LA-UR-23-21950

Approved for public release; distribution is unlimited.

Title:	Multi-phenomenology Yield Characterization
Author(s):	Williams, Brian J. Picard, Richard Roy Anderson, Dale N.
Intended for:	Report
Issued:	2023-03-06 (rev.1)

Multi-phenomenology Yield Characterization

Brian J. Williams[1], Richard R. Picard[1], and Dale N. Anderson[2]

[1]*Los Alamos National Laboratory*
[2]*Nevada National Security Site*

February 23, 2023

1 Executive Summary

This report serves as the first delivery of a four-year applied science effort to transform and advance the error bounds for the yield estimate of an explosion. Each year's delivery will be in this form, culminating in the submission of this work for peer review to a scientific journal. Importantly, the yearly progress reports can then also be viewed as expanding drafts working towards a formal journal article submission.

For the first tranche of funding, we collaborated with Air Force Technical Applications Center (AFTAC) scientists to identify unclassified real-world data that demonstrate and validate our advanced error propagation methods. Collaboration includes visits to AFTAC and telecons. For this development, we illustrate the fusion of seismic, acoustic, optical, and surface effect signatures from an explosion. The mathematics and code being adapted to this specific application (Williams et al., 2021) involves physics models of multiple sensor signatures. We have also identified related physics models and have integrated them into code.

Current methods of underground explosion yield estimation for the Threshold Test Ban Treaty (TTBT) have served the US treaty monitoring mission well for decades. A research objective of the Defense Nuclear Nonproliferation Research and Development (DNN R&D) office of the National Nuclear Security Administration (NNSA) has always been to provide new technical capabilities for monitoring lower thresholds. The general model and error propagation code to be developed in this project is based on significant advances in error modeling and propagation needed to analyze data at lower detection thresholds.

The second tranche of funding for this project began on May 1, 2022, and planned work for the second tranche includes:

i) completing the integration of physical model code into the general error model framework; this code accommodates a wide range of linear/nonlinear source models, fixed/random effects, and frequentist/Bayesian analyses (the purpose of which is *not* to dictate to users how to analyze data, but instead to allow users the maximum flexibility in their work),

ii) illustrative application of code to identified data, and

iii) initial planning with AFTAC researchers on delivery of code to the Common Development Environment at AFTAC, and continued writing of the planned final journal article submission (year two of this progress report), with particular emphasis on descriptions of data identified for this effort.

2 Introduction

This project will demonstrate multi-physics (multivariate) yield estimation using disparate physical sensor signatures. Importantly, this methodology correctly aligns error propagation for yield estimation with first-order physical basis from explosion source to propagation to sensor. This means that signature weighting in a yield calculation is done correctly, including accounting for error correlation between disparate signatures. It also means that the corresponding advanced error budget includes errors for source, propagation, and sensor site models — errors that masquerade as other uncertainties in models that do not directly incorporate these effects. These components of error in our advanced error budget, expressed as statistical variances/standard deviations, provide metrics on the technical validation of ensemble model components (e.g., source, propagation, and sensor).

The scientific basis and general linear modeling approach is described in Williams et al. (2021), and is now generalized to include nonlinear forward models. This research also expands the advanced univariate error modeling for explosions described in Anderson et al. (2009) to a nonlinear multivariate setting, demonstrating the development with seismic and acoustic data observed from surface chemical explosions. Similar developments to study regional phase variability can be found in Taylor (2007).

For this project, we expand the general multivariate version (multi-phenomenology or MultiPEM) in the Williams et al. (2021) analysis to include seismic, acoustic, optical and surface effect signatures to demonstrate yield estimation for above-ground explosions. This project involves a close collaboration with the U.S. Air Force treaty monitoring mission.

A conceptual technical development of the univariate error model is provided in the next section. Particular attention is given to the interpretation of the model error variances in terms of their utility in model validation. Following the univariate review is a review of multi-phenomenology error model development. A preliminary application of our model then follows for estimating the yield of above ground explosions. The simulated data used in this analysis were based on unclassified data kindly provided by Sean Ford (LLNL) as well as other unclassified data sources as indicated.

Finally, we summarize the work completed with the first two tranches of funding for this project.

3 Scientific and Statistical Basis for the Framework

The advanced error model is consistent with established models and physical basis of waveform signal generation and propagation for many physical explosion signals. Presentation of the related material is necessarily mathematics-intensive.

Consider seismic or acoustic calibration data. Physically meaningful segments of an explosion time series are modeled as a convolution of source (S) and propagation (P) time

series, denoted in the frequency (ω) domain as amplitude.

$$E[Y(\omega)] = S(\omega; \boldsymbol{\beta}_S, \boldsymbol{v}_S) \times P(\omega; \boldsymbol{\beta}_P, \boldsymbol{v}_P)$$

(Stein and Wysession, 2013). Here $\boldsymbol{\beta}_S$ is a vector of (unknown) parameters important to the description of the source (such as forward model coefficients), and $\boldsymbol{v}_S$ is a vector of measured covariates describing the source (such as coupling variation, water content in rock, yield, and height-of-burst/depth-of-burial (HOB/DOB)); and $\boldsymbol{\beta}_P$ and $\boldsymbol{v}_P$ are the corresponding descriptions of a path, with $E[\cdot]$ denoting expected amplitude. For new event data, some "covariates" become "parameters" when their values are unknown, such as yield and HOB/DOB. The advanced error model that follows will be developed in the frequency domain from this basic physical model and context.

Adding appropriate subscripts provides a description of this model for observed amplitudes $a_{hijrk}(\omega)$. Here, the nested subscripts indicate the k-th sensor measurement from the r-th amplitude measurement type along the j-th source-to-sensor propagation path generating observations of the i-th source (explosion) for sensor type h (e.g., seismic, acoustic, etc.). The individual sensor measurements are assumed to arise from a network of sensors deployed in identical and uncorrelated noise.

The amplitude model for this initial formulation is

$$a_{hijrk}(\omega) = S_{hir}(\omega; \boldsymbol{\beta}_{S,hr}, \boldsymbol{v}_{S,hi}) \times P_{hijr}(\omega; \boldsymbol{\beta}_{P,hr}, \boldsymbol{v}_{P,hij}) \times e_{hijrk}\,,$$

and we make the modeling assumption that the noise factor e_{hijrk} is a geometric mean of a large number of multiplicative perturbations. This implies by the central limit theorem for geometric means that observed amplitudes $a_{hijrk}(\omega)$ have a lognormal probability distribution (Galton, 1879; McAlister, 1879).

In log-space this spectral domain model is

$$\log a_{hijrk}(\omega) = \log S_{hir}(\omega; \boldsymbol{\beta}_{S,hr}, \boldsymbol{v}_{S,hi}) + \log P_{hijr}(\omega; \boldsymbol{\beta}_{P,hr}, \boldsymbol{v}_{P,hij}) + \log e_{hijrk}, \qquad (1)$$

which implies that $\log a_{hijrk}(\omega)$ can be modeled with a Gaussian probability distribution, that is, $\log e_{hijrk} = \epsilon_{hijrk}$ are independently Gaussian distributed. The source and path parameter vectors are allowed to depend on sensor and measurement types, as well as possibly on clustering factors that dictate unique source or path forward models, such as emplacement condition (e.g. hard, soft, or wet rock). Information specific to individual sources and paths needed for forward model calculations is incorporated into the relevant covariate vectors.

Two important features of the modeling involve the use of "random effects" versus "fixed effects" and the direct incorporation of yield uncertainties. Regarding the former, the physical terms in the model

$$\log a_{hijrk}(\omega) = \log S_{hir}(\omega; \boldsymbol{\beta}_{S,hr}, \boldsymbol{v}_{S,hi}) + \log P_{hijr}(\omega; \boldsymbol{\beta}_{P,hr}, \boldsymbol{v}_{P,hij}) + \epsilon_{hijrk},$$

are often not fully adequate for describing observed data. Bias in the source and path model terms $\log S_{hir}(\omega; \boldsymbol{\beta}_{S,hr}, \boldsymbol{v}_{S,hi})$ and $\log P_{hijr}(\omega; \boldsymbol{\beta}_{P,hr}, \boldsymbol{v}_{P,hij})$ are composed of i) bias that can be calibrated and ii) bias that must be thought of as dynamic, e.g. unknown characteristics of emplacement propagation that vary from explosion to explosion. Accommodating dynamic bias terms in the base model gives

$$\log a_{hijrk}(\omega) = \log S_{hir}(\omega; \boldsymbol{\beta}_{S,hr}, \boldsymbol{v}_{S,hi}) + E_{S,hir} + \log P_{hijr}(\omega; \boldsymbol{\beta}_{P,hr}, \boldsymbol{v}_{P,hij}) + E_{P,hijr} + \epsilon_{hijrk}, \quad (2)$$

where the terms $E_{S,hir}$ and $E_{P,hijr}$ represent the dynamic biases in the physically motivated source and path models, respectively.

The subscripts on the dynamic bias terms are important for physical interpretation of the model. Fully consistent with physical basis, for explosion i the bias in the physical source term $\log S_{hir}(\omega; \boldsymbol{\beta}_{S,hr}, \boldsymbol{v}_{S,hi})$ applies to the prediction of all observed data, $\log a_{hijrk}(\omega)$. For path j observing explosion i, the bias in the physical path term $\log P_{hijr}(\omega; \boldsymbol{\beta}_{P,hr}, \boldsymbol{v}_{P,hij})$ applies only to observed data for path j. Changes in source term bias apply to all data, while changes in path j for explosion i only apply to data observed on that path.

Note that this involves a philosophical change in the interpretation of the dynamic bias terms $E_{S,hir}$ and $E_{P,hijr}$, describing these unknown physical parameters as random effects. This change is called for because of the difference in monitoring regimes. Throughout TTBT history, most underground nuclear testing was confined to fixed, known locations – i.e., the underground test sites of the respective nuclear powers. Similarly, the same fixed teleseismic monitoring stations supplied data for test after test. As such, the same source-to-sensor paths were repeatedly involved in monitoring, and a fixed-effects approach was well founded. Looking to the future, however, the location of a clandestine nuclear event may not be known in advance, nor would the corresponding source-to-sensor paths. Consequently, path effects are not as predictable as for the TTBT case.

The random effects treatment of dynamic biases allows explosions and source-to-sensor propagation pathways involved in an analysis of interest to be regarded as samples from statistical populations of these entities, allowing for new event data to be seamlessly incorporated into the analysis.

Another component of the error model involves yield estimation. Estimation methods for the TTBT were adequate for testing the larger yields of concern then, although related uncertainty quantification was at times problematic (Picard and Bryson, 1992). Because such yields were always large, a simple error model captured the dominant error sources. For lower yield monitoring, other sources of error can become important.

Observed yields are typically estimates obtained from one of two sources: (a) design yield as obtained from calculations for a device prior to testing, such as for chemical explosions, and (b) measurement data, such as from radiochemistry on drillback samples after a nuclear test. Both sources are subject to uncertainty: in (a), computations are imperfect, conventional-to-nuclear yield conversion factors are not known exactly, and materials are subject to intrinsic manufacturing variability, and (b) involves analysis of data subject to aleatory uncertainty.

A classic statistical result (Levi, 1973) is that a regression coefficient corresponding to the least squares fit mistakenly treating an independent covariate as "known" produces a corresponding parameter estimate that is biased towards zero even in the presence of other known covariates, and also produces biased predicted values when covariate uncertainties are nonnegligible relative to other uncertainties. Noninformative Bayesian estimates that might further shrink a yield parameter coefficient estimate towards zero only magnify this bias.

Yield uncertainty is addressed by modeling calibration yields as being estimated up to a user-specified error. Specifically, a common *errors-in-variables* treatment provides that

$$\widetilde{W}_s \sim w_s + \epsilon_s\,, \quad \epsilon_s \sim \mathcal{N}(0, \sigma_s^2)\,, \quad s \in \mathcal{S}\,, \tag{3}$$

where $\widetilde{W}_s$ is the measured/design log-yield, and w_s is the actual log-yield for the source, for calibration source s included in the catalog $\mathcal{S}$ of unique sources across sensor types. The

standard deviation parameter σ_s is specified by the user, and controls the degree of deviation ϵ_s permitted between the actual and assumed yields for source s.

Validation efforts relative to the model herein involves important technical lanes between the physical and statistical aspects of the model. For each sensor type, defining the signal transition from source to propagation path to sensor, and the transition across paths, is fundamentally specified by physics with the physical model terms $\log S_{hir}(\omega; \boldsymbol{\beta}_{S,hr}, \boldsymbol{v}_{S,hi})$ and $\log P_{hijr}(\omega; \boldsymbol{\beta}_{P,hr}, \boldsymbol{v}_{P,hij})$. Validating these terms and conducting data analyses to discover unknown quantities describing events of interest can be established with statistical theory using multiple explosions, each observed with multiple paths that lead to a network of sensors. Validation across differing physical context, e.g., explosions in differing geologies, requires data from each of the different emplacement conditions.

Consider two explosions with two differing propagation-to-sensors configurations, and with roughly equivalent emplacement. A well-designed collection of validation data: i) statistically confirms that a single model is appropriate for both, or ii) determines that distinct models are needed for each and in each case calibrate all physical and statistical model parameters. Importantly, once validation data are acquired as specified by the full model, improvements in model terms can be re-validated with these data. Additionally, the validation analysis ranks model terms based on contribution to total error, potentially suggesting research priorities.

4 Model Interpretation and Application

Equation (2) can be written as

$$y_{hijrk} = f_{hijr}(\boldsymbol{\beta}_{hr}, \boldsymbol{v}_{hij}) + E_{S,hir} + E_{P,hijr} + \epsilon_{hijrk}\,, \tag{4}$$

where $y_{hijrk} = \log a_{hijrk}$ and in the case of the previous section,

$$f_{hijr}(\boldsymbol{\beta}_{hr}, \boldsymbol{v}_{hij}) = \log S_{hir}(\omega; \boldsymbol{\beta}_{S,hr}, \boldsymbol{v}_{S,hi}) + \log P_{hijr}(\omega; \boldsymbol{\beta}_{P,hr}, \boldsymbol{v}_{P,hij})\,.$$

The frequency ω has been omitted from the forward model $f_{hijr}(\boldsymbol{\beta}_{hr}, \boldsymbol{v}_{hij})$ to highlight that the form of Equation (4) is also utilized for sensor types where physical motivation (say, the above frequency domain arguments) is not applicable as a justification.

Specifically, given sensor type h and measurement type r, this random effects treatment of dynamic bias takes the $\{E_{S,hir}\}_i$ to be independent and identically distributed (iid) Gaussian variables having mean zero and variance $\tau^2_{S,hr}$; the $\{E_{P,hijr}\}_{ij}$ are taken to be iid Gaussian variables having mean zero and variance $\tau^2_{P,hr}$; and the collection of random variables $\{\{E_{S,hir}\}_i, \{E_{P,hijr}\}_{ij}\}$ is assumed to be mutually independent. The observational error terms $\{\epsilon_{hijrk}\}_{ijk}$ are taken to be iid Gaussian variables having mean zero and variance σ^2_{hr}, independent of all dynamic bias terms. Random effects terms for different measurement types are assumed to be mutually independent, while observational error terms for different measurement types may be correlated. The parameters $\{\tau^2_{S,hr}\}_{hr}$, $\{\tau^2_{P,hr}\}_{hr}$, and $\{\sigma^2_{hr}\}_{hr}$ can be estimated statistically from relevant calibration data, providing information to significantly advance model validation and yield estimation.

Equation (4) establishes a baseline for the notation of the general (nonlinear) multivariate version of our model discussed in Section 5. Collecting physical model terms in this way is

common in explosion monitoring research because these terms are nonlinear and a Taylor series expansion (nonlinear regression) of the physics models is a common approach to model calibration and analysis (Fagan et al., 2009; Ford et al., 2021).

When $S_{hir}(\omega; \boldsymbol{\beta}_{S,hr}, \boldsymbol{v}_{S,hi})$ and $P_{hijr}(\omega; \boldsymbol{\beta}_{P,hr}, \boldsymbol{v}_{P,hij})$ provide sound structural (mathematical) descriptions of data, then $E_{S,hir}$ and $E_{P,hijr}$ are statistically negligible.

For discussion, consider the source model taking the form of a simple linear model $\log S_{hir}(\omega; \boldsymbol{\beta}_{S,hr}, \boldsymbol{v}_{S,hi}) = \mu_{hr}(\omega) + \beta_{hr}(\omega) \times w_i$, for $w = \log \text{yield}$, $\boldsymbol{\beta}_{S,hr} = (\mu_{hr}(\omega), \beta_{hr}(\omega))$, and $\boldsymbol{v}_{S,hi} = w_i$. The full univariate model is then

$$y_{hijrk}(\omega) = \mu_{hr}(\omega) + \beta_{hr}(\omega) \times w_i + E_{S,hir} + \log P_{hijr}(\omega; \boldsymbol{\beta}_{P,hr}, \boldsymbol{v}_{P,hij}) + E_{P,hijr} + \epsilon_{hijrk}\,.$$

And while the estimate of the parameter $\beta_{hr}(\omega)$ may be specific to an operational application, it is nonetheless an estimate possessing estimation error that must be propagated to the estimated log yield for a new event, w_0. This error propagation calculation enables the transport of our model, possibly awaiting calibration data. For example, the source model for an above ground explosion includes a model component to represent energy coupling from air into rock. Appropriate calibration data for this more sophisticated source model provides an estimate of $\{\tau^2_{S,hr}\}_{hr}$ (and $\{\tau^2_{P,hr}\}_{hr}$, $\{\sigma^2_{hr}\}_{hr}$), which enables transport of the model awaiting additional calibration data.

An important aspect of the random effects model is that estimation of the variance components $\{\tau^2_{S,hr}\}_{hr}$, $\{\tau^2_{P,hr}\}_{hr}$ and $\{\sigma^2_{hr}\}_{hr}$ uses fit errors for source, path and sensor site (observation error). These fit errors become diagnostic residuals to identify potential problems with model terms in describing observed data. For example, residual outliers associated with each of the variance components can identify suspect data and/or physical inadequacies in model terms.

This simplified model illustrates an important element of statistical experimental design. For given sensor type h and measurement type r, the ideal calibration data needed for a validation analysis consist of at least two explosions, with each explosion observed by at least two sensor networks (paths), and each network having at least two sensors. Additional observations are necessary to estimate all forward model parameters; in this case, two additional data points to estimate $\mu_{hr}(\omega)$ and $\beta_{hr}(\omega)$. In real applications, the amount of available calibration data will place constraints on the sophistication of forward and error models that may be considered, guided by the statistical theory of parameter estimation.

For an observed explosion i given sensor type h and measurement type r, the full model defines the variance of each observation as $\mathrm{V}[y_{hijrk}(\omega)] = \tau^2_{S,hr} + \tau^2_{P,hr} + \sigma^2_{hr}$. The covariance between observations from different networks (paths) is $\mathrm{Cov}[y_{hijrk}(\omega), y_{hij'rk}(\omega)] = \tau^2_{S,hr}$, $j \neq j'$, because these data probabilistically co-vary only through a shared source error $E_{S,hir}$. The covariance between observations from the same network (path) incorporates both contributors $E_{S,hir}$ and $E_{P,hijr}$ to dynamic bias, $\mathrm{Cov}[y_{hijrk}(\omega), y_{hijrk'}(\omega)] = \tau^2_{S,hr} + \tau^2_{P,hr}$, $k \neq k'$. The individual random effect terms are not directly fit in the statistical estimation process; rather only the variance components of the above covariance structure are estimated. The individual dynamic bias terms can then be *predicted* utilizing maximum likelihood methodology. A consequence of this random effects structure is that calibration or new event data can be readily added or deleted from the statistical inference process. Furthermore, the additive nature of the source and path dynamic bias terms is consistent with the physical basis resulting in Equation (1).

5 The Multi-Phenomenology (MultiPEM) Framework

Highlights of the MultiPEM statistical model are provided in this section, with full details of these developments available in Appendix A for interested readers.

5.1 Calibration Data

Collecting the observed data $\{y_{hijrk}\}_k$ into a n_{hijr}-vector of observations extends Equation (4) to vector form,

$$\boldsymbol{y}_{hijr} = \boldsymbol{f}_{hijr}(\boldsymbol{\beta}_{hr}, \boldsymbol{v}_{hij}) + \boldsymbol{E}_{S,hir} + \boldsymbol{E}_{P,hijr} + \boldsymbol{\epsilon}_{hijr} \, .$$

The parameter $\boldsymbol{\beta}_{hr}$ is generalized to allow for dependence on clustering factor t_{hi} of source i for sensor type h, as mentioned in Section 3. In the case of seismic and acoustic data generated by explosions originating in or near different rock types (e.g. hard, soft, wet), enhanced modeling fidelity can be achieved by designating emplacement condition as a clustering factor. The parameter associated with cluster t is designated $\boldsymbol{\beta}_{htr}$. The condition $t = 0$ is reserved for the subset of parameters that have common value across all clusters, so that $\boldsymbol{\beta}_{hr}^{(t)} = (\boldsymbol{\beta}_{h0r}, \boldsymbol{\beta}_{htr})$ is the forward model parameter for calculations involving cluster t. In the absence of a clustering factor, $\boldsymbol{\beta}_{hr}^{(t)} = \boldsymbol{\beta}_{h0r}$. If clustering factor(s) are present with no common parameters among them, $\boldsymbol{\beta}_{hr}^{(t)} = \boldsymbol{\beta}_{htr}$ for each cluster t. In some scenarios, the specific cluster for a new event may be unknown, complicating the analysis.

The source and path dynamic bias vectors $\boldsymbol{E}_{S,hir}$ and $\boldsymbol{E}_{P,hijr}$ are assumed to take the form of linear models,

$$\boldsymbol{E}_{S,hir} = \boldsymbol{Z}_{hir,j}\boldsymbol{b}_{hr}^{(S)} \;\text{ and }\; \boldsymbol{E}_{P,hijr} = \boldsymbol{Z}_{hijr}\boldsymbol{b}_{hr}^{(P)} \, ,$$

where $\boldsymbol{b}_{hr}^{(S)}$ is a $q_{S,hr}$-vector of source-level random coefficients applied to the columns of the fixed covariate matrix $\boldsymbol{Z}_{hir,j}$, and $\boldsymbol{b}_{hr}^{(P)}$ is a $q_{P,hr}$-vector of path-within-source-level random coefficients applied to the columns of the fixed covariate matrix $\boldsymbol{Z}_{hijr}$. The random vectors $\boldsymbol{b}_{hr}^{(S)}$ and $\boldsymbol{b}_{hr}^{(P)}$ are taken to be independent and mean-zero Gaussian distributed,

$$\boldsymbol{b}_{hr}^{(S)} \sim \mathcal{N}(\boldsymbol{0}_{q_{S,hr}}, \boldsymbol{\Sigma}_{hr}^{(S)}) \;\text{ and }\; \boldsymbol{b}_{hr}^{(P)} \sim \mathcal{N}(\boldsymbol{0}_{q_{P,hr}}, \boldsymbol{\Sigma}_{hr}^{(P)}) \tag{5}$$

with $\boldsymbol{0}_n$ the n-vector of zeros, and $\boldsymbol{\Sigma}_{hr}^{(\ell)}$ for $\ell \in \{S, P\}$ the unknown covariance matrices of the source and path-within-source random coefficient vectors. This statistical model for dynamic bias is referred to as a two-level *random effects* model (Pinheiro and Bates, 2000), with "source" as the first level and "path nested within source" as the second level.

The observational error vector $\boldsymbol{\epsilon}_{hijr}$ is assumed to be mean-zero Gaussian distributed, independent of the source and path dynamic bias vectors, having covariance matrix $\sigma_{hrr}\boldsymbol{I}_{n_{hijr}}$ for $\boldsymbol{I}_n$ the $n \times n$ identity matrix. Paired observations across sensor measurement types r_1 and r_2 within sensor type h are allowed to be correlated, with covariance designated by $\sigma_{hr_1r_2}$, subject to the constraint that the matrix $\boldsymbol{\Sigma}_h = (\sigma_{hr_1r_2})$ is positive definite. For given sensor type h and measurement type r, calibration data are assumed to be mutually independent across sources i. In this development, observations across sensor types h are assumed to be mutually independent, although it would be straightforward (and possibly challenging

computationally) to admit correlation across sensor types should an appropriate, physically motivated covariance model become available.

Assembling the elements of the above discussion leads to the general form of the Multi-PEM statistical model for the data used to calibrate all unknown parameters,

$$\boldsymbol{y}_{hijr} = \boldsymbol{f}_{hijr}(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{v}_{hij}) + \boldsymbol{Z}_{hir,j}\boldsymbol{b}_{hr}^{(S)} + \boldsymbol{Z}_{hijr}\boldsymbol{b}_{hr}^{(P)} + \boldsymbol{\epsilon}_{hijr}\,, \tag{6}$$

combined with all distributional assumptions stated above. Two reduced models are also of interest; the first is obtained by eliminating the second-level (path) random effects,

$$\boldsymbol{y}_{hijr} = \boldsymbol{f}_{hijr}(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{v}_{hij}) + \boldsymbol{Z}_{hir,j}\boldsymbol{b}_{hr}^{(S)} + \boldsymbol{\epsilon}_{hijr}\,, \tag{7}$$

and the second is obtained by eliminating all random effects,

$$\boldsymbol{y}_{hijr} = \boldsymbol{f}_{hijr}(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{v}_{hij}) + \boldsymbol{\epsilon}_{hijr}\,. \tag{8}$$

The path random effects are assumed to be nested within the source random effects, precluding inclusion of the former in the model if the latter are absent. Additional details are provided in Appendix A.1.

5.2 New Event Data

New event data is modeled statistically in a fashion analogous to the calibration data. The unknown quantities of interest describing the new event are denoted by the parameter vector $\boldsymbol{\theta}_0$, and potentially include event time, surface location, HOB/DOB, and yield. The ultimate goal of the statistical analysis described in this report is to jointly estimate the elements of $\boldsymbol{\theta}_0$ with rigorous uncertainty quantification, utilizing both the available calibration and new event data simultaneously or in sequence. Note that $\boldsymbol{v}_{h0j} = (\boldsymbol{\theta}_0, \boldsymbol{v}'_{h0j})$ because the quantities contained in $\boldsymbol{\theta}_0$ are unknown for the new event, and are thus separated from the vector of known covariates, now denoted $\boldsymbol{v}'_{h0j}$.

In the ensuing discussion, the index $i = 0$ is reserved for the new event, so that Equation (6) carries over to the new event data as follows,

$$\boldsymbol{y}_{h0jr} = \boldsymbol{f}_{h0jr}((\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0), \boldsymbol{v}'_{h0j}) + \boldsymbol{Z}_{h0r,j}\boldsymbol{b}_{hr}^{(S)} + \boldsymbol{Z}_{h0jr}\boldsymbol{b}_{hr}^{(P)} + \boldsymbol{\epsilon}_{h0jr}\,.$$

Equations (7) and (8) are also modified in the same fashion. New event data is assumed to be distributed independently of calibration data. Additional details are provided in Appendix A.2.

6 Application

The development in this project focuses on analysis of data from four physical signatures generated by an above-ground explosion. Specifically, acoustic waves (compressional waves) in the atmosphere, compressional and shear waves (seismic waves) in the earth, optical (electromagnetic) emissions from an explosion, and the surface effects (crater dimensions) caused by the acoustic wave interaction with the earth (shock and elastic).

Data from each of these four physical signatures was utilized to demonstrate the new event characterization methodology of this report. The size and structure of the data sets for the acoustic, seismic, and optical signatures generally follow the development in Section 2 of Ford et al. (2021); in particular, see Table 1 and Figure 7 (calibration data) and Table 5 (new event data) of that article. The surface effects data unique to this report are described below in Section 6.4, with data provided in Table 2.

In this report, rather than utilizing observed signatures, all signatures were *simulated* from one of the statistical models (6)-(8) based on the covariate information collected in the data sets described above. This allows the results of statistical model selection and new event inference to be compared against the ground truth established for this simulation study. Further, nonlinear models that have appeared in the literature (Ford et al., 2021) are used for illustration of the software, though it is acknowledged that such models may not be optimal. Additional details are given in the following subsections.

6.1 Acoustic

For acoustic signatures, a large above-ground explosion generates pressure that the atmosphere is unable to propagate easily. This atmospheric shock front compresses the air sufficient to cause optical distortion of the atmosphere which is visual before the sound wave from the explosion can be sensed. As this shock front moves through the atmosphere it is creating a void behind which is filled with air. What is sensed acoustically is the compression of the air commonly measured in Pascals (see Figure 1). Eventually the acoustic shock will spread and attenuate so that its compression of the air can be propagated at the speed of sound with no optical distortion.

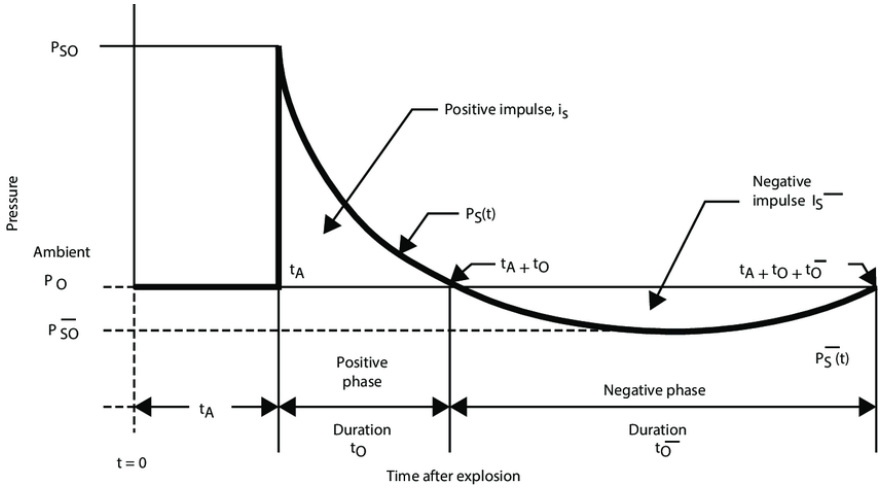

Figure 1: Notional acoustic wave form highlighting relevant features.

Compression on the order of 10^5 Pascals is sufficient to cause optical distortion. The source function of the explosion represents the conversion/partition of the burn of the explosive fuel into compression, with to first order the other partition of energy being heat. It is common to represent the burn energy of the explosive fuel as being equivalent to kilograms/kilotons TNT (explosion yield).

An empirical forward model for signatures measured from the atmospheric compression due to an explosion is

$$f_{ar} \equiv \log(\tilde{d}_{ar}) = \beta_{ar,1} + \beta_{ar,2}\log(\tilde{\delta}_a) + \beta_{ar,3}\tilde{h}_a - \log\left(1 + \exp(\beta_{ar,3}\tilde{h}_a)\right)$$

(Ford et al., 2021). The scaled signatures and covariates of this forward model are given by

$$\begin{aligned}
\tilde{d}_{a1} &= d_{a1}\exp(-w/3)(P/P_0)^{-2/3}(T/T_0)^{1/2} \\
\tilde{d}_{a2} &= d_{a2}\exp(-w/3)(P/P_0)^{-1/3}(T/T_0)^{1/2} \\
\tilde{\delta}_a &= \delta\exp(-w/3)(P/P_0)^{1/3} \\
\tilde{h}_a &= h\exp(-w/3)(P/P_0)^{1/3}\,.
\end{aligned}$$

Table 1 defines the fundamental physical quantities constituting this acoustic forward model. The acoustic error model assumes dynamic bias terms for each signature that modify $\beta_{ar,1}$ for each source. The overpressure impulse and duration signatures are therefore simulated from Equation (7) with the random effects covariate matrices $\boldsymbol{Z}_{hir,j}$ set to all-ones vectors of the appropriate length.

Table 1: Variables of acoustic forward model

Quantity	Description	Quantity	Description
Signatures			
d_{a1}	Overpressure impulse	d_{a2}	Overpressure duration
Covariates			
w	log Yield	h	HOB/DOB
P	Ambient pressure	T	Ambient temperature
P_0	Standard pressure (101,325 Pa)	T_0	Standard temperature (288 K)
δ	Range		

6.2 Seismic

Seismic waves from an above-ground explosion result from the acoustic shock front hitting the earth surface, and are also acoustic compression waves in a solid-yet-elastic media – rock. A strong acoustic compression wave interacting with the surface of the earth can also bend and twist the rock creating shear waves which, like compression waves, can be propagated by rock (see Figure 2).

The earth is a response function taking atmospheric acoustic compression into geologic compression and shear. As a response function, crustal rock is relatively rigid/elastic in that the right combination of explosion height and magnitude is needed to transfer acoustic energy into the rock. This means that this crustal response function should be included in the forward model for seismic energy observed at a sensor. An acoustic shock front hitting the surface of earth can be so strong as to damage and break the rock as well as compressing

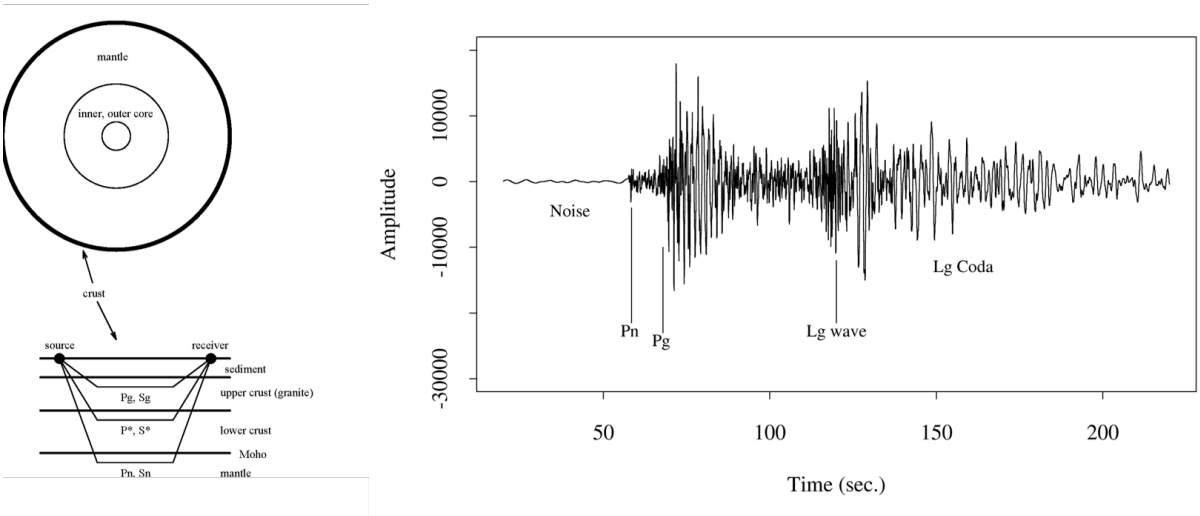

Figure 2: Crustal (regional) seismic wave paths. The notation Pn, Pg, Sn and Lg represents the segment of a seismic wave that follows the associated path.

and shearing the rock, creating a crater. This physical process is also included in a crustal response, either implicitly or explicitly.

Compression seismic energy is observed and modeled in the time/frequency domain as the actual motion of the elastic rock (velocity or acceleration). Ford et al. (2021) propose the seismic forward model from an above-ground explosion as

$$f_{sr} = \log(\tilde{d}_{sr}) = \beta_{sr,1} + \beta_{sr,2}\log(\tilde{\delta}_s) + \beta_{sr,3}\,\text{logistic}(\beta_{sr,4}\tilde{h}_s + \beta_{sr,5})$$

for

$$\text{logistic}(x) = \frac{1}{1+\exp(-x)}\,.$$

The scaled signatures and covariates of this forward model are given by

$$\tilde{d}_{s1} = d_{s1}\exp(-w/3) \qquad \tilde{d}_{s2} = d_{s2}$$

$$\tilde{\delta}_s = \delta\exp(-w/3) \qquad \tilde{h}_s = h\exp(-w/3)\,,$$

where d_{s1} and d_{s2} are P-wave displacement and maximum velocity, respectively, and the covariates are defined in Table 1. The seismic error model assumes dynamic bias terms for each signature that modify $\beta_{sr,1}$ for each source and, additionally, each propagation path (within source). Both signatures are therefore simulated from Equation (6) with the random effects covariate matrices $\boldsymbol{Z}_{hir,j}$ and $\boldsymbol{Z}_{hijr}$ set to all-ones vectors of the appropriate length.

6.3 Optical

When a nuclear device is detonated above ground, a strong light flash is emitted. A special ("streak") camera allows for very fine time resolution to capture the pixel intensity of the light flash as a function of time. Figure 3 illustrates the time dependence of the pixel intensity for the nuclear test Jangle Sugar. Note the "double hump" optical signature, where the

first hump is related to the initial shock front and the second hump is related to thermal radiation. This double-hump shape is characteristic of nuclear explosions, and related data from optical sensors have been an essential component in detecting clandestine above-ground nuclear events (Scott, 1997).

From empirical observations and physical basis, two specific times — $t_{\min}$, the time of the minimum trough between the two humps and $t^{(2)}_{\max}$, the time of the maximum of the second hump — are related to yield. This is apparent in Figure 4, which is based on the unclassified data in Table 2 from seven 1950s-era near-surface nuclear tests, six of which were conducted in the Pacific and the other at NTS. In the table, optical data are provided courtesy of Ford et al. (2021), and surface effects data are from Circeo and Nordyke (1964).

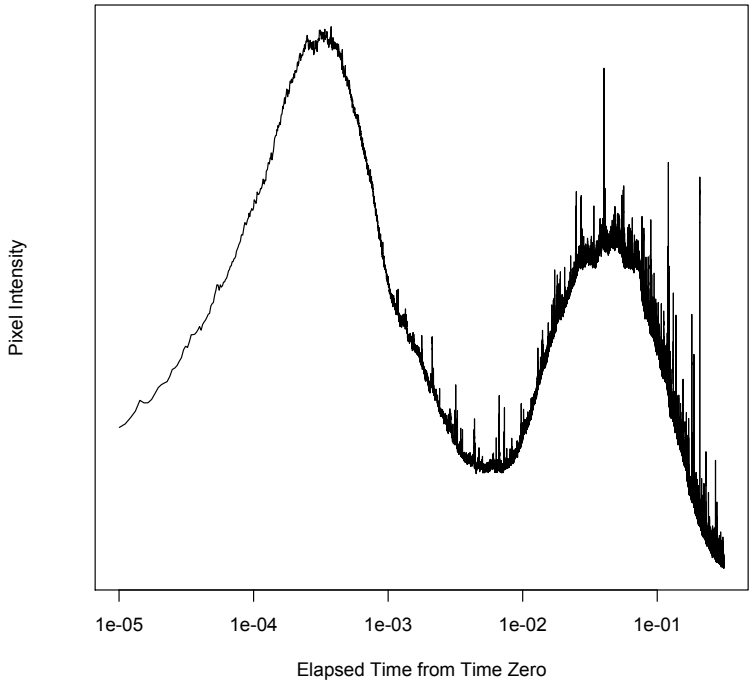

Figure 3: Light flash for Sugar.

Table 2: Optical and surface effects data for near-surface nuclear tests

Test Location	Sugar NTS	Cactus Eniwetok	LaCrosse Eniwetok	Koon Bikini	Zuni Bikini	Mike Eniwetok	Bravo Bikini
Yield (kt)	1.2	18	39.5	110	3530	10400	15000
Height of Burst (ft)	3.5	3	17	13.6	9	35	15.5
Optical t_{min} (sec)	0.006418	0.012	0.026	–	0.16	0.29	0.3315
Optical $t^{(2)}_{max}$ (sec)	0.057343	0.13	0.222	–	1.705	3.25	3.745
Crater Radius r (ft)	45	185	202	400	1240	2800	3000
Crater Depth d (ft)	21	36	44	40	115	164	240

Ford et al. (2021) proposed a optical forward model for optical signatures across a range of HOBs,

$$f_{or} = \log(\tilde{d}_{or}) = \beta_{or,1} + \beta_{or,2}\exp(-|\tilde{h}_o|) + 100(\tilde{h}_o + 0.1) - \log(1 + \exp(100(\tilde{h}_o + 0.1)))\,.$$

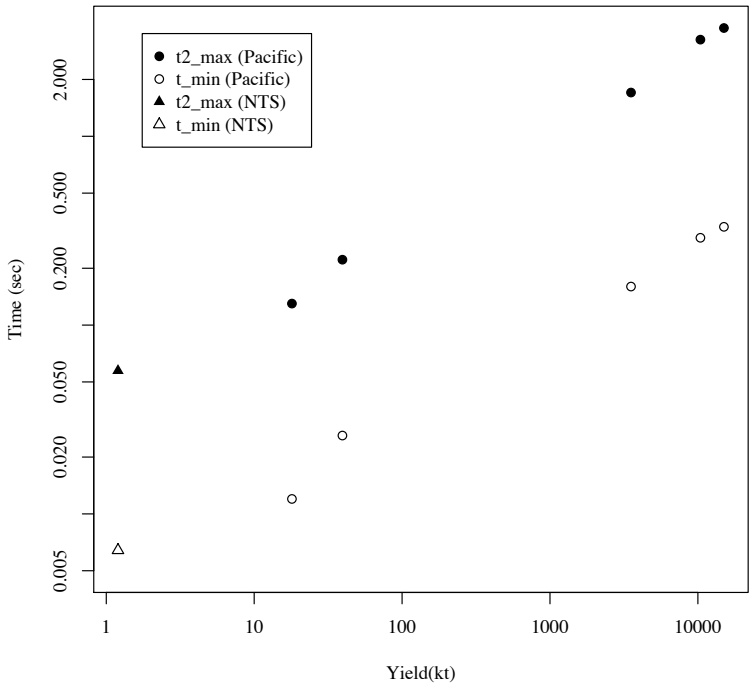

Figure 4: Optical data for near-surface events in Table 2.

The scaled signatures and covariates of this forward model are given by

$$\tilde{d}_{o1} = t_{\min} \exp(-w/3)(P/P_0)^{-1/3}(T/T_0)^{1/2}$$
$$\tilde{d}_{o2} = t^{(2)}_{\max} \exp(-w/3)(P/P_0)^{-1/3}(T/T_0)^{1/2}$$
$$\tilde{h}_o = h \exp(-w/3)(P/P_0)^{1/3}\,,$$

with covariates w and h defined in Table 1. For each nuclear event, only a single observation of each optical signature is available. Therefore, source and path dynamic bias terms cannot be estimated, so the optical signatures used in this report are simulated from Equation (8).

6.4 Surface Effects

For a nuclear device detonated above ground and but at sufficiently low heights, a crater is formed upon explosion. Crater dimensions are strongly related to yield. Empirically, the dimensions of a crater caused by an acoustic shock front on the earth for a near-ground explosion are log-linearly related to the yield of the explosion.

Data from Table 2 for 1950s-era nuclear tests conducted above ground were published in several open literature sources (shortly thereafter, the U.S. ceased open-air nuclear testing, precluding availability of additional data for this type of event). Data taken from the events in Table 2 are plotted in Figure 5. Although these tests were not originally intended for yield estimation purposes, they have obvious value in that regard; note the strong relations between crater radius (more formally defined as the radius of a best-fitting circle to the crater lip position data) and depth with yield in log-log space. Empirical linearity in log-log scale

for monitoring metrics is a common phenomenon owing to various physics considerations (Stein and Wysession, 2013). Also note that these data do not conform to the idealized slope 1/3 that would follow from "cube root scaling" of yield as is assumed for the acoustic data.

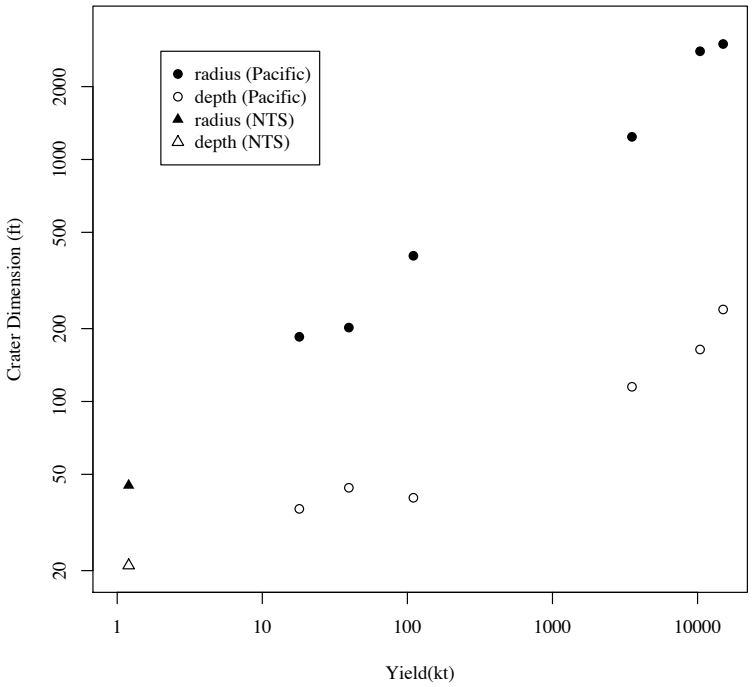

Figure 5: Surface effects data for near-surface events.

As with the optical signatures, crater dimensions are single measurements per event and are therefore simulated from Equation (8) for the analysis in this report.

For completeness, were the new event to occur below the ground surface in an attempt to avoid detection, subsidence and/or "throwout" craters would be formed when the earth collapses into cavities formed by the detonations and/or is ejected by the blast. The physical mechanism for such cratering is more complex for below-ground tests than for tests slightly above ground, although crater dimensions still behave as functions of yield and geology (Higgins, 1970). Such below-ground modeling is not relevant to the present above-ground application.

6.5 Methods

Data analysis consists of fitting one of the statistical models (6)-(8) to each signature collected from each sensor type. This involves selecting both forward and error models, a process that can be automated as described below. Once choices have been made, statistical inference for unknown model parameters is carried out by either *maximum likelihood* estimation or *Bayesian* techniques. The software is designed with the flexibility to accommodate both approaches.

These approaches require calculating the *log-likelihood function* derived in Appendix A, given by Equations (12) for the calibration data and (16) for the new event data, with

the joint (calibration and new event data) log-likelihood function given by Equation (18). When the errors-in-variables treatment of calibration data yields introduced in Section 3 is utilized, the log-likelihood function implied by Equation (3) is added to the joint log-likelihood function as given by Equation (20).

In the maximum likelihood (ML) approach, unknown model parameters are estimated by the values that maximize the joint log-likelihood function. Uncertainties in the maximum likelihood estimator (MLE) are described by the inverse of the *Fisher information matrix* (see, for example, Equation (19) of Appendix A for joint estimation of new event quantities of interest and statistical model parameters). The MLE and its estimated covariance matrix are used to form asymptotic *confidence intervals* (or *confidence regions*) for the unknown new event quantities of interest.

In the Bayesian approach, the unknown model parameters are assigned a *prior distribution* describing expert knowledge about the values of these parameters prior to being informed by any data. This distribution is updated by information contained in the calibration and new event data as described by the likelihood function, resulting in the *posterior distribution* of the unknown model parameters. This distribution is not generally available in closed form and must be sampled via methods such as Markov chain Monte Carlo (MCMC; Gelman et al. (2013)). Posterior samples of the new event quantities of interest are used to form *credible intervals* (or *credible regions*) that provide probabilistic bounds on their unknown true values.

Assessment of a new event consists of the following steps:

1. Determine the best-fitting model to calibration signatures from each sensor type.

2. Infer the quantities of interest describing a new event for each sensor type.

3. Fuse data across multiple sensor types to obtain an integrated new event inference.

The first step involves determining the proper forward model and error model combination for each sensor type. Different candidate models can be found in the literature, and the point of the example is simply to illustrate that the software is capable of accommodating them, and *not* to advocate that any particular model is uniquely "best" for future use.

In addition to a visual inspection of data fit and diagnostic plots, model fidelity can be formally evaluated by information criteria calculated from both maximum likelihood and Bayesian analyses. In particular, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are computed directly from the maximum likelihood estimate. AIC and BIC select the preferred model from the perspectives of predictive capability and asymptotic consistency, respectively (Akaike, 1973; Schwarz, 1978). BIC assumes the observed signatures are independent and identically distributed. This assumption is violated when dynamic bias terms are included in the statistical model (e.g. Equations (6) and (7)) or correlations are allowed among signatures (as in Appendix A). In these cases, an *effective sample size* may be used to calculate BIC (Shen and González, 2021).

The Deviance Information Criterion (DIC) and the Predictive Information Criterion (PIC) are computed directly from Bayesian posterior samples. These criteria are based on selecting a model that maximizes the posterior mean of the log-likelihood function, modified by a complexity penalty (Spiegelhalter et al., 2002; Ando, 2011). The PIC was introduced

to correct the tendency of DIC to select overfitted models (Ando, 2011). The AIC and PIC are preferred for selection of models when the focus is on out-of-sample predictive capability. The chosen information criterion is calculated for each candidate model, and the statistical model associated with the smallest value is selected.

In Section 5.2 and Appendix A, the parameter vector $\boldsymbol{\theta}_0$ collects the quantities of interest for new event inference. In the example application above, $\boldsymbol{\theta}_0 = (w, h)$ (log yield and HOB/DOB). The next subsection presents results on model selection and statistical inference for $\boldsymbol{\theta}_0$ in this application.

6.6 Analysis

The acoustic and seismic calibration data of Ford et al. (2021) were collected from chemical explosions conducted in three different emplacements — soft, hard, and wet rock. The forward models for these sensor types have the same form across emplacements; however, some or all of the model coefficients are allowed to depend on emplacement as summarized in Table 3. The optical data of Ford et al. (2021) and the surface effects data of Section 6.4 were obtained from nuclear explosions without regard to a designated emplacement. As discussed earlier in this section, the structure of these data sets informs the simulated data utilized in this report for all sensor types.

Table 3: Status of acoustic and seismic forward model coefficients

Sensor Type	Emplacement-Dependent	Common	Total Parameter Count
Acoustic	$\beta_{ar,3}$	$(\beta_{ar,1}, \beta_{ar,2})$	5
Seismic	$(\beta_{sr,1}, \beta_{sr,2}, \beta_{sr,3}, \beta_{sr,4}, \beta_{sr,5})$		15

The model selection criteria DIC and PIC depend on the prior distributions assigned in the Bayesian analysis for the forward and error model parameters. The forward models presented above for each sensor type are the only options considered for this application. The forward model coefficients are assigned mutually independent improper uniform priors on the real line, *except* for the emplacement-dependent seismic coefficients $\beta_{sr,3}$, which are restricted to the negative real line.

Generally speaking, the use of specific Bayesian prior distributions — similar to the use of specific forward models for sensor types — requires justification and empirical checking. As noted earlier, Bayesian priors should embody expert knowledge regarding model parameters of interest. Thus, the justification for realistic informative priors in the present application would likely be classified. Consequently, the priors to follow are chosen solely for the purpose of demonstrating the capabilities of the software and should not be interpreted more broadly.

In the application to follow, improper priors are used to illustrate capabilities of the software. In real applications, extreme care must be taken in assigning so-called noninformative prior distributions to nonlinear forward model coefficients such as those for acoustic data. Not only do such priors completely mischaracterize available expert knowledge, but relatively flat priors on $\boldsymbol{\beta}_{hr}$ in model coefficient space *do not imply* flat priors on $\boldsymbol{f}_{hijr}(\boldsymbol{\beta}_{hr}, \boldsymbol{v}_{hij})$ in nonlinear curve-fitting space. As is well known (Seaman et al., 2012), any use of so-called noninformative priors here requires that diagnostic checks are made in practice to ensure that misleading posterior distributions are not produced.

In the development below, the error model coefficient matrices $\boldsymbol{Z}_{hir,j}$ and $\boldsymbol{Z}_{hijr}$ are taken to be all-ones vectors when fitting error models that include dynamic bias terms (see Equations (6) and (7)). Therefore, the covariance matrices of these dynamic bias terms in Equation (5) reduce to scalar variances. The standard deviations are denoted $\sigma_{hr}^{(S)}$ and $\sigma_{hr}^{(P)}$ and assigned mutually independent half-Cauchy distributions of the form

$$\pi(x|A) = \frac{2}{\pi A}\left(1 + \left(\frac{x}{A}\right)^2\right)^{-1}$$

for $A = 20$ (Gelman, 2006). The observational error covariance matrix $\boldsymbol{\Sigma}_h$ (see Section 5.1) can be decomposed as $\boldsymbol{\Sigma}_h = \boldsymbol{S}_h \boldsymbol{C}_h \boldsymbol{S}_h$, where $\boldsymbol{S}_h$ is the diagonal matrix of standard deviations $\sqrt{\sigma_{hrr}}$, and $\boldsymbol{C}_h$ is the *correlation matrix* consisting of ones on the diagonal and correlations $\rho_{hr_1r_2} = \sigma_{hr_1r_2}/\sqrt{\sigma_{hr_1r_1}\sigma_{hr_2r_2}}$ off-diagonal. The variances are assigned mutually independent improper inverse gamma distributions, $\pi(\sigma_{hrr}) \propto 1/\sigma_{hrr}$, while the correlation matrices are assigned mutually independent Lewandowski-Kurowicka-Joe priors,

$$\pi(\boldsymbol{C}_h|\eta) \propto [\det(\boldsymbol{C}_h)]^{\eta-1}, \qquad \boldsymbol{C}_h \text{ a correlation matrix},$$

for $\eta = 1$ (the uniform distribution on correlation matrices, Lewandowski et al. (2009)). The prior distributions for all groups of parameters (forward model coefficients, dynamic bias variances, observational error variances, and observational error correlations) are taken to be mutually independent. This specification of joint prior distributions is motivated by a desire to impose minimal *a priori* information on the statistical parameters of the forward and error models, with the caveat as above that these choices do not generally imply uniform probability across model realizations.

Posterior sampling is conducted via the delayed rejection adaptive Metropolis algorithm of Haario et al. (2006), using the `R` package `FME`. The adaptation operates by sequentially updating the proposal covariance matrix at user-specified intervals during the initial phase of the sampling to achieve efficient mixing. A single delayed rejection step is allowed at each iteration, reducing the variance of estimators calculated from the final sample. The first 10,000 samples are discarded, and 20,000 production samples are taken and decorrelated by thinning out the chain every 20-th sample.

Because only a single forward model is considered for each sensor type, all model selection is conducted with regard to the choice of error model. Only Equation (8) is eligible for fitting the signatures of the optical and surface effects sensor types.

For acoustic, only Equations (7) and (8) may be considered to model the signatures, as there are multiple observations but only one propagation path per source (as in the Ford et al. (2021) data). The coefficient matrices $\boldsymbol{Z}_{hir,j}$ are taken to be all-ones vectors when fitting error models that include source dynamic bias. Table 4 presents calculated values for the model selection criteria discussed in the previous subsection. In this table, a check indicates the presence of the indicated dynamic bias term in the error model for the indicated signature. For each criterion, **bold** font indicates the error model choices for each signature representing the optimum out of all nine alternatives considered. For all criteria, selecting Equation (7) to model both acoustic signatures is optimal, confirming the presence of random effects, as is consistent with the simulation scenario described in Section 6.1.

For seismic, each of Equations (6)–(8) may be considered for each signature, as there are multiple measurements per source, and at least two propagation paths within several sources

Table 4: Model selection results for acoustic sensor type

Source Dynamic Bias					
f_{a1}	f_{a2}	AIC	BIC	DIC	PIC
✓	✓	**−3578.6**	**−3544.5**	**−3577.0**	**−3561.8**
✓		−2970.1	−2912.1	−2968.8	−2954.7
	✓	−3175.3	−3117.6	−3175.1	−3161.5
		−2570.1	−2511.3	−2569.6	−2556.4

possessing multiple measurements (this is only true for the simulated scenario considered in this report; the data of Ford et al. (2021) contain only one propagation path per source). The coefficient matrices $\boldsymbol{Z}_{hir,j}$ and $\boldsymbol{Z}_{hijr}$ are taken to be all-ones vectors when fitting error models that include dynamic bias terms.

Table 5 presents calculated values for the model selection criteria discussed in the previous subsection. For all criteria, selecting Equation (6) to model both seismic signatures is optimal, again consistent with the simulation scenario described in Section 6.2.

Table 5: Model selection results for seismic sensor type

Dynamic Bias Terms							
Source		Path					
f_{s1}	f_{s2}	f_{s1}	f_{s2}	AIC	BIC	DIC	PIC
✓	✓	✓	✓	**−2124.5**	**−2010.7**	**−2169.5**	**−2155.5**
✓	✓	✓		−2109.6	−1999.2	−2143.6	−2125.1
✓	✓		✓	−2080.4	−1968.4	−2127.7	−2115.7
✓	✓			−2061.8	−1953.0	−2097.5	−2081.2
✓		✓		−2030.8	−1887.7	−2083.2	−2075.2
✓				−1983.7	−1844.2	−2032.0	−2022.3
	✓		✓	−1990.4	−1847.3	−2034.4	−2022.1
	✓			−1961.0	−1822.3	−2016.0	−2009.9
				−1882.0	−1729.2	−1936.4	−1929.9

The "new event" for which the quantities of interest $\boldsymbol{\theta}_0 = (w, h)$ are inferred is a simulated version of the above-ground nuclear test Sugar. This type of low-yield event is not representative of an evade-detection scenario: being detonated above ground, its optical signal would be promptly detected and the environmental dispersal of radioactive debris that could be sampled post-detonation would be highly incriminating. Instead, this event is more representative of a terrorist act in an urban area, especially where acoustic sensors are placed in advance at local distances.

For the event Sugar, established values (Table 2) are $w = \log(1.2 \text{ kt})$ and $h = 3.5$ ft. In the following, h will be reported in meters for consistency with Ford et al. (2021). The value of h is restricted to the interval $[-10, 160]$ — roughly the envelope of values (in meters) represented in the calibration data across all sensor types — under the presumption that the forward models are only appropriate for near-surface events in this HOB/DOB range.

In the Bayesian analyses, two prior specifications are considered for $\boldsymbol{\theta}_0$. The first assigned

an improper uniform distribution on the real line to w and a proper uniform distribution on the restricted HOB/DOB range to h. The second assigned independent Gaussian distributions to w and h,

$$w \sim \mathcal{N}(\mu_w, \sigma_w^2) \text{ and } h \sim \mathcal{N}(0, \sigma_h^2)\,,$$

the latter being truncated to the restricted interval for HOB/DOB. This prior for w assumes that log yield is centered in the interval between 10 tons and 10 megatons on the log scale ($\mu_w = 4\log(10)$), and with high probability, restricted to this interval ($\sigma_w = \log(10)$). This prior for h is centered at ground-level, reflecting the applicability of the sensor type forward models to near-surface explosions, with values becoming less likely as the upper HOB/DOB limit is approached ($\sigma_h = 160/3$).

An assumption regarding the amount of energy from a chemical versus a nuclear explosion that is converted into seismic and acoustic signatures is necessary to combine signatures across explosion types for conducting new event characterization. The 2-to-1 chemical to nuclear equivalency as per Ford et al. (2021) is adopted here for illustration; the corresponding relative measurement uncertainty for the errors-in-variables yield estimation is arbitrarily assumed to be 10%. Again, this example is carried out to illustrate capabilities of the developed software and should not be taken literally.

In order to avoid the difficulties with mixing and convergence of posterior sampling algorithms often associated with high-dimensional parameter spaces such as those encountered here, only the new event quantities of interest $\boldsymbol{\theta}_0$ are sampled, while all other parameters are fixed at their posterior mode. Table 6 presents ML and Bayesian inference for $\boldsymbol{\theta}_0$ based on the simulated calibration and new event data from each individual sensor type. The improper uniform and Gaussian priors are indicated by 'U' and 'G' respectively.

The ML and Bayesian results are generally consistent with the true values of w and h used to simulate data for the new event. In expressing uncertainties, note that the usual 95% confidence interval (i.e., log estimate $\pm$ 2σ) in the scale of log yield corresponds to a multiplicative confidence interval for yield,

$$(\text{log estimate} - 2\sigma, \text{log estimate} + 2\sigma) \;\Leftrightarrow\; (\text{estimate}/e^{2\sigma}, e^{2\sigma} \times \text{estimate}),$$

where the multiplicative factor $e^{2\sigma}$ reflects uncertainty relative to the estimated value. These factors are listed in Table 6 together with the corresponding yield estimates. For example, the ML acoustic yield estimate in Table 6 has a corresponding uncertainty factor of 1.33, or roughly a 33% relative error as based on the limited calibration data. For HOB/DOB, 2σ uncertainty is provided parenthetically, and correlations between w and h (or their MLEs) are also indicated.

The seismic ML estimator of (w, h) has extremely large uncertainty. As is shown later (Figure 7), the seismic model used herein simply can't navigate the yield-HOB tradeoff: a lower-yield event at ground level gives a similar seismic signature as does a higher-yield event detonated well above the ground, phenominologically similar to the way that firing a device in a large underground cavity can depress a seismic magnitude relative to a fully-coupled event. Other sensor types are less sensitive to the yield-HOB tradeoff in this example.

Conversely, the seismic Bayes estimate of w is biased high slightly for the improper uniform prior and inference about h has large uncertainty. The posterior means of w are all slightly smaller for the Gaussian prior than for the improper uniform prior, due to μ_w being

approximately two standard deviations smaller than the reported Sugar yield. Because σ_h is large relative to the reported Sugar HOB, this shrinkage effect in h is noted particularly for the seismic sensor type and is negligible for the optical and acoustic sensor types.

Table 6: Individual sensor type ML and Bayesian inference for new event parameters.

Method	Prior	Sensor Type	Yield [kt] (RE)	HOB/DOB [m]	Correlation
ML		Acoustic	1.22 (33%)	−1.0 (17.1)	−0.84
ML		Seismic	1.94 ($\approx +\infty$)	55.5 (16751)	1.00
ML		Optical	1.04 (119%)	0.0 (14.3)	−0.95
ML		Crater	1.97 (122%)		
Bayes	U	Acoustic	1.19 (28%)	1.6 (14.3)	−0.78
Bayes	U	Seismic	2.05 (57%)	78.8 (98.6)	0.92
Bayes	U	Optical	1.38 (57%)	2.4 (14.2)	0.48
Bayes	U	Crater	2.00 (67%)		
Bayes	G	Acoustic	1.17 (28%)	2.2 (14.4)	−0.76
Bayes	G	Seismic	1.62 (53%)	30.9 (60.6)	0.93
Bayes	G	Optical	1.32 (52%)	1.3 (12.4)	0.33
Bayes	G	Crater	1.70 (68%)		

Figure 6 shows the fitted versus simulated data for both seismic signatures. Fitted signatures are obtained by replacing unknown parameters with their ML estimate in the seismic forward models and calculating the model results for each signature. These plots show substantial agreement between these data, as expected from the design of this simulation example. In other words, the ML estimate of (w, h) predicts the data used to train it well. However, the forward models do not allow h to estimated precisely even using data with small errors as in this example. This is shown in Figure 7, where the profile log-likelihood in (yield, HOB/DOB) is plotted. The nearly vertical contours suggest that for any given yield, the maximized log-likelihood is nearly constant in h. The large uncertainty in the ML estimator of yield evaluated at the ML estimate is more difficult to explain. It is noted that the profile ML estimator evaluated at the true parameter value possesses much less uncertainty in log yield (standard deviation of 1.01 versus 43.15), while the likelihood value is nearly the same (1107.9 vs. 1108.3), suggesting that stable ML inference for the seismic models requires data with even smaller errors than those utilized in this example.

Table 7 presents maximum-likelihood (ML) and Bayesian inference for $\boldsymbol{\theta}_0$ based on the simulated calibration and new event data integrated across all four sensor types (Multi-PEM). Relative uncertainty (roughly 11%) in the ML yield estimate is substantially smaller than any of the values obtained from the individual sensor types alone, as are the posterior standard deviations under both prior assumptions, highlighting the potential of integrated analysis to meaningfully reduce uncertainty in the assessment of critical new event parameters. Admittedly, the context here is somewhat optimistic, based on data simulated from a known model, but clearly illustrates the potential of combining multiple sensor types.

Figure 8 shows the marginal posterior distributions of yield and HOB/DOB for each individual sensor type and across all four sensor types assuming the Gaussian prior. Although all four marginal distributions corresponding to each sensor type are probabilistically con-

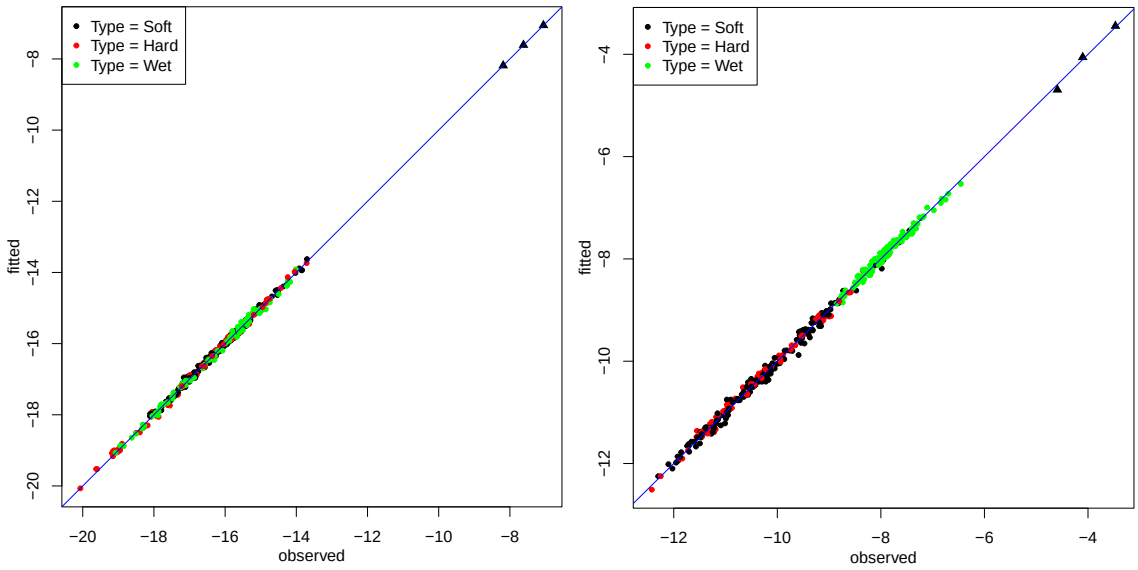

Figure 6: Fitted versus simulated seismic signatures for log P-wave displacement (left) and log maximum velocity (right). Calibration data (circles) and new event data (triangles) are shown by emplacement (soft, hard, or wet rock).

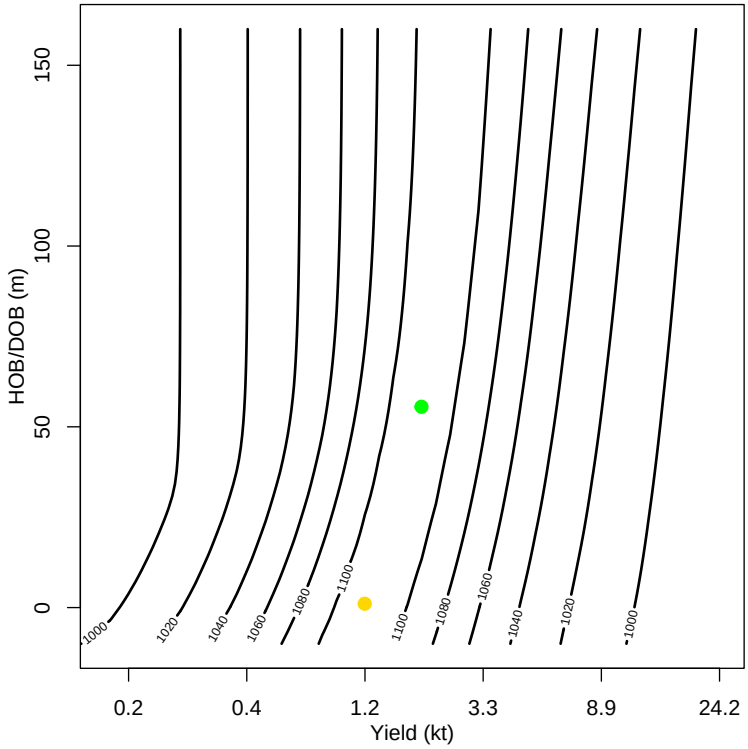

Figure 7: Profile log-likelihood in (yield, HOB/DOB) for seismic signatures. ML estimate (green dot) and true value (gold dot) are plotted for reference.

sistent with the reported Sugar values (gold vertical dashed lines), the seismic distributions are biased high for both yield ($\exp(w)$) and h, as is the surface effects distribution for yield.

Table 7: Multiple sensor type ML and Bayesian inference for new event parameters.

Method	Prior	Yield [kt] (RE)	HOB/DOB [m]	Correlation
ML		1.23 (11%)	−2.5 (4.3)	−0.37
Bayes	U	1.23 (9%)	−0.7 (5.8)	0.00
Bayes	G	1.22 (10%)	−0.5 (5.7)	−0.01

The highest posterior density intervals of the acoustic and optical marginal distributions are aligned more closely with the Sugar values. The MultiPEM marginal posterior distributions for yield and h are more tightly concentrated than any of the corresponding individual sensor type marginal posterior distributions. This results from the integration of signatures across all four sensor types in a way that renders univariate marginal posterior distributions on a tighter volume of (yield, HOB/DOB)-space defined by the probabilistic overlap of the individual sensor type solutions (the "sweet spot"). This example illustrates a desired outcome of MultiPEM analysis, namely that the integrated solution is (with high probability) focused on underlying reality even when one or more of the individual solutions is not particularly informative.

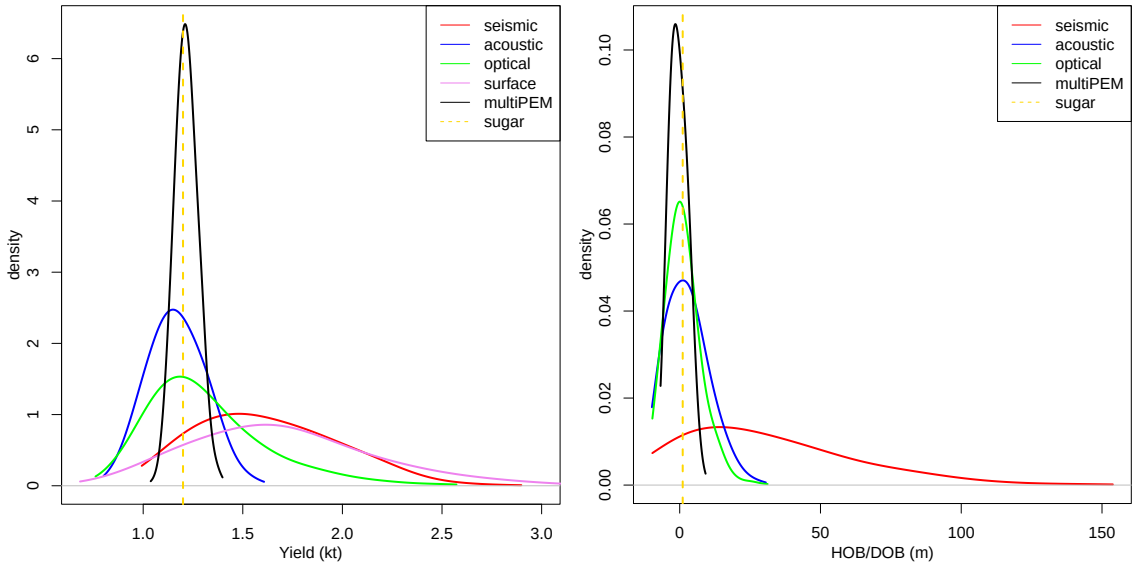

Figure 8: Marginal posterior distributions of yield and HOB/DOB for each sensor type and across all four sensor types assuming the Gaussian prior.

Figure 9 shows joint posterior samples of yield and HOB/DOB for each individual sensor type and across all four sensor types for both prior assumptions. The "sweet spot" described in the previous paragraph is clearly indicated by the MultiPEM joint posterior samples, which are seen to be focused on the region of (yield, HOB/DOB)-space containing the Sugar values (gold dot). The seismic solution is not particularly informative about the values of h consistent with the observed signatures, as could be expected by examination of the profile log-likelihood function in Figure 7. The optical samples branch into "above-ground" and "below-ground" solutions for yields exceeding 1.2 kt, where surface solutions are inconsistent with the observed signatures. Differences between the joint posterior distributions for the

improper uniform and Gaussian priors are negligible for the most data-dominant MultiPEM solution. However, for the Gaussian prior, shrinkage towards the prior means of yield and h can be seen in both the seismic and optical solutions.

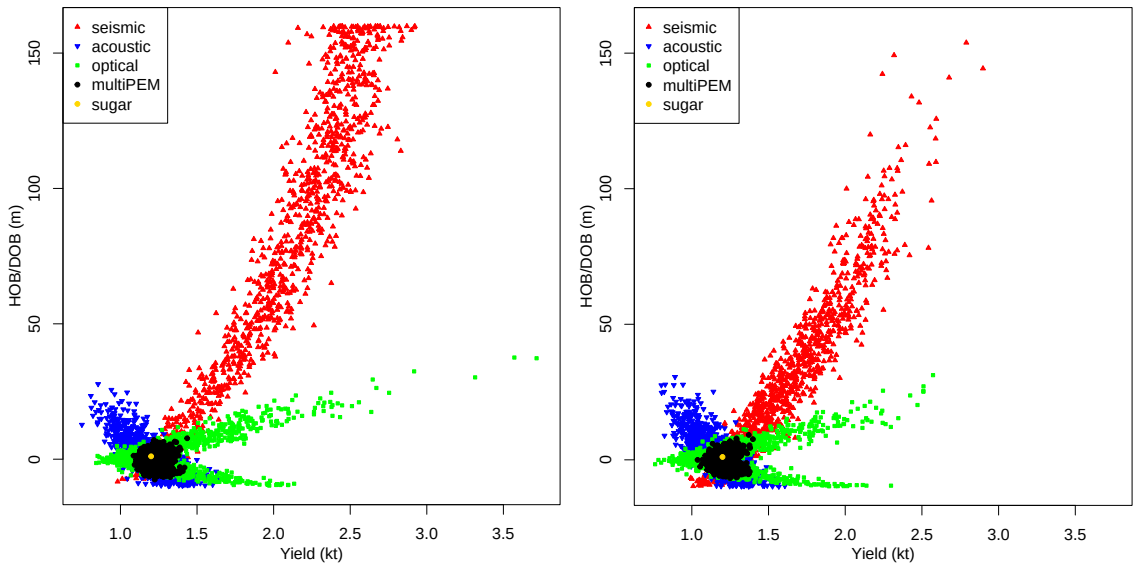

Figure 9: Joint posterior samples of yield and HOB/DOB for each sensor type and across all four sensor types assuming the improper uniform prior (left) and the Gaussian prior (right).

The errors-in-variables approach of Section 3 was implemented for the log yields of the optical and surface effects calibration events by assuming their reported yields were correct up to 10%. That is, the value of σ_s in Equation (3) was set to 0.1/3. This had negligible impact on the quality of regression coefficient estimates as measured by median absolute difference between coefficient ML estimates and true values. For Bayesian analysis, the unknown yield covariate values pertaining to each unique calibration source were independently assigned the *flexible generalized skew-normal* prior distribution of Hossain and Gustafson (2009), given by

$$\pi(x|\mu,\lambda,\upsilon_1,\upsilon_2) = \frac{2}{\lambda}\phi\left(\frac{x-\mu}{\lambda}\right)\Phi\left[\sum_{i=1}^{2}\upsilon_i\left(\frac{x-\mu}{\lambda}\right)^{2i-1}\right]$$

with $\phi(\cdot)$ and $\Phi(\cdot)$ the standard normal density and distribution functions. Regression coefficient estimates are sensitive to the prior choice of unknown covariate distribution(s), and the choice adopted here allows for skewness, bimodality, and heavy tailedness to give a higher degree of flexibility. The hyper-parameters $(\mu,\lambda,\upsilon_1,\upsilon_2)$ are independently assigned the default prior distributions in Section 4.2 of Hossain and Gustafson (2009).

Table 8 presents ML and Bayesian inference for $\boldsymbol{\theta}_0$ from the MultiPEM analysis that incorporates errors-in-variables. These results are consistent with those of Table 7, as expected due to the minimal effect of accounting for errors-in-variables under the above assumption of 10% relative uncertainty in calibration yields.

Table 8: Multiple sensor type ML and Bayesian inference for new event parameters incorporating errors-in-variables for optical and surface effects calibration event yields.

Method	Prior	Yield [kt] (RE)	HOB/DOB [m]	Correlation
ML		1.22 (11%)	−2.6 (3.9)	−0.44
Bayes	U	1.22 (10%)	−0.4 (5.9)	0.02
Bayes	G	1.22 (10%)	−0.8 (5.8)	−0.02

7 Discussion

Any signal from an explosion that can be represented as generated by a source model and propagated to a sensor can be reasonably described by a base model. The worked example in this report (the event Sugar) is intended to illustrate the great flexibility in modeling and data analysis offered by the software that has been developed; the corresponding particulars should *not* be misconstrued as being an endorsement of specific forward source models nor of specific frequentist or Bayesian data analysis techniques.

Benefits of combining results across multiple sensor types are greatest when the sensors have complementary strengths and weaknesses because estimates can be obtained across a wide range of scenarios. When sensors provide largely redundant information, precision is improved for some scenarios, but there may be limited sensitivity for other scenarios. An obvious recommendation is to build as robust a calibration data base as is practical in order to reduce turnaround times for data analyses by thorough analysis of calibration data in advance of the new events.

8 Acknowledgments

All authors contributed written content to the manuscript. The first author thanks Sean Ford of Lawrence Livermore National Laboratory for considerable personal communications regarding Ford et al. (2021) and for providing the data used therein, and Josh Carmichael of Los Alamos National Laboratory for providing Figure 1. This research was funded by the Office of the Assistant Secretary of Defense and supported by the U.S. Department of Energy through the Los Alamos National Laboratory. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001).

References

Akaike, H. (1973), Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrov, B. N. & Csaki, F., Eds., International Symposium on Information Theory, 267-281.

Anderson, D. N., Walter, W. R., Fagan, D. K., Mercier, M., & Taylor, S. R. (2009), Regional multistation discriminants: Yield, distance, and amplitude corrections, and sources of error, *Bull. Seismol. Soc. Am.* **99**, 749-808.

Ando, T. (2011), Predictive Bayesian model selection, *Am. J. Math. Manag. Sci.* **31**, 13-38.

Circeo, L. J. & Nordyke, M. D. (1964), Nuclear Cratering Experience at the Pacific Proving Grounds, Lawrence Radiation Laboratory Technical Report UCRL-12172.

Efron, B. & Hinkley, D. V. (1978), Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information, *Biometrika* **65**, 457-487.

Efron, B. (1982), Maximum likelihood and decision theory, *Ann. Statist.* **10**, 340-356.

Fagan, D. K., Taylor, S. R., Schult, F. R., & Anderson, D. N. (2009), Using ancillary information to improve hypocenter estimation: Bayesian single event location (BSEL), *Pure Appl. Geophys.* **166**, 521-545.

Ford, S. R., Bulaevskaya, V., Ramirez, A., Johannesson, G., & Rodgers, A. J. (2021), Joint Bayesian inference for near-surface explosion yield and height-of-burst, *J. Geophys. Res. Solid Earth* **126**, e2020JB020968.

Galton, F. (1879), The geometric mean, in vital and social statistics, *Proc. R. Soc. Lond.* **29**, 365-367.

Gelman, A. (2006), Prior distributions for variance parameters in hierarchical models, *Bayesian Anal.* **1**, 515-533.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013), *Bayesian Data Analysis*, CRC Press, Boca Raton, ISBN: 978-7-5192-6181-8.

Haario, H., Laine, M., Mira, A., & Saksman, E. (2006), DRAM: Efficient adaptive MCMC, *Stat. Comput.* **16**, 339-354.

Higgins, G. H. (1970), Underground Nuclear Explosions, Symposium on Engineering with Nuclear Explosives, Las Vegas, Nev., *Proceedings, CONF-700101*, 29-42.

Hossain, S. & Gustafson, P. (2009), Bayesian adjustment for covariate measurement errors: A flexible parametric approach, *Statist. Med.* **28**, 1580-1600.

Levi, M. D. (1973), Errors in the variable bias in the presence of correctly measured variables, *Econometrica* **41**, 985-986.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009), Generating random correlation matrices based on vines and extended onion method, *J. Multivar. Anal.* **100**, 1989-2001.

McAlister, D. (1879), The law of the geometric mean. *Proc. R. Soc. Lond.* **29**, 367-376.

Picard, R. & Bryson, M. (1992), Calibrated seismic verification of the Threshold Test Ban Treaty, *J. Am. Stat. Assoc.* **87**, 293-299.

Pinheiro, J. C. & Bates, D. M. (2020), *Mixed-Effects Models in S and S-PLUS*, Springer-Verlag, New York, ISBN: 0-387-98957-9.

Schwarz, G. (1978), Estimating the dimension of a model, *Ann. Stat.* **6**, 461-464.

Scott, W. B. (1997), Admission of 1979 nuclear test finally validates Vela data, *AW&ST* **147**, 33.

Seaman, J. W. III, Seaman, J. W. Jr., and Stamey, J. D. (2012), "Hidden Dangers of Specifying Noninformative Priors," *Am. Stat.*, **66**, 77-84.

Shen, N. & González, B. (2021), Bayesian information criterion for linear mixed-effects models, arXiv:2104.14725[v1] [stat.AP] 30 April 2021.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002), Bayesian measures of model complexity and fit (with discussion), *J. R. Stat. Soc. Ser. B* **64**, 583-639.

Stein, S. & Wysession, M. (2013), *An Introduction to Seismology, Earthquakes, and Earth Structure*, John Wiley & Sons, New York, ISBN: 978-1-118-68745-1.

Taylor, S. R. (2007), Isolation of Regional Seismic Phase Source, Path, and Receiver Effects Using Analysis of Variance, Rocky Mountain Geophysics RMG-2007-001.

Williams, B. J., Brug, W. P., Casleton, E. M., Syracuse, E. M., Blom, P. S., Meierbachtol, C. S., Stead, R. J., MacLeod, G. A., Bauer, A. L., Shao, X.-M., & Anderson, D. N. (2021), Multiphenomenology explosion monitoring (MultiPEM): a general framework for data interpretation and yield estimation, *Geophys. J. Int.* **226**, 14-32.

A Explosive Device Data Modeling

This appendix provides details on the statistical modeling of calibration and new event data for the purpose of inferring unknown quantities of interest characterizing a new event (such as event time, surface location, height-of-burst (HOB)/depth-of-burial (DOB), and yield) with rigorous uncertainty quantification. The related material to follow is, necessarily, heavily mathematical and notationally intensive.

A.1 Statistical Modeling of Calibration Event Data

In the following, let h index the various sensor types (seismic, acoustic, ...), i represent the explosion sources, j represent the source-to-sensor path level, r represent the sensor measurement type, and t_{hi} represent the emplacement state of source i measured by sensor type h (surface event, underground test with a certain geology, ...). Consider the following two-level nonlinear mixed effects model for $n_{hijr} \times 1$ response vector $\boldsymbol{y}_{hijr}$:

$$\boldsymbol{y}_{hijr} = \boldsymbol{f}_{hr}((\boldsymbol{\beta}_{h0r}, \boldsymbol{\beta}_{ht_{hi}r}), \boldsymbol{v}_{hij}) + \boldsymbol{Z}_{hir,j}\boldsymbol{b}_{hr}^{(S)} + \boldsymbol{Z}_{hijr}\boldsymbol{b}_{hr}^{(P)} + \boldsymbol{\varepsilon}_{hijr}\,, \tag{9}$$

for $h = 1, \ldots, M$, $i = 1, \ldots, M_h$, $j = 1, \ldots, M_{hi}$, $r = 1, \ldots, R_h$, and $t_{hi} \in \{1, \ldots, T_h\}$. The unknown model parameters $\boldsymbol{\beta}_{h0r}$ are common to every source i within sensor type h, while the parameters $\boldsymbol{\beta}_{ht_{hi}r}$ are common to a set of sources $\mathcal{S}_{ht_{hi}}$ to which source i belongs.

Specifically, there are assumed to be T_h such disjoint sets of sources, where $\mathcal{S}_{h1} \cup \cdots \cup \mathcal{S}_{hT_h}$ partitions the M_h sources by emplacement condition. If $T_h \leq 1$, it is presumed that a common set of parameters pertains for all sources, and the parameters $\boldsymbol{\beta}_{ht_{hi}r}$ are eliminated from (9). If $T_h > 1$ and there are no common parameters across all sources, then the parameter $\boldsymbol{\beta}_{h0r}$ is eliminated from (9). Per the main text, random effects are treated as having Gaussian distributions,

$$\begin{aligned}
\boldsymbol{b}_{hr}^{(S)} &\sim \mathcal{N}(\boldsymbol{0}_{q_{S,hr}}, \boldsymbol{\Sigma}_{hr}^{(S)}) \\
\boldsymbol{b}_{hr}^{(P)} &\sim \mathcal{N}(\boldsymbol{0}_{q_{P,hr}}, \boldsymbol{\Sigma}_{hr}^{(P)}) \\
\boldsymbol{\varepsilon}_{hijr} &\sim \mathcal{N}(\boldsymbol{0}_{n_{hijr}}, \boldsymbol{\Sigma}_{hijr}^{(\varepsilon)})\,,
\end{aligned}$$

where $\boldsymbol{\Sigma}_{hijr}^{(\varepsilon)} = \sigma_{hrr}\mathbf{I}_{n_{hijr}}$. Subsequently, the covariance matrix $\boldsymbol{\Sigma}_h = (\sigma_{hr_1r_2})$ is symmetric and positive-definite across the R_h sensor measurement types. Let $\boldsymbol{\varsigma}_h$ denote the v_h-vector of variance parameters in (9) aggregated across measurement types r for sensor type h.

For sensor type h, index set $\mathcal{S}_{ht}$ designates a set of sources assumed to have common emplacement condition t. For example, suppose seismic ($h = 1$) data are collected from sources subject to $T_1 = 3$ distinct emplacement conditions defined by rock type. In this case, the unknown model parameters are allowed to be specific to rock type, i.e. ($\boldsymbol{\beta}_{11r}$, $\boldsymbol{\beta}_{12r}$, $\boldsymbol{\beta}_{13r}$) for rock types '1', '2', and '3'. The dataset may contain any level of source overlap across sensor types (including none). The vector $\boldsymbol{v}_{hij}$ collects the known/measured covariates utilized in the model for source i and path j, such as origin time, location, HOB/DOB, and yield (which are independent of source to sensor path).

The $q_{S,hr} \times 1$ first level random effect vectors $\boldsymbol{b}_{hr}^{(S)}$ are mutually independent, as are the $q_{P,hr} \times 1$ second level random effect vectors $\boldsymbol{b}_{hr}^{(P)}$. Furthermore, the first and second level

random effect vectors are mutually independent. The covariate matrices $\boldsymbol{Z}_{hir,j}$ and $\boldsymbol{Z}_{hijr}$ are $n_{hijr} \times q_{S,hr}$ and $n_{hijr} \times q_{P,hr}$, respectively. Finally, the elements of each $n_{hijr} \times 1$ error vector $\boldsymbol{\varepsilon}_{hijr}$ are mutually independent and also independent of both levels of random effect vectors.

The statistical model (9) assumes the presence of first and second level random effects for all sensor measurement types. For each measurement type r, reduced random effect models are also considered, obtained by replacing the term $\boldsymbol{Z}_{hir,j}\boldsymbol{b}_{hr}^{(S)} + \boldsymbol{Z}_{hijr}\boldsymbol{b}_{hr}^{(P)}$ by (i) only first-level random effects $\boldsymbol{Z}_{hir,j}\boldsymbol{b}_{hr}^{(S)}$, or (ii) eliminating this term entirely resulting in a fixed effects model. In the subsequent discussion, the ordered sets $\mathcal{R}_h^S = \{r_{S,1}, r_{S,2}, \ldots, r_{S,s_{S,h}}\}$ and $\mathcal{R}_h^P = \{r_{P,1}, r_{P,2}, \ldots, r_{P,s_{P,h}}\}$ index the measurement types that incorporate first and second level random effect terms in their statistical models. Note that $\mathcal{R}_h^P \subset \mathcal{R}_h^S$ and one or both index sets may be empty.

Collating information across the M_{hi} source-to-sensor paths and denoting the unknown model parameter vectors by $\boldsymbol{\beta}_{hr}^{(t)} = (\boldsymbol{\beta}_{h0r}, \boldsymbol{\beta}_{htr})$, consider the following vectors,

$$\begin{aligned}
\boldsymbol{y}_{hir} &= \begin{bmatrix} \boldsymbol{y}_{hi1r}^\top & \boldsymbol{y}_{hi2r}^\top & \cdots & \boldsymbol{y}_{hiM_{hi}r}^\top \end{bmatrix}^\top \\
\boldsymbol{f}_{hir}(\boldsymbol{\beta}_{hr}^{(t_{hi})}) &= \begin{bmatrix} \boldsymbol{f}_{hr}^\top(\boldsymbol{\beta}_{hr}^{(t_{hi})}, \boldsymbol{v}_{hi1}) & \boldsymbol{f}_{hr}^\top(\boldsymbol{\beta}_{hr}^{(t_{hi})}, \boldsymbol{v}_{hi2}) & \cdots & \boldsymbol{f}_{hr}^\top(\boldsymbol{\beta}_{hr}^{(t_{hi})}, \boldsymbol{v}_{hiM_{hi}}) \end{bmatrix}^\top \\
\boldsymbol{b}_{hr}^P &= \begin{bmatrix} \boldsymbol{b}_{h1r}^\top & \boldsymbol{b}_{h2r}^\top & \cdots & \boldsymbol{b}_{hM_{hi}r}^\top \end{bmatrix}^\top \\
\boldsymbol{\varepsilon}_{hir} &= \begin{bmatrix} \boldsymbol{\varepsilon}_{hi1r}^\top & \boldsymbol{\varepsilon}_{hi2r}^\top & \cdots & \boldsymbol{\varepsilon}_{hiM_{hi}r}^\top \end{bmatrix}^\top ,
\end{aligned}$$

where the notation "$\top$" denotes vector/matrix transposition and each $\boldsymbol{b}_{hjr}$ is an independent copy of $\boldsymbol{b}_{hr}^{(P)}$ for $j = 1, \ldots, M_{hi}$. Next, matrices

$$\boldsymbol{Z}_{hir}^S = \begin{cases} \begin{bmatrix} \boldsymbol{Z}_{hir,1}^\top & \boldsymbol{Z}_{hir,2}^\top & \cdots & \boldsymbol{Z}_{hir,M_{hi}}^\top \end{bmatrix}^\top , & r \in \mathcal{R}_h^S \\ \mathbf{0}_{n_{hir}, q_{S,hr}} , & r \notin \mathcal{R}_h^S \end{cases}$$

$$\boldsymbol{Z}_{hir}^P = \begin{cases} \mathrm{Diag}\left(\boldsymbol{Z}_{hi1r}, \boldsymbol{Z}_{hi2r}, \ldots, \boldsymbol{Z}_{hiM_{hi}r}\right) , & r \in \mathcal{R}_h^P \\ \mathbf{0}_{n_{hir}, M_{hi} \cdot q_{P,hr}} , & r \notin \mathcal{R}_h^P \end{cases}$$

for $n_{hir} = \sum_{j=1}^{M_{hi}} n_{hijr}$ allow model (9) concatenation over source-to-sensor paths to be written as follows,

$$\boldsymbol{y}_{hir} = \boldsymbol{f}_{hir}(\boldsymbol{\beta}_{hr}^{(t_{hi})}) + \boldsymbol{Z}_{hir}^S \boldsymbol{b}_{hr}^{(S)} + \boldsymbol{Z}_{hir}^P \boldsymbol{b}_{hr}^P + \boldsymbol{\varepsilon}_{hir} . \tag{10}$$

The vector elements $\{\boldsymbol{b}_{hjr}\}_j$ of $\boldsymbol{b}_{hr}^P$, and $\{\boldsymbol{\varepsilon}_{hijr}\}_j$ of $\boldsymbol{\varepsilon}_{hir}$, are mutually independently distributed. If only common model parameters are present in (10), $\boldsymbol{\beta}_{hr}^{(t)} = \boldsymbol{\beta}_{h0r}$, while if no common model parameters are present, $\boldsymbol{\beta}_{hr}^{(t)} = \boldsymbol{\beta}_{htr}$.

Let $\boldsymbol{\beta}_h^{(t)} = (\boldsymbol{\beta}_{h1}^{(t)}, \ldots, \boldsymbol{\beta}_{hR_h}^{(t)})$ denote the vector of fixed effects coefficients for sources $\mathcal{S}_{ht}$ within sensor type h. By concatenating over the sensor measurement types, it becomes possible to rewrite (10) in terms of independent components $\boldsymbol{y}_{hi}$ (assuming conditional independence across sensor types for common sources),

$$\boldsymbol{y}_{hi} = \boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t_{hi})}) + \boldsymbol{Z}_{hi}^S \boldsymbol{b}_h^S + \boldsymbol{Z}_{hi}^P \boldsymbol{b}_h^P + \boldsymbol{\varepsilon}_{hi} ,$$

with vectors

$$\begin{aligned}
\boldsymbol{y}_{hi} &= \begin{bmatrix} \boldsymbol{y}_{hi1}^\top & \boldsymbol{y}_{hi2}^\top & \cdots & \boldsymbol{y}_{hiR_h}^\top \end{bmatrix}^\top \\
\boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t_{hi})}) &= \begin{bmatrix} \boldsymbol{f}_{hi1}^\top(\boldsymbol{\beta}_{h1}^{(t_{hi})}) & \boldsymbol{f}_{hi2}^\top(\boldsymbol{\beta}_{h2}^{(t_{hi})}) & \cdots & \boldsymbol{f}_{hiR_h}^\top(\boldsymbol{\beta}_{hR_h}^{(t_{hi})}) \end{bmatrix}^\top \\
\boldsymbol{b}_h^S &= \begin{bmatrix} \left(\boldsymbol{b}_{hr_{S,1}}^{(S)}\right)^\top & \left(\boldsymbol{b}_{hr_{S,2}}^{(S)}\right)^\top & \cdots & \left(\boldsymbol{b}_{hr_{S,s_{S,h}}}^{(S)}\right)^\top \end{bmatrix}^\top \\
\boldsymbol{b}_h^P &= \begin{bmatrix} \left(\boldsymbol{b}_{hr_{P,1}}^{P}\right)^\top & \left(\boldsymbol{b}_{hr_{P,2}}^{P}\right)^\top & \cdots & \left(\boldsymbol{b}_{hr_{P,s_{P,h}}}^{P}\right)^\top \end{bmatrix}^\top \\
\boldsymbol{\varepsilon}_{hi} &= \begin{bmatrix} \boldsymbol{\varepsilon}_{hi1}^\top & \boldsymbol{\varepsilon}_{hi2}^\top & \cdots & \boldsymbol{\varepsilon}_{hiR_h}^\top \end{bmatrix}^\top .
\end{aligned}$$

Each matrix $\boldsymbol{Z}_{hi}^k$ for $k \in \{S, P\}$ is composed of $R_h \times s_{k,h}$ blocks. The (r, s) block matrix is set to $\boldsymbol{Z}_{hir}^k$ if $r = r_{k,s}$, otherwise to the zero matrix of size $n_{hir} \times q_{S,hr_{S,s}}$ or $n_{hir} \times M_{hi} \cdot q_{P,hr_{P,s}}$, for $r = 1, \ldots, R_h$ and $s = 1, \ldots, s_{k,h}$. Note that

$$\boldsymbol{\varepsilon}_{hi} \sim \mathcal{N}(\boldsymbol{0}_{n_{hi}}, \boldsymbol{\Sigma}_{hi})$$

where $n_{hi} = \sum_{r=1}^{R_h} n_{hir}$ and $\boldsymbol{\Sigma}_{hi}$ is composed of $R_h \times R_h$ blocks. The (r_1, r_2) block matrix $[\boldsymbol{\Sigma}_{hi}]_{r_1,r_2}$ has size $n_{hir_1} \times n_{hir_2}$ and is constructed as follows: Let $\mathcal{J}_{r_1,r_2}^{hi}$ denote the index set containing pairs of indices (k_1, k_2) specifying measurement types r_1 and r_2 that are observed jointly for source i by sensor type h. If $r_1 = r_2 = r$ this index set contains every measurement type r paired with itself. Let $\boldsymbol{e}_{hir,k}$ denote the $n_{hir} \times 1$ unit vector containing a 1 in the k-th position. Then

$$[\boldsymbol{\Sigma}_{hi}]_{r_1,r_2} = \begin{cases} \displaystyle\sum_{(k_1,k_2)\in\mathcal{J}_{r_1,r_2}^{hi}} \sigma_{hr_1r_2}\, \boldsymbol{e}_{hir_1,k_1} \boldsymbol{e}_{hir_2,k_2}^\top , & r_2 \geq r_1 \\ [\boldsymbol{\Sigma}_{hi}]_{r_2,r_1}^\top & r_2 < r_1 \end{cases} .$$

This source-level concatenated model is

$$\boldsymbol{y}_{hi} = \boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t_{hi})}) + \boldsymbol{\epsilon}_{hi} , \tag{11}$$

where

$$\boldsymbol{\epsilon}_{hi} = \boldsymbol{Z}_{hi}^S \boldsymbol{b}_h^S + \boldsymbol{Z}_{hi}^P \boldsymbol{b}_h^P + \boldsymbol{\varepsilon}_{hi} .$$

The covariance matrix of $\boldsymbol{Z}_{hi}^S \boldsymbol{b}_h^S$ is a block-diagonal matrix given by

$$\boldsymbol{\Xi}_{hi}^{(S)} = \mathrm{Diag}\left(\boldsymbol{Z}_{hi1}^S \boldsymbol{\Sigma}_{h1}^{(S)} \left(\boldsymbol{Z}_{hi1}^S\right)^\top, \ldots, \boldsymbol{Z}_{hiR_h}^S \boldsymbol{\Sigma}_{hR_h}^{(S)} \left(\boldsymbol{Z}_{hiR_h}^S\right)^\top \right) ,$$

and the covariance matrix of $\boldsymbol{Z}_{hi}^P \boldsymbol{b}_h^P$ is also a block-diagonal matrix given by

$$\boldsymbol{\Xi}_{hi}^{(P)} = \mathrm{Diag}\left(\boldsymbol{Z}_{hi1}^P \left(\boldsymbol{I}_{M_{hi}} \otimes \boldsymbol{\Sigma}_{h1}^{(P)}\right) \left(\boldsymbol{Z}_{hi1}^P\right)^\top, \ldots, \boldsymbol{Z}_{hiR_h}^P \left(\boldsymbol{I}_{M_{hi}} \otimes \boldsymbol{\Sigma}_{hR_h}^{(P)}\right) \left(\boldsymbol{Z}_{hiR_h}^P\right)^\top \right) .$$

Any measurement type r that is not observed by sensor type h for source i is excluded from the vectors and matrices above. Any measurement type r that is observed by sensor type h

for source i, but on a reduced set of $1 < M_{hir} < M_{hi}$ paths, has M_{hi} replaced by M_{hir} in the above and subsequent development.

The $n_{hi} \times 1$ error vector $\boldsymbol{\epsilon}_{hi}$ thus has mean vector $\mathbf{0}_{n_{hi}}$ and covariance matrix

$$\boldsymbol{\Omega}_{hi} = \boldsymbol{\Xi}_{hi}^{(S)} + \boldsymbol{\Xi}_{hi}^{(P)} + \boldsymbol{\Sigma}_{hi}\,,$$

and is therefore distributed as

$$\boldsymbol{\epsilon}_{hi} \sim \mathcal{N}(\mathbf{0}_{n_{hi}}, \boldsymbol{\Omega}_{hi})\,.$$

Define the p_{ht}-vector $\boldsymbol{\beta}_{ht} = (\boldsymbol{\beta}_{ht1}, \boldsymbol{\beta}_{ht2}, \ldots, \boldsymbol{\beta}_{htR_h})$ for $t = 0, 1, \ldots, T_h$, and the concatenated model parameter vector $\boldsymbol{\beta}_h = (\boldsymbol{\beta}_{h0}, \boldsymbol{\beta}_{h1}, \ldots, \boldsymbol{\beta}_{hT_h})$ of length $p_h = \sum_{t=0}^{T_h} p_{ht}$. Setting

$$\begin{aligned}
\boldsymbol{\beta} &= (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_M) \\
\boldsymbol{\varsigma} &= (\boldsymbol{\varsigma}_1, \boldsymbol{\varsigma}_2, \ldots, \boldsymbol{\varsigma}_M) \\
\boldsymbol{y} &= (\boldsymbol{y}_{11}, \ldots, \boldsymbol{y}_{1M_1}, \boldsymbol{y}_{21}, \ldots, \boldsymbol{y}_{2M_2}, \ldots, \boldsymbol{y}_{M1}, \ldots, \boldsymbol{y}_{MM_M})\,,
\end{aligned}$$

the log-likelihood function up to an additive constant is given by

$$\begin{aligned}
&\mathcal{L}_c(\boldsymbol{\beta}, \boldsymbol{\varsigma} | \boldsymbol{y}) = \\
&\quad -\frac{1}{2} \sum_{h=1}^{M} \sum_{i=1}^{M_h} \log\det\left(\boldsymbol{\Omega}_{hi}\right) \\
&\quad -\frac{1}{2} \sum_{h=1}^{M} \sum_{t=1}^{T_h} \sum_{i \in \mathcal{S}_{ht}} \left(\boldsymbol{y}_{hi} - \boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t)})\right)^\top \boldsymbol{\Omega}_{hi}^{-1} \left(\boldsymbol{y}_{hi} - \boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t)})\right)\,. \qquad (12)
\end{aligned}$$

Let

$$\boldsymbol{r}'_{hit} = \boldsymbol{y}_{hi} - \boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t)})\,.$$

The components of the gradient vector of the log-likelihood function are given by

$$\begin{aligned}
\frac{\partial \mathcal{L}_c}{\partial \boldsymbol{\beta}_{h0}} &= \sum_{t=1}^{T_h} \sum_{i \in \mathcal{S}_{ht}} \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hi}^\top(\boldsymbol{\beta}_h^{(t)}) \boldsymbol{\Omega}_{hi}^{-1} \boldsymbol{r}'_{hit} \\
\frac{\partial \mathcal{L}_c}{\partial \boldsymbol{\beta}_{ht}} &= \sum_{i \in \mathcal{S}_{ht}} \boldsymbol{J}_{\boldsymbol{\beta}_{ht},hi}^\top(\boldsymbol{\beta}_h^{(t)}) \boldsymbol{\Omega}_{hi}^{-1} \boldsymbol{r}'_{hit}\,,\ t = 1, \ldots, T_h \\
\frac{\partial \mathcal{L}_c}{\partial \varsigma_{hk}} &= -\frac{1}{2} \sum_{t=1}^{T_h} \sum_{i \in \mathcal{S}_{ht}} \operatorname{tr}\left\{ \left[\boldsymbol{\Omega}_{hi}^{-1} - \boldsymbol{\Omega}_{hi}^{-1} \boldsymbol{r}'_{hit} \left(\boldsymbol{r}'_{hit}\right)^\top \boldsymbol{\Omega}_{hi}^{-1}\right] \frac{\partial \boldsymbol{\Omega}_{hi}}{\partial \varsigma_{hk}} \right\}
\end{aligned}$$

where $\boldsymbol{J}_{\boldsymbol{\beta}_{hu},hi}(\boldsymbol{\beta}_h^{(t)})$ for $u \in \{0, t\}$ is the Jacobian matrix of $\boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t)})$ with respect to $\boldsymbol{\beta}_{hu}$, $i \in \mathcal{S}_{ht}$,

$$\boldsymbol{J}_{\boldsymbol{\beta}_{hu},hi}^\top(\boldsymbol{\beta}_h^{(t)}) = \left[\nabla f_{hi,1}(\boldsymbol{\beta}_h^{(t)}) \quad \nabla f_{hi,2}(\boldsymbol{\beta}_h^{(t)}) \quad \cdots \quad \nabla f_{hi,n_{hi}}(\boldsymbol{\beta}_h^{(t)})\right] \qquad (13)$$

for $\nabla f_{hi,k}(\boldsymbol{\beta}_h^{(t)})$ the gradient vector of the k-th element of $\boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t)})$ with respect to $\boldsymbol{\beta}_{hu}$, $k = 1, \ldots, n_{hi}$. Computation of gradients and subsequently quantities derived from them are

modified in an obvious manner for measurement types r in which one of the special cases $\boldsymbol{\beta}_{hr}^{(t)} = \boldsymbol{\beta}_{h0r}$ or $\boldsymbol{\beta}_{hr}^{(t)} = \boldsymbol{\beta}_{htr}$ pertains.

For the combination (h, i, r), it is noted that the Jacobian matrix of $\boldsymbol{f}_{hir}(\boldsymbol{\beta}_{hr}^{(t)})$ with respect to $\boldsymbol{\beta}_{hur}$ is given by

$$\boldsymbol{J}_{\boldsymbol{\beta}_{hur},hir}^{\top}(\boldsymbol{\beta}_{hr}^{(t)}) = \left[\boldsymbol{J}_{\boldsymbol{\beta}_{hur},hi1r}^{\top}(\boldsymbol{\beta}_{hr}^{(t)}) \quad \boldsymbol{J}_{\boldsymbol{\beta}_{hur},hi2r}^{\top}(\boldsymbol{\beta}_{hr}^{(t)}) \quad \cdots \quad \boldsymbol{J}_{\boldsymbol{\beta}_{hur},hiM_{hi}r}^{\top}(\boldsymbol{\beta}_{hr}^{(t)})\right]$$

where $\boldsymbol{J}_{\boldsymbol{\beta}_{hur},hijr}(\boldsymbol{\beta}_{hr}^{(t)})$ is the Jacobian matrix of $\boldsymbol{f}_{hr}(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{v}_{hij})$ with respect to $\boldsymbol{\beta}_{hur}$. The vector $\boldsymbol{f}_{hr}(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{v}_{hij})$ contains n_{hijr} copies of the scalar function $f_{hr}(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{v}_{hij})$, so that each column of the matrix $\boldsymbol{J}_{\boldsymbol{\beta}_{hur},hijr}^{\top}(\boldsymbol{\beta}_{hr}^{(t)})$ contains $\nabla f_{hr}(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{v}_{hij})$ with respect to $\boldsymbol{\beta}_{hur}$, i.e.

$$\boldsymbol{J}_{\boldsymbol{\beta}_{hur},hijr}^{\top}(\boldsymbol{\beta}_{hr}^{(t)}) = \nabla f_{hr}(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{v}_{hij})\mathbf{1}_{n_{hijr}}^{\top}$$

for $\mathbf{1}_m$ a m-vector of ones. With this development, (13) can be written as

$$\boldsymbol{J}_{\boldsymbol{\beta}_{hu},hi}(\boldsymbol{\beta}_{h}^{(t)}) = \text{Diag}\left(\boldsymbol{J}_{\boldsymbol{\beta}_{hu1},hi1}(\boldsymbol{\beta}_{h1}^{(t)}), \boldsymbol{J}_{\boldsymbol{\beta}_{hu2},hi2}(\boldsymbol{\beta}_{h2}^{(t)}), \ldots, \boldsymbol{J}_{\boldsymbol{\beta}_{huR_h},hiR_h}(\boldsymbol{\beta}_{hR_h}^{(t)})\right) .$$

Let

$$\boldsymbol{K}_{h0}(\boldsymbol{\beta}_{h}^{(t)}) = \sum_{t=1}^{T_h}\sum_{i\in\mathcal{S}_{ht}}\left[\frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hi}^{\top}(\boldsymbol{\beta}_{h}^{(t)})}{\partial \beta_{h01}}\boldsymbol{\Omega}_{hi}^{-1}\boldsymbol{r}_{hit}' \;\cdots\; \frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hi}^{\top}(\boldsymbol{\beta}_{h}^{(t)})}{\partial \beta_{h0p_{h0}}}\boldsymbol{\Omega}_{hi}^{-1}\boldsymbol{r}_{hit}'\right]$$

$$\boldsymbol{K}_{h0t}(\boldsymbol{\beta}_{h}^{(t)}) = \sum_{i\in\mathcal{S}_{ht}}\left[\frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hi}^{\top}(\boldsymbol{\beta}_{h}^{(t)})}{\partial \beta_{ht1}}\boldsymbol{\Omega}_{hi}^{-1}\boldsymbol{r}_{hit}' \;\cdots\; \frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hi}^{\top}(\boldsymbol{\beta}_{h}^{(t)})}{\partial \beta_{htp_{ht}}}\boldsymbol{\Omega}_{hi}^{-1}\boldsymbol{r}_{hit}'\right], \; t = 1, \ldots, T_h$$

$$\boldsymbol{K}_{ht}(\boldsymbol{\beta}_{h}^{(t)}) = \sum_{i\in\mathcal{S}_{ht}}\left[\frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{ht},hi}^{\top}(\boldsymbol{\beta}_{h}^{(t)})}{\partial \beta_{ht1}}\boldsymbol{\Omega}_{hi}^{-1}\boldsymbol{r}_{hit}' \;\cdots\; \frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{ht},hi}^{\top}(\boldsymbol{\beta}_{h}^{(t)})}{\partial \beta_{htp_{ht}}}\boldsymbol{\Omega}_{hi}^{-1}\boldsymbol{r}_{hit}'\right], \; t = 1, \ldots, T_h$$

$$\boldsymbol{S}_{hit} = 2\,\boldsymbol{\Omega}_{hi}^{-1}\boldsymbol{r}_{hit}'\left(\boldsymbol{r}_{hit}'\right)^{\top}\boldsymbol{\Omega}_{hi}^{-1} - \boldsymbol{\Omega}_{hi}^{-1} .$$

The second partial derivatives of the log-likelihood function are

$$\frac{\partial^2 \mathcal{L}_c}{\partial \boldsymbol{\beta}_{h0}\partial \boldsymbol{\beta}_{h0}^{\top}} = -\sum_{t=1}^{T_h}\sum_{i\in\mathcal{S}_{ht}} \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hi}^{\top}(\boldsymbol{\beta}_{h}^{(t)})\boldsymbol{\Omega}_{hi}^{-1}\boldsymbol{J}_{\boldsymbol{\beta}_{h0},hi}(\boldsymbol{\beta}_{h}^{(t)}) + \boldsymbol{K}_{h0}(\boldsymbol{\beta}_{h}^{(t)})$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial \boldsymbol{\beta}_{h0}\partial \boldsymbol{\beta}_{ht}^{\top}} = -\sum_{i\in\mathcal{S}_{ht}} \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hi}^{\top}(\boldsymbol{\beta}_{h}^{(t)})\boldsymbol{\Omega}_{hi}^{-1}\boldsymbol{J}_{\boldsymbol{\beta}_{ht},hi}(\boldsymbol{\beta}_{h}^{(t)}) + \boldsymbol{K}_{h0t}(\boldsymbol{\beta}_{h}^{(t)}) , \; t = 1, \ldots, T_h$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial \boldsymbol{\beta}_{ht}\partial \boldsymbol{\beta}_{ht}^{\top}} = -\sum_{i\in\mathcal{S}_{ht}} \boldsymbol{J}_{\boldsymbol{\beta}_{ht},hi}^{\top}(\boldsymbol{\beta}_{h}^{(t)})\boldsymbol{\Omega}_{hi}^{-1}\boldsymbol{J}_{\boldsymbol{\beta}_{ht},hi}(\boldsymbol{\beta}_{h}^{(t)}) + \boldsymbol{K}_{ht}(\boldsymbol{\beta}_{h}^{(t)}) , \; t = 1, \ldots, T_h$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial \boldsymbol{\beta}_{h0}\partial \varsigma_{hk}} = -\sum_{t=1}^{T_h}\sum_{i\in\mathcal{S}_{ht}} \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hi}^{\top}(\boldsymbol{\beta}_{h}^{(t)})\boldsymbol{\Omega}_{hi}^{-1}\frac{\partial \boldsymbol{\Omega}_{hi}}{\partial \varsigma_{hk}}\boldsymbol{\Omega}_{hi}^{-1}\boldsymbol{r}_{hit}'$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial \boldsymbol{\beta}_{ht} \partial \varsigma_{hk}} = -\sum_{i \in \mathcal{S}_{ht}} \boldsymbol{J}^{\top}_{\boldsymbol{\beta}_{ht}, hi}(\boldsymbol{\beta}_h^{(t)}) \boldsymbol{\Omega}_{hi}^{-1} \frac{\partial \boldsymbol{\Omega}_{hi}}{\partial \varsigma_{hk}} \boldsymbol{\Omega}_{hi}^{-1} \boldsymbol{r}'_{hit} \, , \; t = 1, \ldots, T_h$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial \varsigma_{hk} \partial \varsigma_{hl}} = -\frac{1}{2} \sum_{t=1}^{T_h} \sum_{i \in \mathcal{S}_{ht}} \operatorname{tr} \left\{ \boldsymbol{\Omega}_{hi}^{-1} \frac{\partial \boldsymbol{\Omega}_{hi}}{\partial \varsigma_{hk}} \boldsymbol{S}_{hit} \frac{\partial \boldsymbol{\Omega}_{hi}}{\partial \varsigma_{hl}} + \left[\boldsymbol{\Omega}_{hi}^{-1} - \boldsymbol{\Omega}_{hi}^{-1} \boldsymbol{r}'_{hit} \left(\boldsymbol{r}'_{hit} \right)^{\top} \boldsymbol{\Omega}_{hi}^{-1} \right] \frac{\partial^2 \boldsymbol{\Omega}_{hi}}{\partial \varsigma_{hk} \partial \varsigma_{hl}} \right\} .$$

Noting that

$$E \left[\boldsymbol{r}'_{hit} | \boldsymbol{\beta}_h^{(t)}, \boldsymbol{\varsigma}_h \right] = \mathbf{0}$$

$$E \left[\boldsymbol{S}_{hit} | \boldsymbol{\beta}_h^{(t)}, \boldsymbol{\varsigma}_h \right] = \boldsymbol{\Omega}_{hi}^{-1} \, ,$$

the elements of the information matrix $\mathcal{I}^c_{\boldsymbol{\beta}_h, \boldsymbol{\beta}_h}$ involving $\boldsymbol{\beta}_h$ for sensor type h are given by

$$\mathcal{I}^c_{\boldsymbol{\beta}_{h0}, \boldsymbol{\beta}_{h0}} = \sum_{t=1}^{T_h} \sum_{i \in \mathcal{S}_{ht}} \boldsymbol{J}^{\top}_{\boldsymbol{\beta}_{h0}, hi}(\boldsymbol{\beta}_h^{(t)}) \boldsymbol{\Omega}_{hi}^{-1} \boldsymbol{J}_{\boldsymbol{\beta}_{h0}, hi}(\boldsymbol{\beta}_h^{(t)})$$

$$\mathcal{I}^c_{\boldsymbol{\beta}_{h0}, \boldsymbol{\beta}_{ht}} = \sum_{i \in \mathcal{S}_{ht}} \boldsymbol{J}^{\top}_{\boldsymbol{\beta}_{h0}, hi}(\boldsymbol{\beta}_h^{(t)}) \boldsymbol{\Omega}_{hi}^{-1} \boldsymbol{J}_{\boldsymbol{\beta}_{ht}, hi}(\boldsymbol{\beta}_h^{(t)}) \, , \; t = 1, \ldots, T_h$$

$$\mathcal{I}^c_{\boldsymbol{\beta}_{ht}, \boldsymbol{\beta}_{ht}} = \sum_{i \in \mathcal{S}_{ht}} \boldsymbol{J}^{\top}_{\boldsymbol{\beta}_{ht}, hi}(\boldsymbol{\beta}_h^{(t)}) \boldsymbol{\Omega}_{hi}^{-1} \boldsymbol{J}_{\boldsymbol{\beta}_{ht}, hi}(\boldsymbol{\beta}_h^{(t)}) \, , \; t = 1, \ldots, T_h$$

$$\mathcal{I}^c_{\boldsymbol{\beta}_{ht_1}, \boldsymbol{\beta}_{ht_2}} = \mathbf{0}_{p_{ht_1}, p_{ht_2}} \, , \; t_1 \neq t_2 \in \{1, \ldots, T_h\}$$

$$\mathcal{I}^c_{\boldsymbol{\beta}_h, \boldsymbol{\varsigma}_h} = \mathbf{0}_{p_h, v_h} \, .$$

The information matrix for sensor type h, in terms of the parameters $(\boldsymbol{\beta}_h, \boldsymbol{\varsigma}_h)$, is therefore given by

$$\mathcal{I}^c_h = \begin{bmatrix} \mathcal{I}^c_{\boldsymbol{\beta}_h, \boldsymbol{\beta}_h} & \mathbf{0}_{p_h, v_h} \\ \mathbf{0}_{v_h, p_h} & \mathcal{I}^c_{\boldsymbol{\varsigma}_h, \boldsymbol{\varsigma}_h} \end{bmatrix} ,$$

where the (t_1, t_2) block of $\mathcal{I}^c_{\boldsymbol{\beta}_h, \boldsymbol{\beta}_h}$ is taken to be $\mathcal{I}^c_{\boldsymbol{\beta}_{ht_1}, \boldsymbol{\beta}_{ht_2}}$ for $t_1, t_2 \in \{0, 1, \ldots, T_h\}$, and the elements of $\mathcal{I}^c_{\boldsymbol{\varsigma}_h, \boldsymbol{\varsigma}_h}$ are obtained as follows:

$$\mathcal{I}^c_{\varsigma_{hk}, \varsigma_{hl}} = \frac{1}{2} \sum_{i=1}^{M_h} \operatorname{tr} \left\{ \boldsymbol{\Omega}_{hi}^{-1} \frac{\partial \boldsymbol{\Omega}_{hi}}{\partial \varsigma_{hk}} \boldsymbol{\Omega}_{hi}^{-1} \frac{\partial \boldsymbol{\Omega}_{hi}}{\partial \varsigma_{hl}} \right\} .$$

Define the following quantities:

$$\mathcal{I}^c_{\boldsymbol{\beta}, \boldsymbol{\beta}} = \operatorname{Diag} \left(\mathcal{I}^c_{\boldsymbol{\beta}_1, \boldsymbol{\beta}_1}, \mathcal{I}^c_{\boldsymbol{\beta}_2, \boldsymbol{\beta}_2}, \ldots, \mathcal{I}^c_{\boldsymbol{\beta}_M, \boldsymbol{\beta}_M} \right)$$

$$\mathcal{I}^c_{\boldsymbol{\varsigma}, \boldsymbol{\varsigma}} = \operatorname{Diag} \left(\mathcal{I}^c_{\boldsymbol{\varsigma}_1, \boldsymbol{\varsigma}_1}, \mathcal{I}^c_{\boldsymbol{\varsigma}_2, \boldsymbol{\varsigma}_2}, \ldots, \mathcal{I}^c_{\boldsymbol{\varsigma}_M, \boldsymbol{\varsigma}_M} \right) .$$

Taking advantage of assumed independence across sensor types, the information matrix for the entire dataset in terms of the parameters $(\boldsymbol{\beta}, \boldsymbol{\varsigma})$ is given by

$$\mathcal{I}^c = \begin{bmatrix} \mathcal{I}^c_{\boldsymbol{\beta}, \boldsymbol{\beta}} & \mathbf{0} \\ \mathbf{0} & \mathcal{I}^c_{\boldsymbol{\varsigma}, \boldsymbol{\varsigma}} \end{bmatrix} .$$

Per standard statistical theory (Efron and Hinkley, 1978; Efron, 1982), the approximate covariance matrix of the maximum likelihood estimators $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\varsigma}})$ is

$$(\mathcal{I}^c)^{-1} = \begin{bmatrix} \left(\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \left(\mathcal{I}^c_{\boldsymbol{\varsigma},\boldsymbol{\varsigma}}\right)^{-1} \end{bmatrix} .$$

The corresponding distribution of the maximum likelihood estimators is

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\varsigma}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\varsigma} \end{pmatrix}, (\mathcal{I}^c)^{-1} \right) .$$

A.2 Statistical Modeling of New Event Data

Now consider a *new* event. For sensor type h, (11) becomes

$$\boldsymbol{y}_{h0} = \boldsymbol{f}_{h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0) + \boldsymbol{\epsilon}_{h0} , \tag{14}$$

where

$$\boldsymbol{\epsilon}_{h0} \sim \mathcal{N}(\mathbf{0}_{n_{h0}}, \boldsymbol{\Omega}_{h0})$$

for

$$\boldsymbol{\Omega}_{h0} = \boldsymbol{\Xi}^{(S)}_{h0} + \boldsymbol{\Xi}^{(P)}_{h0} + \boldsymbol{\Sigma}_{h0} .$$

The q-dimensional parameter $\boldsymbol{\theta}_0$ contains quantities of inferential interest for the new event, such as origin time, location, HOB/DOB, and yield (or some subset of these features). These inputs are known covariates in the analysis of calibration data discussed earlier (i.e., contained within the covariate vectors $\boldsymbol{v}_{hij}$), but some of them may be unknown for a new event.

The error terms $\boldsymbol{\epsilon}_{h0}$ for $h = 1, \ldots, M$ are assumed to be mutually independent, and independent of all error terms in (11). Define

$$\begin{aligned} \boldsymbol{y}_0 &= \begin{bmatrix} \boldsymbol{y}_{10}^\top & \boldsymbol{y}_{20}^\top & \cdots & \boldsymbol{y}_{M0}^\top \end{bmatrix}^\top \\ \boldsymbol{f}_0(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) &= \begin{bmatrix} \boldsymbol{f}_{10}^\top(\boldsymbol{\beta}_1^{(t_{10})}, \boldsymbol{\theta}_0) & \boldsymbol{f}_{20}^\top(\boldsymbol{\beta}_2^{(t_{20})}, \boldsymbol{\theta}_0) & \cdots & \boldsymbol{f}_{M0}^\top(\boldsymbol{\beta}_M^{(t_{M0})}, \boldsymbol{\theta}_0) \end{bmatrix}^\top \\ \boldsymbol{\epsilon}_0 &= \begin{bmatrix} \boldsymbol{\epsilon}_{10}^\top & \boldsymbol{\epsilon}_{20}^\top & \cdots & \boldsymbol{\epsilon}_{M0}^\top \end{bmatrix}^\top \end{aligned}$$

for $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_1^{(t_{10})}, \boldsymbol{\beta}_2^{(t_{20})}, \ldots, \boldsymbol{\beta}_M^{(t_{M0})})$. Concatenated across phenomenologies, (14) becomes

$$\boldsymbol{y}_0 = \boldsymbol{f}_0(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + \boldsymbol{\epsilon}_0 \tag{15}$$

where

$$\boldsymbol{\epsilon}_0 \sim \mathcal{N}(\mathbf{0}_{n_0}, \boldsymbol{\Omega}_0)$$

for $n_0 = \sum_{h=1}^M n_{h0}$ and

$$\boldsymbol{\Omega}_0 = \mathrm{Diag}\left(\boldsymbol{\Omega}_{10}, \boldsymbol{\Omega}_{20}, \ldots, \boldsymbol{\Omega}_{M0}\right) .$$

The log-likelihood function up to an additive constant for the new event data (15) is given by

$$\begin{aligned}\mathcal{L}_0(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \boldsymbol{\varsigma}|\boldsymbol{y}_0) = \\ -\frac{1}{2}\sum_{h=1}^{M}\log\det\left(\boldsymbol{\Omega}_{h0}\right) \\ -\frac{1}{2}\sum_{h=1}^{M}\left(\boldsymbol{y}_{h0}-\boldsymbol{f}_{h0}(\boldsymbol{\beta}_h^{(t_{h0})},\boldsymbol{\theta}_0)\right)^\top\boldsymbol{\Omega}_{h0}^{-1}\left(\boldsymbol{y}_{h0}-\boldsymbol{f}_{h0}(\boldsymbol{\beta}_h^{(t_{h0})},\boldsymbol{\theta}_0)\right). \end{aligned} \tag{16}$$

The components of the gradient vector of the log-likelihood function for the new event data are given by

$$\begin{aligned}\frac{\partial\mathcal{L}_0}{\partial\boldsymbol{\beta}_{hu}} &= \boldsymbol{J}_{\boldsymbol{\beta}_{hu},h0}^\top(\boldsymbol{\beta}_h^{(t_{h0})},\boldsymbol{\theta}_0)\boldsymbol{\Omega}_{h0}^{-1}\boldsymbol{r}'_{h0t_{h0}},\ u\in\{0,t_{h0}\} \\ \frac{\partial\mathcal{L}_0}{\partial\boldsymbol{\theta}_0} &= \sum_{h=1}^{M}\boldsymbol{J}_{\boldsymbol{\theta}_0,h0}^\top(\boldsymbol{\beta}_h^{(t_{h0})},\boldsymbol{\theta}_0)\boldsymbol{\Omega}_{h0}^{-1}\boldsymbol{r}'_{h0t_{h0}} \\ \frac{\partial\mathcal{L}_0}{\partial\varsigma_{hk}} &= -\frac{1}{2}\mathrm{tr}\left\{\left[\boldsymbol{\Omega}_{h0}^{-1}-\boldsymbol{\Omega}_{h0}^{-1}\boldsymbol{r}'_{h0t_{h0}}\left(\boldsymbol{r}'_{h0t_{h0}}\right)^\top\boldsymbol{\Omega}_{h0}^{-1}\right]\frac{\partial\boldsymbol{\Omega}_{h0}}{\partial\varsigma_{hk}}\right\}\end{aligned}$$

where the Jacobian matrix of $\boldsymbol{f}_{h0}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_0)$ with respect to $(\boldsymbol{\beta}_{h0},\boldsymbol{\beta}_{ht},\boldsymbol{\theta}_0)$ is given by

$$\boldsymbol{J}_{h0}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_0) = \left[\boldsymbol{J}_{\boldsymbol{\beta}_{h0},h0}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_0)\quad\boldsymbol{J}_{\boldsymbol{\beta}_{ht},h0}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_0)\quad\boldsymbol{J}_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_0)\right] \tag{17}$$

for

$$\begin{aligned}\boldsymbol{J}_{\boldsymbol{\beta}_{hu},h0}^\top(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_0) &= \left[\nabla_{\boldsymbol{\beta}_{hu}}f_{h0,1}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_0)\quad\nabla_{\boldsymbol{\beta}_{hu}}f_{h0,2}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_0)\quad\cdots\quad\nabla_{\boldsymbol{\beta}_{hu}}f_{h0,n_{h0}}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_0)\right] \\ \boldsymbol{J}_{\boldsymbol{\theta}_0,h0}^\top(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_0) &= \left[\nabla_{\boldsymbol{\theta}_0}f_{h0,1}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_0)\quad\nabla_{\boldsymbol{\theta}_0}f_{h0,2}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_0)\quad\cdots\quad\nabla_{\boldsymbol{\theta}_0}f_{h0,n_{h0}}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_0)\right].\end{aligned}$$

Here $\nabla_{\boldsymbol{\eta}}f_{h0,k}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_0)$ is the gradient vector of the k-th element of $\boldsymbol{f}_{h0}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_0)$ with respect to $\boldsymbol{\eta}$, $k=1,\ldots,n_{h0}$, for $\boldsymbol{\eta}$ one of $\boldsymbol{\beta}_{hu}$ for $u\in\{0,t\}$ or $\boldsymbol{\theta}_0$.

For the combination $(h,0,r)$, the Jacobian matrices of $\boldsymbol{f}_{h0r}(\boldsymbol{\beta}_{hr}^{(t)},\boldsymbol{\theta}_0)$ with respect to $\boldsymbol{\beta}_{hur}$ for $u\in\{0,t\}$ and $\boldsymbol{\theta}_0$ are given by

$$\begin{aligned}\boldsymbol{J}_{\boldsymbol{\beta}_{hur},h0r}^\top(\boldsymbol{\beta}_{hr}^{(t)},\boldsymbol{\theta}_0) &= \left[\boldsymbol{J}_{\boldsymbol{\beta}_{hur},h01r}^\top(\boldsymbol{\beta}_{hr}^{(t)},\boldsymbol{\theta}_0)\quad\boldsymbol{J}_{\boldsymbol{\beta}_{hur},h02r}^\top(\boldsymbol{\beta}_{hr}^{(t)},\boldsymbol{\theta}_0)\quad\cdots\quad\boldsymbol{J}_{\boldsymbol{\beta}_{hur},h0M_{h0}r}^\top(\boldsymbol{\beta}_{hr}^{(t)},\boldsymbol{\theta}_0)\right] \\ \boldsymbol{J}_{\boldsymbol{\theta}_0,h0r}^\top(\boldsymbol{\beta}_{hr}^{(t)},\boldsymbol{\theta}_0) &= \left[\boldsymbol{J}_{\boldsymbol{\theta}_0,h01r}^\top(\boldsymbol{\beta}_{hr}^{(t)},\boldsymbol{\theta}_0)\quad\boldsymbol{J}_{\boldsymbol{\theta}_0,h02r}^\top(\boldsymbol{\beta}_{hr}^{(t)},\boldsymbol{\theta}_0)\quad\cdots\quad\boldsymbol{J}_{\boldsymbol{\theta}_0,h0M_{h0}r}^\top(\boldsymbol{\beta}_{hr}^{(t)},\boldsymbol{\theta}_0)\right],\end{aligned}$$

where $\boldsymbol{J}_{\boldsymbol{\eta},h0jr}(\boldsymbol{\beta}_{hr}^{(t)},\boldsymbol{\theta}_0)$ is the Jacobian matrix of $\boldsymbol{f}_{hr}((\boldsymbol{\beta}_{hr}^{(t)},\boldsymbol{\theta}_0),\boldsymbol{v}'_{h0j})$ with respect to $\boldsymbol{\eta}$ for $\boldsymbol{\eta}$ one of $\boldsymbol{\beta}_{hur}$ with $u\in\{0,t\}$ or $\boldsymbol{\theta}_0$. Note that the covariate vector $\boldsymbol{v}'_{h0j}$ is the subset of $\boldsymbol{v}_{h0j}$ formed by removing the unknown inputs $\boldsymbol{\theta}_0$ for the new event (i.e. quantities such as origin

time, location, HOB/DOB, and yield that are known covariates in the calibration data but may be unknown for a new event) — in other words, for the new event $\boldsymbol{v}_{h0j} = (\boldsymbol{\theta}_0, \boldsymbol{v}'_{h0j})$. The vector $\boldsymbol{f}_{hr}((\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0), \boldsymbol{v}'_{h0j})$ contains n_{h0jr} copies of the scalar function $f_{hr}((\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0), \boldsymbol{v}'_{h0j})$, so that each column of the matrix $\boldsymbol{J}^{\top}_{\boldsymbol{\eta},h0jr}(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0)$ contains $\nabla_{\boldsymbol{\eta}} f_{hr}((\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0), \boldsymbol{v}'_{h0j})$ with respect to $\boldsymbol{\eta}$, for $\boldsymbol{\eta}$ one of $\boldsymbol{\beta}_{hur}$ with $u \in \{0, t\}$ or $\boldsymbol{\theta}_0$, i.e.

$$\boldsymbol{J}^{\top}_{\boldsymbol{\eta},h0jr}(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0) = \nabla_{\boldsymbol{\eta}} f_{hr}((\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0), \boldsymbol{v}'_{h0j}) \mathbf{1}^{\top}_{n_{h0jr}}$$

for $\mathbf{1}_m$ a m-vector of ones.

With this development, the constituent elements of (17) can be written as

$$\boldsymbol{J}_{\boldsymbol{\beta}_{hu},h0}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_0) = \text{Diag}\left(\boldsymbol{J}_{\boldsymbol{\beta}_{hu}1,h01}(\boldsymbol{\beta}_{h1}^{(t)}, \boldsymbol{\theta}_0), \boldsymbol{J}_{\boldsymbol{\beta}_{hu}2,h02}(\boldsymbol{\beta}_{h2}^{(t)}, \boldsymbol{\theta}_0), \ldots, \boldsymbol{J}_{\boldsymbol{\beta}_{hu}R_h,h0R_h}(\boldsymbol{\beta}_{hR_h}^{(t)}, \boldsymbol{\theta}_0)\right)$$

$$\boldsymbol{J}^{\top}_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_0) = \left[\boldsymbol{J}^{\top}_{\boldsymbol{\theta}_0,h01}(\boldsymbol{\beta}_{h1}^{(t)}, \boldsymbol{\theta}_0) \quad \boldsymbol{J}^{\top}_{\boldsymbol{\theta}_0,h02}(\boldsymbol{\beta}_{h2}^{(t)}, \boldsymbol{\theta}_0) \quad \cdots \quad \boldsymbol{J}^{\top}_{\boldsymbol{\theta}_0,h0R_h}(\boldsymbol{\beta}_{hR_h}^{(t)}, \boldsymbol{\theta}_0)\right].$$

Let

$$\boldsymbol{K}_{(\boldsymbol{\theta}_0,\boldsymbol{\theta}_0)}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = \sum_{h=1}^{M} \left[\frac{\partial \boldsymbol{J}^{\top}_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0)}{\partial \theta_{01}} \boldsymbol{\Omega}_{h0}^{-1} \boldsymbol{r}'_{h0t_{h0}} \quad \cdots \quad \frac{\partial \boldsymbol{J}^{\top}_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0)}{\partial \theta_{0q}} \boldsymbol{\Omega}_{h0}^{-1} \boldsymbol{r}'_{h0t_{h0}}\right]$$

$$\boldsymbol{K}_{(\boldsymbol{\beta}_{hu},\boldsymbol{\theta}_0)}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0) = \left[\frac{\partial \boldsymbol{J}^{\top}_{\boldsymbol{\beta}_{hu},h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0)}{\partial \theta_{01}} \boldsymbol{\Omega}_{h0}^{-1} \boldsymbol{r}'_{h0t_{h0}} \quad \cdots \quad \frac{\partial \boldsymbol{J}^{\top}_{\boldsymbol{\beta}_{hu},h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0)}{\partial \theta_{0q}} \boldsymbol{\Omega}_{h0}^{-1} \boldsymbol{r}'_{h0t_{h0}}\right]$$

for $u \in \{0, t_{h0}\}$. The second partial derivatives of the log-likelihood function for the new event data follow from the previous development for the calibration data (replacing index i with '0'), with new elements relating to $\boldsymbol{\theta}_0$:

$$\frac{\partial^2 \mathcal{L}_0}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0^{\top}} = -\sum_{h=1}^{M} \boldsymbol{J}^{\top}_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0) \boldsymbol{\Omega}_{h0}^{-1} \boldsymbol{J}_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0) + \boldsymbol{K}_{(\boldsymbol{\theta}_0,\boldsymbol{\theta}_0)}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$$

$$\frac{\partial^2 \mathcal{L}_0}{\partial \boldsymbol{\beta}_{hu} \partial \boldsymbol{\theta}_0^{\top}} = -\boldsymbol{J}^{\top}_{\boldsymbol{\beta}_{hu},h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0) \boldsymbol{\Omega}_{h0}^{-1} \boldsymbol{J}_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0) + \boldsymbol{K}_{(\boldsymbol{\beta}_{hu},\boldsymbol{\theta}_0)}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0)\,,\ u \in \{0, t_{h0}\}$$

$$\frac{\partial^2 \mathcal{L}_0}{\partial \boldsymbol{\theta}_0 \partial \varsigma_{hk}} = -\boldsymbol{J}^{\top}_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0) \boldsymbol{\Omega}_{h0}^{-1} \frac{\partial \boldsymbol{\Omega}_{h0}}{\partial \varsigma_{hk}} \boldsymbol{\Omega}_{h0}^{-1} \boldsymbol{r}'_{h0t_{h0}}\,.$$

The log-likelihood function for the joint calibration ('c') and new event ('0') data is additive, as conditional on the parameters $(\boldsymbol{\beta}, \boldsymbol{\theta}_0, \boldsymbol{\varsigma})$ the two sources of data are independently distributed,

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}_0, \boldsymbol{\varsigma}) = \mathcal{L}_c(\boldsymbol{\beta}, \boldsymbol{\varsigma}) + \mathcal{L}_0(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \boldsymbol{\varsigma})\,. \tag{18}$$

This implies that the information matrix from the combined data in terms of the parameters $(\boldsymbol{\beta}, \boldsymbol{\theta}_0, \boldsymbol{\varsigma})$ is given by

$$\mathcal{I}^{c,0} = \begin{bmatrix} \mathcal{I}^{c}_{\boldsymbol{\beta},\boldsymbol{\beta}} + \mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\beta}} & \mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\theta}_0} & \mathbf{0}_{p,v} \\ \left(\mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^{\top} & \mathcal{I}^{0}_{\boldsymbol{\theta}_0,\boldsymbol{\theta}_0} & \mathbf{0}_{q,v} \\ \mathbf{0}_{v,p} & \mathbf{0}_{v,q} & \mathcal{I}^{c}_{\boldsymbol{\varsigma},\boldsymbol{\varsigma}} + \mathcal{I}^{0}_{\boldsymbol{\varsigma},\boldsymbol{\varsigma}} \end{bmatrix}$$

where

$$\begin{aligned}
\mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\beta}} &= \text{Diag}\left(\mathcal{I}^0_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_1}, \mathcal{I}^0_{\boldsymbol{\beta}_2,\boldsymbol{\beta}_2}, \ldots, \mathcal{I}^0_{\boldsymbol{\beta}_M,\boldsymbol{\beta}_M}\right) \\
\left[\mathcal{I}^0_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right]_{t_1 t_2} &= \begin{cases} \mathcal{I}^0_{\boldsymbol{\beta}_{ht_1},\boldsymbol{\beta}_{ht_2}}\,, & t_1, t_2 \in \{0, t_{h0}\} \\ \mathbf{0}_{p_{ht_1},p_{ht_2}}\,, & (t_1,t_2) \notin \{0,t_{h0}\} \times \{0,t_{h0}\} \end{cases} \\
\mathcal{I}^0_{\boldsymbol{\beta}_{ht_1},\boldsymbol{\beta}_{ht_2}} &= \boldsymbol{J}^\top_{\boldsymbol{\beta}_{ht_1},h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0)\boldsymbol{\Omega}^{-1}_{h0}\boldsymbol{J}_{\boldsymbol{\beta}_{ht_2},h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0)\,,\ t_1, t_2 \in \{0, t_{h0}\} \\
\left(\mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^\top &= \left[\left(\mathcal{I}^0_{\boldsymbol{\beta}_1,\boldsymbol{\theta}_0}\right)^\top \quad \left(\mathcal{I}^0_{\boldsymbol{\beta}_2,\boldsymbol{\theta}_0}\right)^\top \quad \cdots \quad \left(\mathcal{I}^0_{\boldsymbol{\beta}_M,\boldsymbol{\theta}_0}\right)^\top\right] \\
\left[\mathcal{I}^0_{\boldsymbol{\beta}_h,\boldsymbol{\theta}_0}\right]_{t1} &= \begin{cases} \boldsymbol{J}^\top_{\boldsymbol{\beta}_{ht},h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0)\boldsymbol{\Omega}^{-1}_{h0}\boldsymbol{J}_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0)\,, & t \in \{0, t_{h0}\} \\ \mathbf{0}_{p_{ht},q}\,, & t \notin \{0, t_{h0}\} \end{cases} \\
\mathcal{I}^0_{\boldsymbol{\theta}_0,\boldsymbol{\theta}_0} &= \sum_{h=1}^{M} \boldsymbol{J}^\top_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0)\boldsymbol{\Omega}^{-1}_{h0}\boldsymbol{J}_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0) \\
\mathcal{I}^0_{\boldsymbol{\varsigma},\boldsymbol{\varsigma}} &= \text{Diag}\left(\mathcal{I}^0_{\boldsymbol{\varsigma}_1,\boldsymbol{\varsigma}_1}, \mathcal{I}^0_{\boldsymbol{\varsigma}_2,\boldsymbol{\varsigma}_2}, \ldots, \mathcal{I}^0_{\boldsymbol{\varsigma}_M,\boldsymbol{\varsigma}_M}\right) \\
\mathcal{I}^0_{\varsigma_{hk},\varsigma_{hl}} &= \frac{1}{2}\text{tr}\left\{\boldsymbol{\Omega}^{-1}_{h0}\frac{\partial \boldsymbol{\Omega}_{h0}}{\partial \varsigma_{hk}}\boldsymbol{\Omega}^{-1}_{h0}\frac{\partial \boldsymbol{\Omega}_{h0}}{\partial \varsigma_{hl}}\right\}\,.
\end{aligned}$$

In the above, $[\boldsymbol{A}]_{ab}$ refers to the (a,b) block of the matrix $\boldsymbol{A}$ where blocks are formed from the vector elements of $\boldsymbol{\beta}_h$, indexed by $0, 1, \ldots, T_h$. Consider the following definitions,

$$\begin{aligned}
\boldsymbol{J}_{(0,t_{h0}),h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0) &= \left[\boldsymbol{J}_{\boldsymbol{\beta}_{h0},h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0) \quad \boldsymbol{J}_{\boldsymbol{\beta}_{ht_{h0}},h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0)\right] \\
\mathcal{C}^c_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h} &= \left(\mathcal{I}^c_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right)^{-1} \\
\mathcal{C}^c_{(0,t_{h0}),h0} &= \begin{bmatrix} \left[\mathcal{C}^c_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right]_{00} & \left[\mathcal{C}^c_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right]_{0t_{h0}} \\ \left[\mathcal{C}^c_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right]^\top_{0t_{h0}} & \left[\mathcal{C}^c_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right]_{t_{h0}t_{h0}} \end{bmatrix} \\
\mathcal{P}^c_{(0,t_{h0}),h0} &= \boldsymbol{J}_{(0,t_{h0}),h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0)\,\mathcal{C}^c_{(0,t_{h0}),h0}\,\boldsymbol{J}^\top_{(0,t_{h0}),h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0)\,,
\end{aligned}$$

where

$$\begin{aligned}
\left[\mathcal{C}^c_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right]_{00} &= \left(\mathcal{I}^c_{\boldsymbol{\beta}_{h0},\boldsymbol{\beta}_{h0}} - \sum_{t=1}^{T_h}\mathcal{I}^c_{\boldsymbol{\beta}_{h0},\boldsymbol{\beta}_{ht}}\left(\mathcal{I}^c_{\boldsymbol{\beta}_{ht},\boldsymbol{\beta}_{ht}}\right)^{-1}\left(\mathcal{I}^c_{\boldsymbol{\beta}_{h0},\boldsymbol{\beta}_{ht}}\right)^\top\right)^{-1} \\
\left[\mathcal{C}^c_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right]_{0t_{h0}} &= -\left[\mathcal{C}^c_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right]_{00}\mathcal{I}^c_{\boldsymbol{\beta}_{h0},\boldsymbol{\beta}_{ht_{h0}}}\left(\mathcal{I}^c_{\boldsymbol{\beta}_{ht_{h0}},\boldsymbol{\beta}_{ht_{h0}}}\right)^{-1} \\
\left[\mathcal{C}^c_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right]_{t_{h0}t_{h0}} &= \left(\mathcal{I}^c_{\boldsymbol{\beta}_{ht_{h0}},\boldsymbol{\beta}_{ht_{h0}}}\right)^{-1} \\
&\quad + \left(\mathcal{I}^c_{\boldsymbol{\beta}_{ht_{h0}},\boldsymbol{\beta}_{ht_{h0}}}\right)^{-1}\left(\mathcal{I}^c_{\boldsymbol{\beta}_{h0},\boldsymbol{\beta}_{ht_{h0}}}\right)^\top\left[\mathcal{C}^c_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right]_{00}\mathcal{I}^c_{\boldsymbol{\beta}_{h0},\boldsymbol{\beta}_{ht_{h0}}}\left(\mathcal{I}^c_{\boldsymbol{\beta}_{ht_{h0}},\boldsymbol{\beta}_{ht_{h0}}}\right)^{-1}\,.
\end{aligned}$$

Let

$$\begin{aligned}
\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}} &= \mathcal{I}^{c}_{\boldsymbol{\beta},\boldsymbol{\beta}} + \mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\beta}} \\
\mathcal{I}^{c,0}_{\boldsymbol{\theta}_0|\boldsymbol{\beta}} &= \mathcal{I}^{0}_{\boldsymbol{\theta}_0,\boldsymbol{\theta}_0} - \left(\mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^\top \left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1} \mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\theta}_0} \\
&= \sum_{h=1}^{M} \boldsymbol{J}^\top_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0) \left[\boldsymbol{\Omega}_{h0} + \mathcal{P}^{c}_{(0,t_{h0}),h0}\right]^{-1} \boldsymbol{J}_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0) \\
\mathcal{I}^{c,0}_{\boldsymbol{\varsigma},\boldsymbol{\varsigma}} &= \mathcal{I}^{c}_{\boldsymbol{\varsigma},\boldsymbol{\varsigma}} + \mathcal{I}^{0}_{\boldsymbol{\varsigma},\boldsymbol{\varsigma}} \,.
\end{aligned}$$

The approximate covariance matrix $\left(\mathcal{I}^{c,0}\right)^{-1}$ of the maximum likelihood estimators $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\varsigma}})$ is

$$\begin{bmatrix}
\left(\mathcal{I}^{c,0}_{\boldsymbol{\beta}|\boldsymbol{\theta}_0}\right)^{-1} & -\left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1} \mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\theta}_0} \left(\mathcal{I}^{c,0}_{\boldsymbol{\theta}_0|\boldsymbol{\beta}}\right)^{-1} & \mathbf{0} \\
-\left(\mathcal{I}^{c,0}_{\boldsymbol{\theta}_0|\boldsymbol{\beta}}\right)^{-1} \left(\mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^\top \left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1} & \left(\mathcal{I}^{c,0}_{\boldsymbol{\theta}_0|\boldsymbol{\beta}}\right)^{-1} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \left(\mathcal{I}^{c,0}_{\boldsymbol{\varsigma},\boldsymbol{\varsigma}}\right)^{-1}
\end{bmatrix}$$

where

$$\left(\mathcal{I}^{c,0}_{\boldsymbol{\beta}|\boldsymbol{\theta}_0}\right)^{-1} = \left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1} + \left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1} \mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\theta}_0} \left(\mathcal{I}^{c,0}_{\boldsymbol{\theta}_0|\boldsymbol{\beta}}\right)^{-1} \left(\mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^\top \left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1} ,$$

resulting in a corresponding Gaussian distribution for the maximum likelihood estimators,

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\theta}}_0 \\ \widehat{\boldsymbol{\varsigma}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\theta}_0 \\ \boldsymbol{\varsigma} \end{pmatrix}, \left(\mathcal{I}^{c,0}\right)^{-1} \right) . \tag{19}$$

A.3 Example: Linear Model

Consider the linear model formulation in Williams et al. (2021), where $\boldsymbol{\theta}_0$ reduces to $\log_{10}$ yield W_0:

$$\begin{aligned}
\boldsymbol{f}_{hi}(\boldsymbol{\beta}_h) &= \begin{bmatrix} \boldsymbol{X}_{hi} & \mathbf{1}_{n_{hi}} W_i \end{bmatrix} \boldsymbol{\beta}_h \\
\boldsymbol{f}_{h0}(\boldsymbol{\beta}_h, W_0) &= \begin{bmatrix} \boldsymbol{X}_{h0} & \mathbf{1}_{n_{h0}} W_0 \end{bmatrix} \boldsymbol{\beta}_h \,.
\end{aligned}$$

Here $T_h = 0$. It is straightforward to calculate

$$\begin{aligned}
\boldsymbol{J}_{hi}(\boldsymbol{\beta}_h) &= \begin{bmatrix} \boldsymbol{X}_{hi} & \mathbf{1}_{n_{hi}} W_i \end{bmatrix} \\
\boldsymbol{J}_{\boldsymbol{\beta}_h,h0}(\boldsymbol{\beta}_h, W_0) &= \begin{bmatrix} \boldsymbol{X}_{h0} & \mathbf{1}_{n_{h0}} W_0 \end{bmatrix} \\
\boldsymbol{J}_{W_0,h0}(\boldsymbol{\beta}_h, W_0) &= \mathbf{1}_{n_{h0}} \omega_h
\end{aligned}$$

resulting in

$$\mathcal{I}^{c,0}_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h} = \sum_{i=0}^{M_h} \begin{bmatrix} \boldsymbol{X}_{hi}^\top \boldsymbol{\Omega}_{hi}^{-1} \boldsymbol{X}_{hi} & W_i \boldsymbol{X}_{hi}^\top \boldsymbol{\Omega}_{hi}^{-1} \mathbf{1}_{n_{hi}} \\ W_i \mathbf{1}_{n_{hi}}^\top \boldsymbol{\Omega}_{hi}^{-1} \boldsymbol{X}_{hi} & W_i^2 \mathbf{1}_{n_{hi}}^\top \boldsymbol{\Omega}_{hi}^{-1} \mathbf{1}_{n_{hi}} \end{bmatrix}$$

$$= \begin{bmatrix} \mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\boldsymbol{\mu}_h} & \mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\omega_h} \\ \left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\omega_h}\right)^\top & \mathcal{I}^{c,0}_{\omega_h,\omega_h} \end{bmatrix},$$

where $\boldsymbol{\mu}_h = \boldsymbol{\beta}_h(-p_h)$ (all elements of $\boldsymbol{\beta}_h$ except for the last one) and $\omega_h = \boldsymbol{\beta}_h(p_h)$ (the last element of $\boldsymbol{\beta}_h$). This leads to the following expression for $\left(\mathcal{I}^{c,0}_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right)^{-1}$:

$$\begin{bmatrix} \left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h|\omega_h}\right)^{-1} & -\left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\boldsymbol{\mu}_h}\right)^{-1}\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\omega_h}\left(\mathcal{I}^{c,0}_{\omega_h|\boldsymbol{\mu}_h}\right)^{-1} \\ -\left(\mathcal{I}^{c,0}_{\omega_h|\boldsymbol{\mu}_h}\right)^{-1}\left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\omega_h}\right)^\top\left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\boldsymbol{\mu}_h}\right)^{-1} & \left(\mathcal{I}^{c,0}_{\omega_h|\boldsymbol{\mu}_h}\right)^{-1} \end{bmatrix}$$

for

$$\begin{aligned} \mathcal{I}^{c,0}_{\omega_h|\boldsymbol{\mu}_h} &= \mathcal{I}^{c,0}_{\omega_h,\omega_h} - \left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\omega_h}\right)^\top\left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\boldsymbol{\mu}_h}\right)^{-1}\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\omega_h} \\ \left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h|\omega_h}\right)^{-1} &= \left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\boldsymbol{\mu}_h}\right)^{-1} + \left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\boldsymbol{\mu}_h}\right)^{-1}\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\omega_h}\left(\mathcal{I}^{c,0}_{\omega_h|\boldsymbol{\mu}_h}\right)^{-1}\left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\omega_h}\right)^\top\left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\boldsymbol{\mu}_h}\right)^{-1}. \end{aligned}$$

Utilizing the following definitions,

$$\begin{aligned} \boldsymbol{X}_0 &= \text{Diag}\left(\boldsymbol{X}_{10}, \boldsymbol{X}_{20}, \ldots, \boldsymbol{X}_{M0}\right) \\ \boldsymbol{U}_0 &= \text{Diag}\left(\mathbf{1}_{n_{10}}, \mathbf{1}_{n_{20}}, \ldots, \mathbf{1}_{n_{M0}}\right) \\ \overline{\boldsymbol{X}}_0 &= \boldsymbol{U}_0^\top\boldsymbol{\Omega}_0^{-1}\boldsymbol{X}_0 \\ \mathcal{I}^{c,0}_{\boldsymbol{\mu},\boldsymbol{\mu}} &= \text{Diag}\left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_1}, \mathcal{I}^{c,0}_{\boldsymbol{\mu}_2,\boldsymbol{\mu}_2}, \ldots, \mathcal{I}^{c,0}_{\boldsymbol{\mu}_M,\boldsymbol{\mu}_M}\right) \\ \mathcal{I}^{c,0}_{\boldsymbol{\mu},\boldsymbol{\omega}} &= \text{Diag}\left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_1,\omega_1}, \mathcal{I}^{c,0}_{\boldsymbol{\mu}_2,\omega_2}, \ldots, \mathcal{I}^{c,0}_{\boldsymbol{\mu}_M,\omega_M}\right) \\ \widetilde{\boldsymbol{X}}_0 &= \overline{\boldsymbol{X}}_0\left(\mathcal{I}^{c,0}_{\boldsymbol{\mu},\boldsymbol{\mu}}\right)^{-1}\mathcal{I}^{c,0}_{\boldsymbol{\mu},\boldsymbol{\omega}} \\ \mathcal{I}^{c,0}_{\boldsymbol{\omega}|\boldsymbol{\mu}} &= \text{Diag}\left(\mathcal{I}^{c,0}_{\omega_1|\boldsymbol{\mu}_1}, \mathcal{I}^{c,0}_{\omega_2|\boldsymbol{\mu}_2}, \ldots, \mathcal{I}^{c,0}_{\omega_M|\boldsymbol{\mu}_M}\right) \\ \boldsymbol{\omega} &= \left(\omega_1, \omega_2, \ldots, \omega_M\right), \end{aligned}$$

the maximum likelihood estimator (MLE) $\widehat{W}_0$ of W_0 has asymptotic variance $V\left[\widehat{W}_0\right]$ given by

$$\begin{aligned} V^{-1}\left[\widehat{W}_0\right] = \boldsymbol{\omega}^\top &\left[\boldsymbol{U}_0^\top\boldsymbol{\Omega}_0^{-1}\boldsymbol{U}_0 - \overline{\boldsymbol{X}}_0\left(\mathcal{I}^{c,0}_{\boldsymbol{\mu},\boldsymbol{\mu}}\right)^{-1}\overline{\boldsymbol{X}}_0^\top\right]\boldsymbol{\omega} \\ &- \boldsymbol{\omega}^\top\left(\widetilde{\boldsymbol{X}}_0 - W_0\boldsymbol{U}_0^\top\boldsymbol{\Omega}_0^{-1}\boldsymbol{U}_0\right)\left(\mathcal{I}^{c,0}_{\boldsymbol{\omega}|\boldsymbol{\mu}}\right)^{-1}\left(\widetilde{\boldsymbol{X}}_0^\top - W_0\boldsymbol{U}_0^\top\boldsymbol{\Omega}_0^{-1}\boldsymbol{U}_0\right)\boldsymbol{\omega}, \end{aligned}$$

where MLEs replace parameters in this expression to obtain the estimate $\hat{V}\left[\widehat{W}_0\right]$.

A.4 Errors-in-Variables Treatment of Calibration Yields

The above framework assumes yields for sources in the calibration dataset(s) are known with certainty (and thus included in the covariate vectors $\boldsymbol{v}_{hij}$). This may not be the case in general

as design (e.g., calculated via chemical-to-nuclear equivalence factors for HE tests) or measured (e.g. radiochemistry) yields may differ from the actual device yields. To accommodate this situation, the previous development may be extended to take an "errors-in-variables" approach for the yield covariate.

Let $\mathcal{S}$ denote the index set of unique sources (across sensor types) contained in the calibration dataset(s). For $s \in \mathcal{S}$ it is assumed as in Equation (3) that

$$\widetilde{W}_s \sim W_s + \epsilon_s \,, \qquad \epsilon_s \sim \mathcal{N}\left(0, \sigma_s^2\right) \,,$$

where $\widetilde{W}_s$ and W_s are the design/measured log-yield and actual device log-yield for source s, respectively, and σ_s^2 is the variance parameter that specifies an assumed degree of deviation between these quantities. The variance parameters $\{\sigma_s^2\}$ are *assumed known*. If *replicate data* exist for any source $s \in \mathcal{S}$, then it is possible to statistically estimate σ_s^2. Additionally, independence of design/measured log-yields across devices is assumed.

For each source s, the unknown parameter value W_s replaces the calculated/measured quantity $\widetilde{W}_s$ in the log-likelihood function $\mathcal{L}_c(\boldsymbol{\beta}, \boldsymbol{\varsigma})$ for the calibration data. That is, log-yield is removed from the covariate vectors $\boldsymbol{v}_{hij}$ and the calibration log-likelihood function is written $\mathcal{L}_c(\boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{\varsigma})$ to represent its additional dependence on the unknown log-yields $\boldsymbol{w}$ for each calibration source.

The joint distribution of the new event, calibration, and calculated/measured yield data vectors $(\boldsymbol{y}_0, \boldsymbol{y}, \widetilde{\boldsymbol{w}})$ given all model parameters $(\boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{\theta}_0, \boldsymbol{\varsigma})$ is given by

$$[\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{\varsigma}]\,[\boldsymbol{y}_0|\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \boldsymbol{\varsigma}]\,[\widetilde{\boldsymbol{w}}|\boldsymbol{w}] \,,$$

so that the log-likelihood function from the previous development is augmented to accommodate a term informing on the actual device yields,

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{\theta}_0, \boldsymbol{\varsigma}) = \mathcal{L}_c(\boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{\varsigma}) + \mathcal{L}_0(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \boldsymbol{\varsigma}) + \mathcal{L}_w(\boldsymbol{w}) \,. \tag{20}$$

The information matrices required for estimating the asymptotic covariance matrix of the MLE for $\boldsymbol{\theta}_0$ carry over from the previous development, with extra block components to accommodate the additional parameters $\boldsymbol{w}$. New notation is required to compute derivatives of the augmented log-likelihood function. Let $\mathcal{H}_s$ denote the index set of sensor types measured for source $s \in \mathcal{S}$, and $i_h(s)$ the index $i \in \{1, 2, \ldots, M_h\}$ associated with source s for sensor type h. Let

$$\boldsymbol{g}_{hi}(\boldsymbol{\beta}_h^{(t_{hi})}, w) = \frac{\partial \boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t_{hi})}, w)}{\partial w}$$

where the notation $\boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t_{hi})}, w)$ acknowledges that log-yield w is now viewed as a parameter rather than as a fixed covariate in $\boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t_{hi})})$. The vector $\boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t_{hi})}, w)$ ultimately contains n_{hij} copies of scalar elements $f_{hir}(\boldsymbol{\beta}_{hr}^{(t_{hi})}, w, \boldsymbol{v}''_{hij})$ (where log-yield w has been removed from $\boldsymbol{v}_{hij}$, $\boldsymbol{v}_{hij} = (w, \boldsymbol{v}''_{hij})$), having partial derivatives

$$g_{hir}(\boldsymbol{\beta}_{hr}^{(t_{hi})}, w, \boldsymbol{v}''_{hij}) = \frac{\partial f_{hir}(\boldsymbol{\beta}_{hr}^{(t_{hi})}, w, \boldsymbol{v}''_{hij})}{\partial w} \,.$$

In similar fashion, the Jacobian of $\boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t_{hi})}, w)$ with respect to $\boldsymbol{\beta}_{hu}$ for $u \in \{0, t_{hi}\}$ is given by

$$\boldsymbol{J}_{\boldsymbol{\beta}_{hu},hi}(\boldsymbol{\beta}_h^{(t_{hi})}, w) = \left[\frac{\partial \boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t_{hi})}, w)}{\partial \beta_{hu1}} \quad \frac{\partial \boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t_{hi})}, w)}{\partial \beta_{hu2}} \quad \cdots \quad \frac{\partial \boldsymbol{f}_{hi}(\boldsymbol{\beta}_h^{(t_{hi})}, w)}{\partial \beta_{hup_{hu}}} \right], \; u \in \{0, t_{hi}\}.$$

The partial derivatives of the log-likelihood function for the calibration data with respect to log-yield w_s for $s \in \mathcal{S}$ are given by

$$\frac{\partial \mathcal{L}_c}{\partial w_s} = \sum_{h \in \mathcal{H}_s} \left(\boldsymbol{g}_{hi_h(s)}(\boldsymbol{\beta}_h^{(t_{hi_h(s)})}, w_s) \right)^\top \boldsymbol{\Omega}_{hi_h(s)}^{-1} \boldsymbol{r}'_{hi_h(s)t_{hi_h(s)}},$$

with relevant second partial derivatives given by

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}_c}{\partial w_s^2} =& - \sum_{h \in \mathcal{H}_s} \left(\boldsymbol{g}_{hi_h(s)}(\boldsymbol{\beta}_h^{(t_{hi_h(s)})}, w_s) \right)^\top \boldsymbol{\Omega}_{hi_h(s)}^{-1} \boldsymbol{g}_{hi_h(s)}(\boldsymbol{\beta}_h^{(t_{hi_h(s)})}, w_s) \\
& + \sum_{h \in \mathcal{H}_s} \left(\frac{\partial \boldsymbol{g}_{hi_h(s)}(\boldsymbol{\beta}_h^{(t_{hi_h(s)})}, w_s)}{\partial w_s} \right)^\top \boldsymbol{\Omega}_{hi_h(s)}^{-1} \boldsymbol{r}'_{hi_h(s)t_{hi_h(s)}} \\
\frac{\partial^2 \mathcal{L}_c}{\partial w_s w_t} =& \; 0 \quad \text{for } s \neq t \\
\frac{\partial^2 \mathcal{L}_c}{\partial w_s \partial \boldsymbol{\beta}_{hu}^\top} =& - \left(\boldsymbol{g}_{hi_h(s)}(\boldsymbol{\beta}_h^{(t_{hi_h(s)})}, w_s) \right)^\top \boldsymbol{\Omega}_{hi_h(s)}^{-1} \boldsymbol{J}_{\boldsymbol{\beta}_{hu},hi_h(s)}(\boldsymbol{\beta}_h^{(t_{hi_h(s)})}, w_s) \\
& + \left(\frac{\partial \boldsymbol{g}_{hi_h(s)}(\boldsymbol{\beta}_h^{(t_{hi_h(s)})}, w_s)}{\partial \boldsymbol{\beta}_{hu}^\top} \right)^\top \boldsymbol{\Omega}_{hi_h(s)}^{-1} \boldsymbol{r}'_{hi_h(s)t_{hi_h(s)}} \\
& \text{for } h \in \mathcal{H}_s \text{ and } u \in \{0, t_{hi_h(s)}\} \\
\frac{\partial^2 \mathcal{L}_c}{\partial w_s \partial \boldsymbol{\beta}_{hu}^\top} =& \; \mathbf{0}_{1,p_{hu}} \quad \text{for } h \notin \mathcal{H}_s \text{ or } h \in \mathcal{H}_s \text{ and } u \notin \{0, t_{hi_h(s)}\} \\
\frac{\partial^2 \mathcal{L}_c}{\partial w_s \partial \boldsymbol{\theta}_0^\top} =& \; \mathbf{0}_{1,q} \\
\frac{\partial^2 \mathcal{L}_c}{\partial w_s \partial \varsigma_{hk}} =& - \left(\boldsymbol{g}_{hi_h(s)}(\boldsymbol{\beta}_h^{(t_{hi_h(s)})}, w_s) \right)^\top \boldsymbol{\Omega}_{hi_h(s)}^{-1} \frac{\partial \boldsymbol{\Omega}_{hi_h(s)}}{\partial \varsigma_{hk}} \boldsymbol{\Omega}_{hi_h(s)}^{-1} \boldsymbol{r}'_{hi_h(s)t_{hi_h(s)}} \\
& \text{for } h \in \mathcal{H}_s \\
\frac{\partial^2 \mathcal{L}_c}{\partial w_s \partial \varsigma_{hk}} =& \; 0 \quad \text{for } h \notin \mathcal{H}_s.
\end{aligned}$$

The log-likelihood function for design/measured log-yields $\widetilde{w}_s$ with $s \in \mathcal{S}$ corresponding to the calibration data in this errors-in-variables approach is given up to an additive constant

by

$$\mathcal{L}_w(\boldsymbol{w}) = -\frac{1}{2}\sum_{s=1}^{|\mathcal{S}|}\left(\widetilde{w}_s - w_s\right)^2/\sigma_s^2$$

with partial derivatives given by

$$\frac{\partial \mathcal{L}_w}{\partial w_s} = (\widetilde{w}_s - w_s)/\sigma_s^2$$

and second partial derivatives given by

$$\frac{\partial^2 \mathcal{L}_w}{\partial w_s^2} = -1/\sigma_s^2$$

$$\frac{\partial^2 \mathcal{L}_w}{\partial w_s \partial w_t} = 0 \;\text{ for } s \neq t\,.$$

Incorporating the above calculations involving new parameters $\boldsymbol{w}$ results in modifications to the previously computed information matrix from the combined data in terms of the parameters $(\boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{\theta}_0, \boldsymbol{\varsigma})$:

$$\mathcal{I}^{c,0,w} = \begin{bmatrix} \mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\beta}} + \mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\beta}} & \mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{w}} & \mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\theta}_0} & \mathbf{0}_{p,v} \\ \left(\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{w}}\right)^\top & \mathcal{I}^c_{\boldsymbol{w},\boldsymbol{w}} + \mathcal{I}^w_{\boldsymbol{w},\boldsymbol{w}} & \mathbf{0}_{|\mathcal{S}|,q} & \mathbf{0}_{|\mathcal{S}|,v} \\ \left(\mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^\top & \mathbf{0}_{q,|\mathcal{S}|} & \mathcal{I}^0_{\boldsymbol{\theta}_0,\boldsymbol{\theta}_0} & \mathbf{0}_{q,v} \\ \mathbf{0}_{v,p} & \mathbf{0}_{v,|\mathcal{S}|} & \mathbf{0}_{v,q} & \mathcal{I}^c_{\boldsymbol{\varsigma},\boldsymbol{\varsigma}} + \mathcal{I}^0_{\boldsymbol{\varsigma},\boldsymbol{\varsigma}} \end{bmatrix}$$

where $|\mathcal{S}|$ is the number of unique sources contained in the calibration dataset(s), and

$$\mathcal{I}^c_{\boldsymbol{w},\boldsymbol{w}} = \text{Diag}\left(\mathcal{I}^c_{w_1,w_1}, \mathcal{I}^c_{w_2,w_2}, \ldots, \mathcal{I}^c_{w_{|\mathcal{S}|},w_{|\mathcal{S}|}}\right)$$

$$\mathcal{I}^c_{w_s,w_s} = \sum_{h\in\mathcal{H}_s}\left(\boldsymbol{g}_{hi_h(s)}(\boldsymbol{\beta}_h^{(t_{hi_h(s)})}, w_s)\right)^\top \boldsymbol{\Omega}^{-1}_{hi_h(s)}\boldsymbol{g}_{hi_h(s)}(\boldsymbol{\beta}_h^{(t_{hi_h(s)})}, w_s)$$

$$\mathcal{I}^c_{\boldsymbol{\beta}_h,\boldsymbol{w}} = \left[\mathcal{I}^c_{\boldsymbol{\beta}_{hu},w_s}\right]_{u=0,1,\ldots,T_h,s=1,\ldots,|\mathcal{S}|}$$

$$\mathcal{I}^c_{\boldsymbol{\beta}_{hu},w_s} = \begin{cases} \boldsymbol{J}^\top_{\boldsymbol{\beta}_{hu},hi_h(s)}(\boldsymbol{\beta}_h^{(t_{hi_h(s)})}, w_s)\boldsymbol{\Omega}^{-1}_{hi_h(s)}\boldsymbol{g}_{hi_h(s)}(\boldsymbol{\beta}_h^{(t_{hi_h(s)})}, w_s)\,, & h\in\mathcal{H}_s\,,\; u\in\{0, t_{hi_h(s)}\} \\ \mathbf{0}_{p_{hu},1}\,, & h\notin\mathcal{H}_s \text{ or} \\ & h\in\mathcal{H}_s\,,\; u\notin\{0, t_{hi_h(s)}\} \end{cases}$$

$$\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{w}} = \left[\mathcal{I}^c_{\boldsymbol{\beta}_h,\boldsymbol{w}}\right]_{h=1,\ldots,M}$$

$$\mathcal{I}^w_{\boldsymbol{w},\boldsymbol{w}} = \text{Diag}\left(1/\sigma_1^2, 1/\sigma_2^2, \ldots, 1/\sigma^2_{|\mathcal{S}|}\right)\,.$$

Analogous to the previous development, the asymptotic covariance matrix of the MLE $\widehat{\boldsymbol{\theta}}_0$ of $\boldsymbol{\theta}_0$ is given by $\left(\mathcal{I}^{c,w,0}_{\boldsymbol{\theta}_0|(\boldsymbol{\beta},\boldsymbol{w})}\right)^{-1}$, where

$$\mathcal{I}^{c,w,0}_{\boldsymbol{\theta}_0|(\boldsymbol{\beta},\boldsymbol{w})} = \mathcal{I}^0_{\boldsymbol{\theta}_0,\boldsymbol{\theta}_0} - \left(\mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^\top\left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1}\mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\theta}_0}$$

$$- \left(\mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^{\top} \left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1} \mathcal{I}^{c}_{\boldsymbol{\beta},\boldsymbol{w}} \left(\mathcal{I}^{c,w,0}_{\boldsymbol{w}|\boldsymbol{\beta}}\right)^{-1} \left(\mathcal{I}^{c}_{\boldsymbol{\beta},\boldsymbol{w}}\right)^{\top} \left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1} \mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\theta}_0}$$

$$= \mathcal{I}^{c,0}_{\boldsymbol{\theta}_0|\boldsymbol{\beta}} - \left(\mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^{\top} \left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1} \mathcal{I}^{c}_{\boldsymbol{\beta},\boldsymbol{w}} \left(\mathcal{I}^{c,w,0}_{\boldsymbol{w}|\boldsymbol{\beta}}\right)^{-1} \left(\mathcal{I}^{c}_{\boldsymbol{\beta},\boldsymbol{w}}\right)^{\top} \left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1} \mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\theta}_0} \tag{21}$$

for

$$\mathcal{I}^{c,w,0}_{\boldsymbol{w}|\boldsymbol{\beta}} = \mathcal{I}^{c}_{\boldsymbol{w},\boldsymbol{w}} + \mathcal{I}^{w}_{\boldsymbol{w},\boldsymbol{w}} - \left(\mathcal{I}^{c}_{\boldsymbol{\beta},\boldsymbol{w}}\right)^{\top} \left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1} \mathcal{I}^{c}_{\boldsymbol{\beta},\boldsymbol{w}} \,.$$

Computation of the second term in (21) simplifies by noting that

$$\left(\mathcal{I}^{c}_{\boldsymbol{\beta},\boldsymbol{w}}\right)^{\top} \left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1} \mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\theta}_0} = \sum_{h=1}^{M} \left(\mathcal{I}^{c}_{\boldsymbol{\beta}_h,\boldsymbol{w}}\right)^{\top} \left(\mathcal{I}^{c,0}_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right)^{-1} \mathcal{I}^{0}_{\boldsymbol{\beta}_h,\boldsymbol{\theta}_0}$$

$$\left(\mathcal{I}^{c}_{\boldsymbol{\beta},\boldsymbol{w}}\right)^{\top} \left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1} \mathcal{I}^{c}_{\boldsymbol{\beta},\boldsymbol{w}} = \sum_{h=1}^{M} \left(\mathcal{I}^{c}_{\boldsymbol{\beta}_h,\boldsymbol{w}}\right)^{\top} \left(\mathcal{I}^{c,0}_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right)^{-1} \mathcal{I}^{c}_{\boldsymbol{\beta}_h,\boldsymbol{w}} \,.$$

From (21) it is seen that the second term on the right-hand side of the equality represents the loss of information in estimating $\boldsymbol{\theta}_0$ due to estimating log-yields $\boldsymbol{w}$ rather than assuming they are known and fixed at their design/measured values. Expression (21) can also be written more elegantly as follows:

$$\mathcal{I}^{c,w,0}_{\boldsymbol{\theta}_0|(\boldsymbol{\beta},\boldsymbol{w})} = \mathcal{I}^{0}_{\boldsymbol{\theta}_0,\boldsymbol{\theta}_0} - \left(\mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^{\top} \left(\mathcal{I}^{c,w,0}_{\boldsymbol{\beta}|\boldsymbol{w}}\right)^{-1} \mathcal{I}^{0}_{\boldsymbol{\beta},\boldsymbol{\theta}_0}$$

for

$$\mathcal{I}^{c,w,0}_{\boldsymbol{\beta}|\boldsymbol{w}} = \mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}} - \mathcal{I}^{c}_{\boldsymbol{\beta},\boldsymbol{w}} \left(\mathcal{I}^{c}_{\boldsymbol{w},\boldsymbol{w}} + \mathcal{I}^{w}_{\boldsymbol{w},\boldsymbol{w}}\right)^{-1} \left(\mathcal{I}^{c}_{\boldsymbol{\beta},\boldsymbol{w}}\right)^{\top} \,.$$

In practice, increased standard errors in estimates of $\boldsymbol{\theta}_0$ suggested by expression (21) for the errors-in-variables treatment may be mitigated by the fact that the maximum likelihood parameter estimates used to compute the information matrices will be different than those obtained from the assumed known yields analysis, resulting in analysis-dependent values of $\mathcal{I}^{c,0}_{\boldsymbol{\theta}_0|\boldsymbol{\beta}}$. The consequence is an inability to arrive at any firm conclusion regarding the impact of errors-in-variables treatment on uncertainty quantification for $\boldsymbol{\theta}_0$.

A.4.1 Example: Linear Model with Errors-in-Variables

Returning to the linear model formulation, define

$$\overline{\boldsymbol{x}}_{hu} = \boldsymbol{X}^{\top}_{hu} \boldsymbol{\Omega}^{-1}_{hu} \mathbf{1}_{n_{hu}}$$

$$\widetilde{X}_{hu} = \overline{\boldsymbol{x}}^{\top}_{hu} \left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\boldsymbol{\mu}_h}\right)^{-1} \mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\omega_h}$$

and

$$q_h(u,v) = \overline{\boldsymbol{x}}^{\top}_{hu} \left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\boldsymbol{\mu}_h}\right)^{-1} \overline{\boldsymbol{x}}_{hv}$$
$$+ \left(\widetilde{X}_{hu} - W_u \mathbf{1}^{\top}_{n_{hu}} \boldsymbol{\Omega}^{-1}_{hu} \mathbf{1}_{n_{hu}}\right) \left(\mathcal{I}^{c,0}_{\omega_h|\boldsymbol{\mu}_h}\right)^{-1} \left(\widetilde{X}_{hv} - W_v \mathbf{1}^{\top}_{n_{hv}} \boldsymbol{\Omega}^{-1}_{hv} \mathbf{1}_{n_{hv}}\right) \,.$$

Consider the $M \times |\mathcal{S}|$ matrix $\boldsymbol{Q}_0$ defined to have u-th column

$$\left[\left(\omega_1 q_1(0,u) \quad \omega_2 q_2(0,u) \quad \cdots \quad \omega_M q_M(0,u)\right) \boldsymbol{C}_u\right]^\top$$

for $u = 1, \ldots, |\mathcal{S}|$, where

$$\boldsymbol{C}_u = \sum_{h \in \mathcal{H}_u} \boldsymbol{e}_h \boldsymbol{e}_h^\top$$

with $\boldsymbol{e}_h$ the M-vector having a 1 in the h-th position and zeros elsewhere. Then

$$V_{\text{eiv}}^{-1}\left[\widehat{W}_0\right] = V^{-1}\left[\widehat{W}_0\right] - \boldsymbol{\omega}^\top \boldsymbol{Q}_0 \left(\mathcal{I}^{c,w,0}_{\boldsymbol{w}|\boldsymbol{\beta}}\right)^{-1} \boldsymbol{Q}_0^\top \boldsymbol{\omega}$$

is the asymptotic precision (inverse variance) of the log-yield MLE adjusted for errors-in-variables (eiv), where the (u,v) element of $\mathcal{I}^{c,w,0}_{\boldsymbol{w}|\boldsymbol{\beta}}$ is given by

$$\left[\sum_{h \in \mathcal{H}_u} \omega_h^2 \, \mathbf{1}_{n_{hu}}^\top \boldsymbol{\Omega}_{hu}^{-1} \mathbf{1}_{n_{hu}} + \frac{1}{\sigma_u^2}\right] \chi\left[u = v\right] - \sum_{h \in \mathcal{H}_u \cap \mathcal{H}_v} \omega_h^2 \, q_h(u,v)$$

for $u, v = 1, \ldots, |\mathcal{S}|$, and $\chi[A]$ is the indicator function of condition A.

A.5 Covariance Matrix Specification and Differentiation

Optimization of the log-likelihood function involves the parameters of several covariance matrices for each sensor type h, namely $\boldsymbol{\Sigma}^{(S)}_{h1}, \ldots, \boldsymbol{\Sigma}^{(S)}_{hR_h}$, $\boldsymbol{\Sigma}^{(P)}_{h1}, \ldots, \boldsymbol{\Sigma}^{(P)}_{hR_h}$, and $\boldsymbol{\Sigma}_h$. To begin, treatment of possible biases in the source and path (within source) physical/empirical models will be simplified to remove covariate information from the model for source and path effects. That is, both $\boldsymbol{Z}_{hir,j}$ and $\boldsymbol{Z}_{hijr}$ are set to $\mathbf{1}_{n_{hijr}}$. This implies that for each sensor measurement type r, the random effects vectors $\boldsymbol{b}^{(S)}_{hr}$ and $\boldsymbol{b}^{(P)}_{hr}$ simplify to scalars (i.e. $q_{S,hr} = q_{P,hr} = 1$) with the result

$$\boldsymbol{\Sigma}^{(S)}_{hr} = \sigma^2_{S_{hr}} \quad \text{and} \quad \boldsymbol{\Sigma}^{(P)}_{hr} = \sigma^2_{P_{hr}} \, .$$

Variation in the source ("S") and path-within-source ("P") biases is allowed to depend on sensor type h and sensor measurement type r, implying up to a total of $2\sum_{h=1}^{M} R_h$ variance parameters from the random effects model to be optimized if a two-level random effects model is fit to all responses.

To facilitate unconstrained optimization, these variances are reparameterized using the log-transform: $\tau_{S_{hr}} = \log\left(\sigma^2_{S_{hr}}\right)$ and $\tau_{P_{hr}} = \log\left(\sigma^2_{P_{hr}}\right)$. The partial derivatives of $\boldsymbol{\Omega}_{hi}$ corresponding to the source and path errors are given as follows,

$$\frac{\partial \boldsymbol{\Omega}_{hi}}{\partial \tau_{S_{hr}}} = \frac{\partial \boldsymbol{\Xi}^{(S)}_{hi}}{\partial \tau_{S_{hr}}} = \text{Diag}\left(\mathbf{0}_{n_{hi1}, n_{hi1}}, \ldots, \sigma^2_{S_{hr}} \boldsymbol{Z}^S_{hir} \left(\boldsymbol{Z}^S_{hir}\right)^\top, \ldots, \mathbf{0}_{n_{hiR_h}, n_{hiR_h}}\right)$$

$$\frac{\partial \boldsymbol{\Omega}_{hi}}{\partial \tau_{P_{hr}}} = \frac{\partial \boldsymbol{\Xi}^{(P)}_{hi}}{\partial \tau_{P_{hr}}} = \text{Diag}\left(\mathbf{0}_{n_{hi1}, n_{hi1}}, \ldots, \sigma^2_{P_{hr}} \boldsymbol{Z}^P_{hir} \left(\boldsymbol{Z}^P_{hir}\right)^\top, \ldots, \mathbf{0}_{n_{hiR_h}, n_{hiR_h}}\right) \, ,$$

where

$$\boldsymbol{Z}^S_{hir}\left(\boldsymbol{Z}^S_{hir}\right)^\top = \begin{cases} \boldsymbol{J}_{n_{hir}} \, , & r \in \mathcal{R}^S_h \\ \mathbf{0}_{n_{hir}, n_{hir}} \, , & r \notin \mathcal{R}^S_h \end{cases}$$

$$\boldsymbol{Z}_{hir}^{P}\left(\boldsymbol{Z}_{hir}^{P}\right)^{\top}=\begin{cases}\text{Diag}\left(\boldsymbol{J}_{n_{hi1r}},\boldsymbol{J}_{n_{hi2r}},\ldots,\boldsymbol{J}_{n_{hiM_{hi}r}}\right), & r\in\mathcal{R}_{h}^{P}\\ \boldsymbol{0}_{n_{hir},n_{hir}}, & r\notin\mathcal{R}_{h}^{P}\end{cases}$$

with $\boldsymbol{J}_m$ the $m\times m$ matrix of ones. To obtain an unconstrained parameterization of $\boldsymbol{\Sigma}_h$, set $\boldsymbol{\Sigma}_h=\boldsymbol{L}_h^{\top}\boldsymbol{L}_h$ where $\boldsymbol{L}_h$ is an upper triangular matrix,

$$\boldsymbol{L}_h=\begin{bmatrix}l_{h,11} & l_{h,12} & l_{h,13} & \cdots & l_{h,1R_h}\\ \cdot & l_{h,22} & l_{h,23} & \cdots & l_{h,2R_h}\\ \cdot & \cdot & l_{h,33} & \cdots & l_{h,3R_h}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \cdot & \cdot & \cdot & \cdots & l_{h,R_hR_h}\end{bmatrix}.$$

This factorization is unique if the diagonal elements of $\boldsymbol{L}_h$ are required to be positive. Hence consider $\ell_{h,rr}=\log\left(l_{h,rr}\right)$, $r=1,\ldots,R_h$, as a reparameterization of the diagonal elements to achieve an unconstrained set of parameters. For υ an element of $\boldsymbol{L}_h$, note that

$$\frac{\partial\boldsymbol{\Omega}_{hi}}{\partial\upsilon}=\frac{\partial\boldsymbol{\Sigma}_{hi}}{\partial\upsilon}=\sum_{r_1\leq r_2}\frac{\partial\boldsymbol{\Sigma}_{hi}}{\partial\sigma_{h,r_1r_2}}\frac{\partial\sigma_{h,r_1r_2}}{\partial\upsilon}$$

where

$$\left[\frac{\partial\boldsymbol{\Sigma}_{hi}}{\partial\sigma_{h,r_1r_2}}\right]_{r_1,r_2}=\sum_{(k_1,k_2)\in\mathcal{J}_{r_1,r_2}^{hi}}\boldsymbol{e}_{hir_1,k_1}\boldsymbol{e}_{hir_2,k_2}^{\top}$$

$$\left[\frac{\partial\boldsymbol{\Sigma}_{hi}}{\partial\sigma_{h,r_1r_2}}\right]_{r_2,r_1}=\sum_{(k_1,k_2)\in\mathcal{J}_{r_1,r_2}^{hi}}\boldsymbol{e}_{hir_2,k_2}\boldsymbol{e}_{hir_1,k_1}^{\top}$$

$$\left[\frac{\partial\boldsymbol{\Sigma}_{hi}}{\partial\sigma_{h,r_1r_2}}\right]_{rs}=\boldsymbol{0}_{n_{hir},n_{his}},\ (r,s)\notin\{(r_1,r_2),(r_2,r_1)\}$$

and

$$\frac{\partial\sigma_{h,r_1r_2}}{\partial\ell_{h,rr}}=\begin{cases}2\,l_{h,rr}^{2}, & r_1=r,\ r_2=r\\ l_{h,rr}\,l_{h,rr_2}, & r_1=r,\ r_2>r\\ 0, & \text{otherwise}\end{cases}$$

$$\frac{\partial\sigma_{h,r_1r_2}}{\partial\ell_{h,rs}}=\begin{cases}l_{h,rr_1} & r\leq r_1<s,\ r_2=s\\ 2\,l_{h,rs}, & r_1=s,\ r_2=s\\ l_{h,rr_2}, & r_1=s,\ r_2>s\\ 0, & \text{otherwise}\end{cases}$$

for $s>r$. In this treatment,

$$\boldsymbol{\varsigma}_h=\left(\tau_{S_{h1}},\ldots,\tau_{S_{hR_h}},\tau_{P_{h1}},\ldots,\tau_{P_{hR_h}},\ell_{h,11},l_{h,12},\ldots,l_{h,1R_h},\ldots,\ell_{h,R_hR_h}\right).$$