \mathbf{r})$ maximizes the Shannon [21] information entropy, S , as defined below:

$$S = - \int_D p(\boldsymbol{\alpha}, \mathbf{r}) \ln[p(\boldsymbol{\alpha}, \mathbf{r})] d\boldsymbol{\alpha} d\mathbf{r}, \tag{101}$$

2) $p(\boldsymbol{\alpha}, \mathbf{r})$ satisfies the “moments constraints” defined by Equations (96)-(100);

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

$$\int_D p(\boldsymbol{\alpha}, \mathbf{r}) d\boldsymbol{\alpha} d\mathbf{r} = 1. \tag{102}$$

The MaxEnt distribution $p(\boldsymbol{\alpha}, \mathbf{r})$ 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:

$$H(p) = - \int_D p(\boldsymbol{\alpha}, \mathbf{r}) \ln[p(\boldsymbol{\alpha}, \mathbf{r})] d\boldsymbol{\alpha} d\mathbf{r} - \lambda_0 \left[\int_D p(\boldsymbol{\alpha}, \mathbf{r}) d\boldsymbol{\alpha} d\mathbf{r} - 1 \right] - \sum_{k=1}^{TR} \lambda_k^{(1)} \left[\int_D r_k p(\boldsymbol{\alpha}, \mathbf{r}) d\boldsymbol{\alpha} d\mathbf{r} - r_k^0 \right] - \sum_{i=1}^{TP} \lambda_i^{(2)} \left[\int_D \alpha_i p(\boldsymbol{\alpha}, \mathbf{r}) d\boldsymbol{\alpha} d\mathbf{r} - \alpha_i^0 \right]$$

$$\begin{aligned}
 & -\frac{1}{2} \sum_{k=1}^{TR} \sum_{\ell=1}^{TR} \lambda_{k\ell}^{(11)} \left[\int_D r_k r_\ell p(\boldsymbol{\alpha}, \mathbf{r}) d\boldsymbol{\alpha} d\mathbf{r} - \text{cov}(r_k, r_\ell) - r_k^0 r_\ell^0 \right] \\
 & -\sum_{k=1}^{TR} \sum_{i=1}^{TP} \lambda_{ki}^{(12)} \left[\int_D r_k \alpha_i p(\boldsymbol{\alpha}, \mathbf{r}) d\boldsymbol{\alpha} d\mathbf{r} - \text{cov}(r_k, \alpha_i) - r_k^0 \alpha_i^0 \right] \\
 & -\frac{1}{2} \sum_{i=1}^{TP} \sum_{j=1}^{TP} \lambda_{ij}^{(22)} \left[\int_D \alpha_i \alpha_j p(\boldsymbol{\alpha}, \mathbf{r}) d\boldsymbol{\alpha} d\mathbf{r} - \text{cov}(\alpha_i, \alpha_j) - \alpha_i^0 \alpha_j^0 \right].
 \end{aligned} \tag{103}$$

In Equation (103), 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 $p(\mathbf{z})$:

$$p_c(\mathbf{z}) = \frac{1}{Z(\mathbf{b}, \boldsymbol{\Lambda})} \exp\left(-\mathbf{b}^\dagger \mathbf{z} - \frac{1}{2} \mathbf{z}^\dagger \boldsymbol{\Lambda} \mathbf{z}\right), \tag{104}$$

where the various vectors and matrices are defined as follows:

$$\mathbf{z} = \begin{pmatrix} \mathbf{r} \\ \boldsymbol{\alpha} \end{pmatrix}; \quad \mathbf{b} \triangleq \begin{pmatrix} \boldsymbol{\lambda}^{(1)} \\ \boldsymbol{\lambda}^{(2)} \end{pmatrix}; \quad \boldsymbol{\lambda}^{(1)} \triangleq \begin{pmatrix} \lambda_1^{(1)} \\ \cdot \\ \lambda_{TR}^{(1)} \end{pmatrix}; \quad \boldsymbol{\lambda}^{(2)} \triangleq \begin{pmatrix} \lambda_1^{(2)} \\ \cdot \\ \lambda_{TP}^{(1)} \end{pmatrix}; \tag{105}$$

$$\begin{aligned}
 \boldsymbol{\Lambda} & \triangleq \begin{pmatrix} \boldsymbol{\Lambda}^{(11)} & \boldsymbol{\Lambda}^{(12)} \\ \left[\boldsymbol{\Lambda}^{(12)}\right]^\dagger & \boldsymbol{\Lambda}^{(22)} \end{pmatrix}; \quad \boldsymbol{\Lambda}^{(11)} \triangleq \begin{pmatrix} \lambda_{11}^{(11)} & \cdot & \lambda_{1,TR}^{(11)} \\ \cdot & \lambda_{k\ell}^{(11)} & \cdot \\ \lambda_{TR,1}^{(11)} & \cdot & \lambda_{TR,TR}^{(11)} \end{pmatrix}; \\
 \boldsymbol{\Lambda}^{(12)} & \triangleq \begin{pmatrix} \lambda_{11}^{(12)} & \cdot & \lambda_{1,TP}^{(12)} \\ \cdot & \lambda_{k\ell}^{(12)} & \cdot \\ \lambda_{TR,1}^{(12)} & \cdot & \lambda_{TR,TP}^{(12)} \end{pmatrix}; \quad \boldsymbol{\Lambda}^{(22)} \triangleq \begin{pmatrix} \lambda_{11}^{(22)} & \cdot & \lambda_{1,TP}^{(22)} \\ \cdot & \lambda_{k\ell}^{(22)} & \cdot \\ \lambda_{TP,1}^{(22)} & \cdot & \lambda_{TP,TP}^{(22)} \end{pmatrix}.
 \end{aligned} \tag{106}$$

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

$$Z(\mathbf{b}, \boldsymbol{\Lambda}) = \int_D \exp\left(-\mathbf{b}^\dagger \mathbf{z} - \frac{1}{2} \mathbf{z}^\dagger \boldsymbol{\Lambda} \mathbf{z}\right) d\mathbf{z}; \quad d\mathbf{z} \equiv d\boldsymbol{\alpha} d\mathbf{r}. \tag{107}$$

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 $p(\mathbf{z})$ 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:

$$Z_c(\mathbf{b}, \boldsymbol{\Lambda}_c) = \int_{\mathbb{R}^N} \exp\left(-\mathbf{b}^\dagger \mathbf{z} - \frac{1}{2} \mathbf{z}^\dagger \boldsymbol{\Lambda} \mathbf{z}\right) d\mathbf{z} = \frac{(2\pi)^{N/2}}{\sqrt{\text{Det}(\boldsymbol{\Lambda})}} e^{\left(\frac{1}{2} \mathbf{b}^\dagger \boldsymbol{\Lambda}^{-1} \mathbf{b}\right)}. \tag{108}$$

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

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

$$\frac{\partial F(\mathbf{b}, \mathbf{C})}{\partial \lambda_k^{(1)}} = \frac{1}{Z_D} \int r_k e^{\left(-b^\dagger \mathbf{z} - \frac{1}{2} \mathbf{z}^\dagger \boldsymbol{\Lambda} \mathbf{z}\right)} d\mathbf{z} = \int_{D_c} r_k p(\boldsymbol{\alpha}, \mathbf{r}) d\mathbf{z} = r_k^0; k = 1, \dots, TR; \quad (109)$$

$$\frac{\partial F(\mathbf{b}, \mathbf{C})}{\partial \lambda_i^{(2)}} = \frac{1}{Z_D} \int \alpha_i e^{\left(-b^\dagger \mathbf{z} - \frac{1}{2} \mathbf{z}^\dagger \boldsymbol{\Lambda} \mathbf{z}\right)} d\mathbf{z} = \int \alpha_i p(\boldsymbol{\alpha}, \mathbf{r}) d\boldsymbol{\alpha} d\mathbf{r} = \alpha_i^0; i = 1, \dots, TP. \quad (110)$$

The results obtained in Equations (109) and (110) can be collectively written in vector-matrix form as follows:

$$\frac{\partial F(\mathbf{b}, \mathbf{C})}{\partial \mathbf{b}} = \mathbf{z}^0; \quad \mathbf{z}^0 \triangleq (\mathbf{r}^0, \boldsymbol{\alpha}^0)^\dagger; \quad \mathbf{r}^0 \triangleq (r_1^0, \dots, r_{TR}^0). \quad (111)$$

On the other hand, it follows from Equation (108) that:

$$\frac{\partial F(\mathbf{b}, \mathbf{C})}{\partial \mathbf{b}} = -\frac{\partial [\ln Z(\mathbf{b}, \mathbf{C})]}{\partial \mathbf{b}} = -\boldsymbol{\Lambda}^{-1} \mathbf{b}. \quad (112)$$

The relations obtained in Equations (111) and (112) imply the following relation:

$$\mathbf{b} = -\boldsymbol{\Lambda} \mathbf{z}^0. \quad (113)$$

Differentiating a second time the relation provided in Equation (109) or (110) yields the following relations:

$$\begin{aligned} \frac{\partial^2 F(\mathbf{b}, \mathbf{C})}{\partial \lambda_j^{(1)} \partial \lambda_k^{(1)}} &= -\frac{1}{Z^2} \frac{\partial Z}{\partial \lambda_j^{(1)}} r_k^0 + \frac{1}{Z} \int (-r_j) r_k e^{\left(-b^\dagger \mathbf{z} - \frac{1}{2} \mathbf{z}^\dagger \boldsymbol{\Lambda} \mathbf{z}\right)} d\mathbf{z} \\ &= -\text{cov}(r_j, r_k); \quad j, k = 1, \dots, TR; \end{aligned} \quad (114)$$

$$\frac{\partial^2 F(\mathbf{b}, \mathbf{C})}{\partial \lambda_i^{(2)} \partial \lambda_k^{(1)}} = -\text{cov}(\alpha_i, r_k); \quad i = 1, \dots, TP; \quad k = 1, \dots, TR; \quad (115)$$

$$\frac{\partial^2 F(\mathbf{b}, \mathbf{C})}{\partial \lambda_i^{(2)} \partial \lambda_j^{(2)}} = -\text{cov}(\alpha_i, \alpha_j); \quad i, j = 1, \dots, TP. \quad (116)$$

The results obtained in Equations (114)-(116) can be collectively written in vector-matrix form as follows:

$$\begin{aligned} \frac{\partial^2 F(\mathbf{b}, \mathbf{C}_c)}{\partial \mathbf{b} \partial \mathbf{b}} &= -\mathbf{C}; \quad \mathbf{C} \triangleq \begin{pmatrix} \mathbf{C}^{rr} & \mathbf{C}^{ra} \\ \mathbf{C}^{ar} & \mathbf{C}^{aa} \end{pmatrix}; \quad \mathbf{C}^{rr} \triangleq [\text{cov}(r_j, r_k)]_{TR \times TR}; \\ \mathbf{C}^{ra} &\triangleq [\text{cov}(r_k, \alpha_i)]_{TR \times TP} = (\mathbf{C}^{ar})^\dagger; \quad \mathbf{C}^{aa} \triangleq [\text{cov}(\alpha_i, \alpha_j)]_{TP \times TP}. \end{aligned} \quad (117)$$

On the other hand, it follows from Equation (112) that:

$$\frac{\partial^2 F(\mathbf{b}, \mathbf{C})}{\partial \mathbf{b} \partial \mathbf{b}} = -\boldsymbol{\Lambda}^{-1}. \quad (118)$$

The relations obtained in Equations (117) and (118) imply the following relation:

$$\boldsymbol{\Lambda}^{-1} = \mathbf{C}. \quad (119)$$

Introducing the results obtained in Equations (113) and (119) into Equation (104) and (108) yields the following expression for the MaxEnt distribution

$p(\mathbf{z}|\mathbf{z}, \mathbf{C})$:

$$p(\mathbf{z}|\mathbf{z}, \mathbf{C}) = \frac{(2\pi)^{-N/2}}{\sqrt{\text{Det}(\mathbf{C})}} \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{z}^0)^\dagger \mathbf{C}^{-1}(\mathbf{z} - \mathbf{z}^0)\right]. \quad (120)$$

Appendix B: Construction of the “Constrained Minimum Discrepancy Distribution with a Given Function”

The aim is to obtain a mathematical expression for a distribution $p(\mathbf{z})$, defined over a domain $\mathbf{z} \in D$, which is normalized to unity and satisfies K independent constraints of the following form:

$$\int_D g_k(\mathbf{z}) p(\mathbf{z}) d\mathbf{z} = c_k < \infty, \quad k = 1, \dots, K; \quad (121)$$

with

$$\int_D p(\mathbf{z}) d\mathbf{z} = 1. \quad (122)$$

In addition, it is required that the normalized distribution $p(\mathbf{z})$ be “minimally discrepant from” (*i.e.*, “fit as well as possible”) a known function $f(\mathbf{z})$, which is also defined over the same domain $\mathbf{z} \in D$.

The discrepancy (*i.e.*, “lack of fit”) between an approximate distribution $f(\mathbf{z})$ and a true distribution $p(\mathbf{z})$, both defined over the same domain $\mathbf{z} \in D$, is denoted below as $\delta(f|p)$, and is defined [see, e.g.,] as follows:

$$\delta(f|p) \triangleq \int_D p(\mathbf{z}) \log \frac{p(\mathbf{z})}{f(\mathbf{z})} d\mathbf{z}. \quad (123)$$

The discrepancy $\delta(f|p)$ is also called, see e.g., [22], the “Kulback-Leibler divergence or relative entropy.” The normalized distribution $p(\mathbf{z})$ which minimizes the discrepancy defined in Equation (123) while satisfying the constraints expressed by Equations (121) and (122) is obtained by minimizing the Lagrangian functional $F(p)$ defined below:

$$F(p) \triangleq \int_D p(\mathbf{z}) \log \frac{p(\mathbf{z})}{f(\mathbf{z})} d\mathbf{z} + \sum_{k=1}^{TR} \theta_k \left[\int_D g_k(\mathbf{z}) p(\mathbf{z}) d\mathbf{z} - c_k \right] + c \left[\int_D p(\mathbf{z}) d\mathbf{z} - 1 \right]. \quad (124)$$

In Equation (124), the quantities c and $\theta_1, \dots, \theta_k, \dots, \theta_{TR}$ are Lagrange multipliers which remain to be determined. The stationary value of $F(p)$ is provided by the probability $p(\mathbf{z})$ where the first variation of $F(p)$ vanishes, *i.e.*, $p(\mathbf{z})$ is the solution of the following equation

$$\begin{aligned} \delta F(p) &\triangleq \frac{d}{d\varepsilon} \left\{ F[p(\mathbf{z}) + \varepsilon \tau(\mathbf{z})] \right\}_{\varepsilon=0} \\ &= \int_D \left[\log \frac{p(\mathbf{z})}{f(\mathbf{z})} + \sum_{k=1}^{TR} \theta_k g_k(\mathbf{z}) + (c+1) \right] \tau(\mathbf{z}) d\mathbf{z} = 0. \end{aligned} \quad (125)$$

The desired posterior distribution is the normalized solution of Equation (125), having the following expression:

$$p(\mathbf{z}) = \frac{1}{Z} f(\mathbf{z}) \exp\left[-\sum_{k=1}^{TR} \theta_k g_k(\mathbf{z})\right]; \quad Z \triangleq \int_D f(\mathbf{z}) \exp\left[-\sum_{k=1}^{TR} \theta_k g_k(\mathbf{z})\right] d\mathbf{z}. \quad (126)$$

It is important to note that in the limiting case when the function $f(z)$ is the uniform distribution $f(z) = \text{constant}$, the distribution $p(z)$ becomes a MaxEnt-distribution. In other words, the distribution which is least discrepant from (*i.e.*, “best-fits”) the uniform distribution is the MaxEnt distribution.