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: 2025-04-22 (rev.4)

Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Multi-phenomenology Yield Characterization

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

[1]*Los Alamos National Laboratory*
[2]*Desert Research Institute*

April 16, 2025

1 Introduction

This report demonstrates 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 report, 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.

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 unclassified data used in this analysis were kindly provided by Sean Ford (LLNL) as well as other unclassified data sources as indicated.

2 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 benchmark 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 has been 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 notation $E_{P,hijr}$ in (2) assumes that paths are *nested* within sources, and is replaced by $E_{P,hjr}$ if instead paths are treated as *crossed* among sources, as described in the following paragraph. Traditional estimation of seismic *magnitudes* involved applying fixed station corrections calculated from an established network of stations to the estimated magnitude of an event at the station in question, that accounted for unmodeled source-to-sensor path effects (Vaněk and Kondorskaya, 1974). In (2) these corrections are instead implemented as a random path effect in the error model for an observed signature, to allow for explosions occurring at previously untested locations and possibly new local and regional seismic and acoustic networks.

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, $\log a_{hijrk}(\omega)$. The situation is more nuanced for path bias. In one scenario, paths are considered to be *nested* within sources. In other words, 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 from explosion i. In the more general scenario, paths are considered to be *crossed* among sources. Specifically, for path j observing (nearly) co-located explosions i_1 and i_2, a common bias $E_{P,hjr}$ in the physical path terms $\log P_{hi_1jr}(\omega; \boldsymbol{\beta}_{P,hr}, \boldsymbol{v}_{P,hi_1j})$ and $\log P_{hi_2jr}(\omega; \boldsymbol{\beta}_{P,hr}, \boldsymbol{v}_{P,hi_2j})$ applies to all data $(\log a_{hi_1jrk}(\omega), \log a_{hi_2jrk}(\omega))$ observed along path j. If data are observed along path j for any additional explosions, then application of the common path bias applies to these observed data as well. Changes in source term bias apply to all data for a given source, while changes in path term bias apply only to data observed on that path, in different ways depending on whether it is assumed to be *nested* within source or *crossed* among sources.

Note that this involves a philosophical change in the interpretation of the dynamic bias terms $E_{S,hir}$ and $E_{P,hijr}$ (or $E_{P,hjr}$), 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) pre-shot design yield as obtained from calculations for a device prior to testing, such as for chemical explosions, and (b) post-shot measurement data, such as from radiochemistry on drillback samples after a nuclear test. Both sources are subject to uncertainty: in (a), computations are imperfect, chemical-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 benchmark 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 benchmark 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.

3 Model Interpretation and Application

In the following, it is assumed that paths are modeled as *crossed* among sources, as the paths *nested* within sources model is mathematically a special case of this approach. Furthermore, the *crossed* paths model is most relevant in the historical context of monitoring foreign and domestic test sites using established networks of teleseismic and acoustic sensors. The *nested* paths within sources model is only relevant to settings in which no source-to-sensor propagation pathways are replicated across distinct sources, which is rare in practice but can apply to temporary local and regional seismic and acoustic networks used for individual small-scale experiments. Equation (2) can be written as

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

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 $\{E_{P,hjr}\}_j$ are taken to be iid Gaussian variables having mean zero and variance $\tau^2_{P,hr}$; and the collection of random variables $\{\{E_{S,hir}\}_i, \{E_{P,hjr}\}_j\}$ is assumed to be mutually independent. Note that for a *crossed* path j, cross-correlation is induced among the applicable sources; the *nested* path model merely zeroes these cross-correlations by construction.

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 benchmark 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 4. 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 $E_{P,hjr}$ 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

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

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 benchmark data. For example, the source model for an above ground explosion includes a model component to represent energy coupling from air into rock. Appropriate benchmark 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 benchmark 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 benchmark data needed for a validation analysis consist of at least two explosions and at least two source-to-sensor pathways (for *nested* paths, at least one explosion must be observed by at least two paths), and at least one site 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 benchmark 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 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 path incorporates both contributors $E_{S,hir}$ and $E_{P,hjr}$ to dynamic bias, $\mathrm{Cov}[y_{hijrk}(\omega), y_{hijrk'}(\omega)] = \tau^2_{S,hr} + \tau^2_{P,hr}$, $k \neq k'$. The covariance between observations from two different explosions i and i' measured along the same path j, given sensor type h and measurement type r, is $\mathrm{Cov}[y_{hijrk}(\omega), y_{hi'jrk}(\omega)] = \tau^2_{P,hr}$, $i \neq i'$, because these data probabilistically co-vary only through a shared path error $E_{P,hjr}$. For *nested* paths, $\mathrm{Cov}[y_{hijrk}(\omega), y_{hi'jrk}(\omega)] = 0$, $i \neq i'$, by construction. 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 benchmark 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).

4 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.

4.1 Benchmark Data

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

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

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 2. 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 $\boldsymbol{E}_{P,hjr}$ are assumed to take the form of linear models,

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

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-level random coefficients applied to the columns of the fixed covariate matrix $\boldsymbol{Z}_{hjr,i}$. 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 random coefficient vectors. This statistical model for dynamic bias is referred to as a *random effects* model (Pinheiro and Bates, 2000).

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, observational errors 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,

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

combined with all distributional assumptions stated above. Three reduced models are also of interest; the first is obtained by eliminating the path random effects,

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

the second is obtained by eliminating the source random effects,

$$\boldsymbol{y}_{hijr} = \boldsymbol{f}_{hijr}(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{v}_{hij}) + \boldsymbol{Z}_{hjr,i}\boldsymbol{b}_{hr}^{(P)} + \boldsymbol{\epsilon}_{hijr}\,, \tag{8}$$

and the third is obtained by eliminating all random effects,

$$\boldsymbol{y}_{hijr} = \boldsymbol{f}_{hijr}(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{v}_{hij}) + \boldsymbol{\epsilon}_{hijr}\,. \tag{9}$$

If path random effects are assumed to be *nested* within the source random effects, the latter must be included in the model if the former are present, precluding Equation (8) as a viable model.

For the *crossed* dynamic path model, sources are partitioned into groups defining collections of data that are assumed to be mutually statistically independent. Specifically, any two sources that contain at least one common path are assigned to the same group, while distinct groups contain no sources having any paths in common. Additional details are provided in Appendix A.1.

4.2 New Event Data

New event data is modeled statistically in a fashion analogous to the benchmark 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 benchmark 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,

$$\boldsymbol{y}_{h0jr} = \boldsymbol{f}_{h0jr}((\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0), \boldsymbol{v}'_{h0j}) + \boldsymbol{Z}_{h0r,j}\boldsymbol{b}_{hr}^{(S)} + \boldsymbol{Z}_{hjr,0}\boldsymbol{b}_{hr}^{(P)} + \boldsymbol{\epsilon}_{h0jr}\,.$$

Equations (7)–(9) are also modified in the same fashion. New event data is assumed to be distributed independently of benchmark data, facilitating rapid inference of $\boldsymbol{\theta}_0$. However, in the *crossed* path model, if one or more paths observing new event data also observe data for at least one benchmark source, a more computationally intensive analysis may be chosen to accommodate the resulting cross-correlations. Additional details are provided in Appendix A.2.

5 Application

The primary purpose of the worked example to follow is to highlight the many capabilities of the MultiPEM Toolbox software, licensed as open source on GitHub: `https://github.com/lanl/MultiPEM_Toolbox.git`. Those capabilities include the ability to

a) handle linear and nonlinear forward models, fusing data from multiple sensor types,

b) handle uncertainties in measured and "equivalent" nuclear yields,

c) quantify source effects and source-to-sensor path effects with either fixed effects or random effects modeling,

d) carry out traditional frequentist as well as Bayesian analyses,

e) compare multiple candidate models via standard goodness-of-fit metrics, and

f) provide graphical displays of data fit and residual diagnostic plots.

The development in this report 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 (benchmark data) and Table 5 (new event data) of that article. The surface effects data unique to this report are described below in Section 5.4, with data provided in Table 2.

In this report, source-to-sensor pathways were identified in the seismic and acoustic data sets to populate the dynamic path effect model. A unique path was defined by the geolocation of its endpoints. Sources nearly colocated within an experiment were assigned to the same location, while sensors not exactly colocated were assigned to different paths. 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. An alternative forward model of the optical signatures (Whitaker and Symbalisty, 2009) was also examined. Additional details are given in the following subsections.

5.1 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.

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.

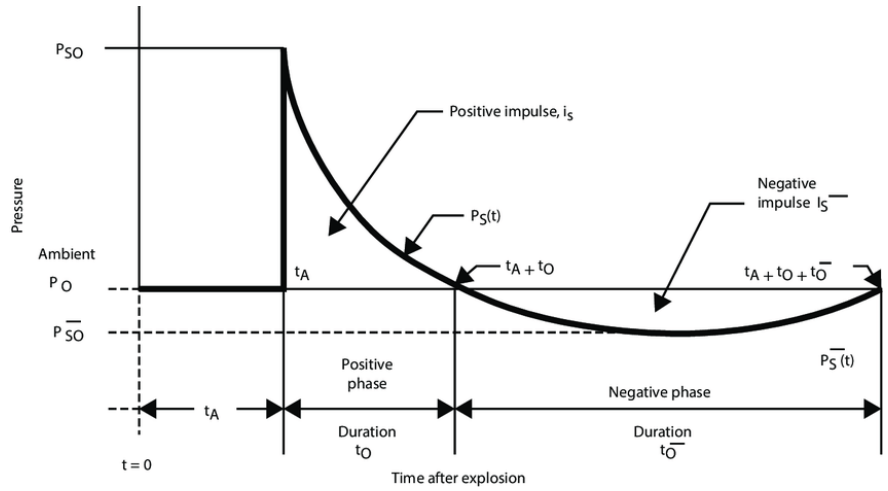

Figure 1: Notional acoustic wave form highlighting relevant features.

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 and path. The overpressure impulse and duration signatures are therefore modeled according to Equation (6) with the random effects covariate matrices $\boldsymbol{Z}_{hir,j}$ and $\boldsymbol{Z}_{hjr,i}$ set to all-ones vectors of the appropriate length.

5.2 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

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		

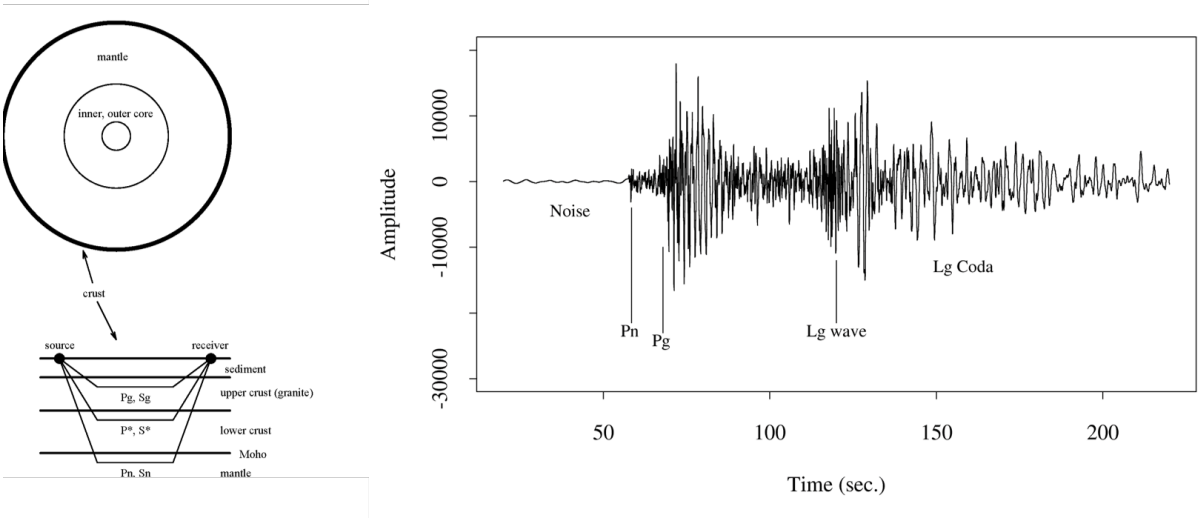

Figure 2: Crustal (regional) seismic wave paths. The notation Pn, Pg, Sn and Lg represents the segment of a seismic wave that follows the associated path.

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 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

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

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

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 path. Both signatures are therefore modeled according to Equation (6) with the random effects covariate matrices $\boldsymbol{Z}_{hir,j}$ and $\boldsymbol{Z}_{hjr,i}$ set to all-ones vectors of the appropriate length.

5.3 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).

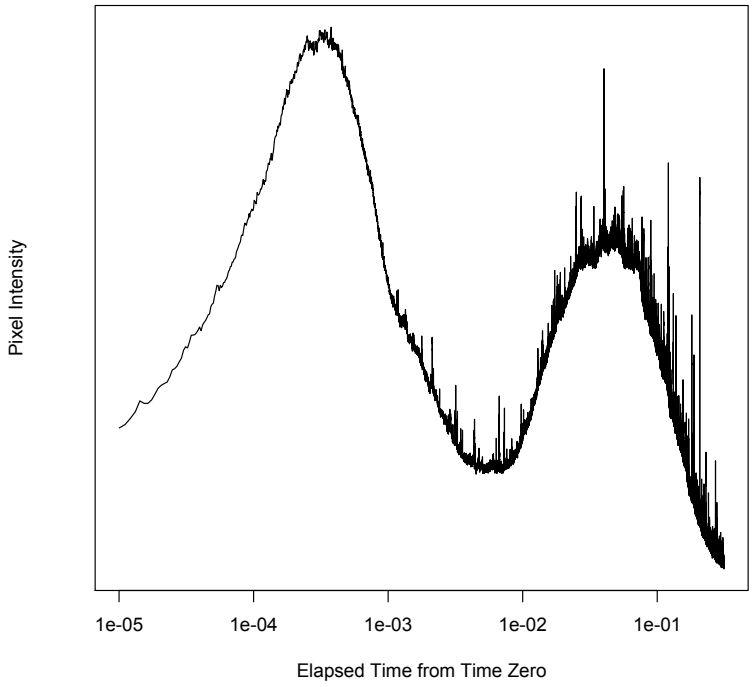

Figure 3: Light flash for Sugar.

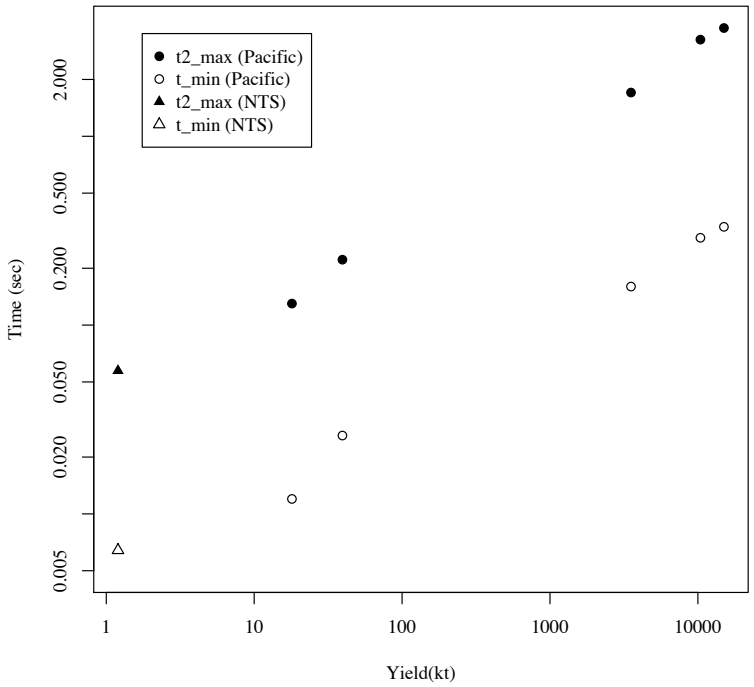

Figure 4: Optical data for near-surface events in Table 2.

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 an optical forward model for optical signatures across a range of HOBs,

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

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

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

Alternatively, Whitaker and Symbalisty (2009) proposed an optical forward model for optical signatures from above-ground explosions,

$$f_{or} = \log(d_{or}) = \beta_{or,1} + \beta_{or,2}\,w + \log(1 + \beta_{or,3}\exp(-(\tilde{h}_s/\beta_{or,4})^2))\,.$$

The scaled height-of-burst $\tilde{h}_s$ is identical to that used in computation of the seismic forward models. Covariates w and h are 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 modeled according to Equation (9).

5.4 Surface Effects

For a nuclear device detonated above ground 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.

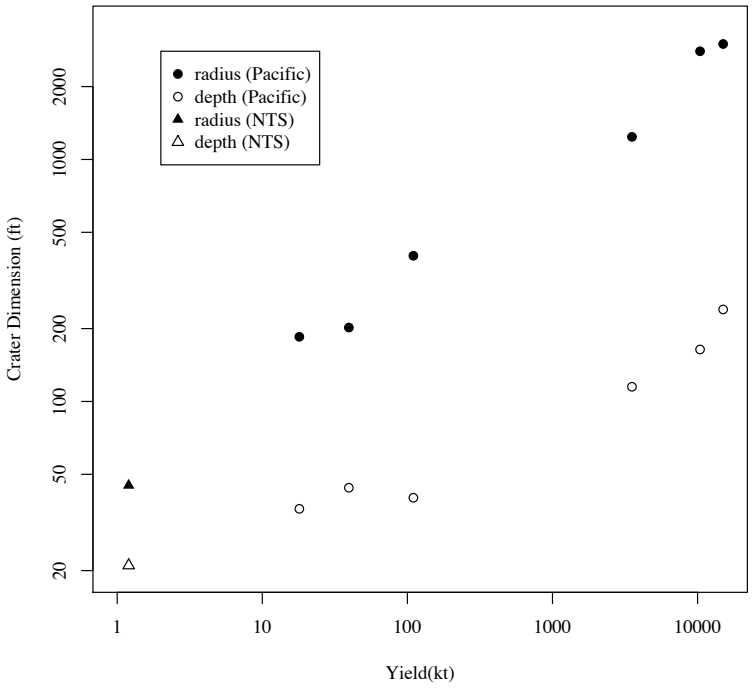

Figure 5: Surface effects data for near-surface events.

As with the optical signatures, crater dimensions are single measurements per event and

are therefore modeled according to Equation (9) 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.

5.5 Methods

Data analysis consists of fitting one of the statistical models (6)-(9) 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 (14) for the benchmark data and (18) for the new event data, with the joint (benchmark and new event data) log-likelihood function given by Equation (20). When the errors-in-variables treatment of benchmark data yields introduced in Section 2 is utilized, the log-likelihood function implied by Equation (3) is added to the joint log-likelihood function as given by Equation (24).

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 (23) 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 benchmark 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 benchmark 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)–(8) 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 4.2 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.

5.6 Analysis

The acoustic and seismic benchmark 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 5.4 were obtained from nuclear explosions without regard to a designated emplacement.

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

Fixed and errors-in-variables treatments of log yields for the optical and surface effects benchmark events are considered. The errors-in-variables approach of Section 2 was implemented by assuming the reported benchmark yields were correct up to 30%. That is, the value of σ_s in Equation (3) was set to $0.3/3 = 0.1$.

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, noting that there are two competing forward models for the optical sensor type. 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 $\boldsymbol{Z}_{hjr,i}$ are taken to be all-ones vectors when fitting error models that include dynamic bias terms (see Equation (6)). 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 4.1) 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 in the example at hand 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. The MultiPEM Toolbox software is capable of handling more realistic priors based on expert knowledge.

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. For higher dimensional parameter spaces, the robust adaptive Metropolis algorithm of Vihola (2012) implemented in the `R` package `adaptMCMC`, may be selected for improved performance by avoiding the potential for instability in the estimated covariance matrix of the target distribution used for proposal generation. In either case, an initial transient sequence of samples is discarded (burnin), and 20,000 production samples are taken and decorrelated by thinning out the chain every 20-th sample.

For both optical signatures, two competing forward models are considered: the Ford (F) and Whitaker-Symbalisty (WS) models as described in Section 5.3. Only the error model of Equation (9) can be fitted in conjunction with these forward model options, so the only available comparison is between the two forward models themselves. Table 4 presents calculated values for the model selection criteria discussed in the previous subsection. For each criterion, **bold** font indicates the chosen forward model. The results indicate that for both fixed and errors-in-variables treatments of benchmark yields, the WS forward model is preferred by every information criterion, and thus it is selected for all subsequent analyses. For the remaining sensor types, only a single forward model is considered, so that all model selection is conducted with regard to the choice of error model. Only Equation (9) is eligible for fitting the signatures of the surface effects sensor type, as was the case above for the optical sensor type.

Table 4: Model selection results for optical sensor type

Benchmark Yields	Forward Model	AIC	BIC	DIC	PIC
Fixed	WS	**−2.57**	**10.94**	**−0.22**	**4.01**
Fixed	F	12.53	21.23	13.47	20.49
Errors-in-variables	WS	**92.98**	**137.22**	**97.81**	**133.95**
Errors-in-variables	F	108.21	148	110.46	142.9

For acoustic, each of Equations (6)–(9) may be considered to model the signatures, as there are multiple measurements per source, and at least two propagation paths possessing multiple measurements (including within source for *nested* paths). The coefficient matrices $\boldsymbol{Z}_{hir,j}$ and $\boldsymbol{Z}_{hjr,i}$ 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. In this table, a check indicates the presence of the indicated dynamic source term in the error model for the indicated signature.

For the dynamic path bias models, "C" and "N" indicate paths that are *crossed* among sources and *nested* within sources, respectively. For *nested* paths, the corresponding dynamic source bias term must also be present in the error model. For each criterion, **bold** font indicates the error model choices for each signature representing the optimum out of all seventeen alternatives considered. For all criteria, selecting Equation (6) with *crossed* paths to model both acoustic signatures is optimal, indicating the presence of clustering by both

Table 5: Model selection results for acoustic sensor type

Dynamic Bias Terms							
Source		Path					
f_{a1}	f_{a2}	f_{a1}	f_{a2}	AIC	BIC	DIC	PIC
✓	✓	C	C	**17.58**	**59.26**	**17.9**	**34.38**
✓	✓	N	N	60.48	103.6	60.54	76.42
✓	✓	C		80.7	121.4	81.59	97.21
✓	✓	N		111	152.3	110.7	125.5
✓	✓		C	48.4	88.3	48.73	64.11
✓	✓		N	59.6	100.3	60.8	76.4
✓	✓			109.3	148	111.4	126.4
		C	C	370.2	436.7	371.4	386.8
✓		C		251	313.2	250.5	264.9
✓		N		281.4	343.8	281	295.3
✓				279.7	337.9	280.6	294.6
		C		400.4	466	399.9	413.5
	✓		C	164.6	226.8	165.8	181
	✓		N	174.8	236.8	175	189.7
	✓			227.3	285.7	228.2	242
			C	381.2	445.3	381.2	395.1
				410.5	473.1	410.9	424

source and path for each signature. In addition, the acoustic data arose from a controlled experiment with crossed paths, and the model comparison metrics in Table 5 show that force-fitting an incorrect (e.g. nested path) error structure in the model produces noticeably inferior results.

Clustering in acoustic signatures by source and path is illustrated as follows. The source and path random effects $\boldsymbol{b}_{hr}^{(S)}$ and $\boldsymbol{b}_{hr}^{(P)}$ of Section 4.1 have Gaussian posterior distributions, conditional on all forward and error model parameters. For each source and path, these effects are estimated by their posterior means with uncertainty provided by their posterior standard errors, which in this analysis are all scalars. Figure 6 shows 95% credible intervals for every source effect on the logarithm of overpressure duration. Sources for which the credible intervals indicate a statistically significant non-zero effect are labeled on the x-axis. ML estimates of the forward and error model parameters are substituted into the posterior mean and standard error expressions.

Figure 7 illustrates the presence of the largest source and path effects for observed logarithm of overpressure duration. For highlighted sources or paths, observed signatures cluster above or below the 45^o line indicating an ideal fit to the forward model. Large clusters may introduce bias into estimates of the forward model coefficients, due to the presence of common and unmodeled source or path biases that correlate the observations. This is alleviated by including the source and path random effects components in the error model as in Equation (6), which has the practical effect of downweighting large clusters of observations by accounting for the dependency among them. The posterior estimates of source

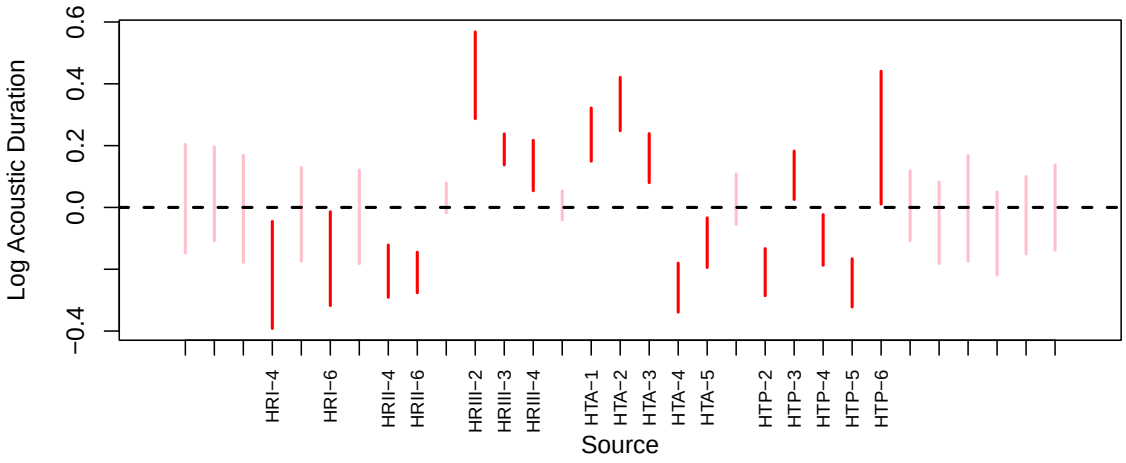

Figure 6: 95% credible intervals for source random effects on $\log(d_{a2})$ of Section 5.1.

i and path j random effects, denoted $\hat{\boldsymbol{b}}^{(S)}_{hr,i}$ and $\hat{\boldsymbol{b}}^{(P)}_{hr,j}$, can be used to compute observed signatures adjusted by source and path effects. For example, with reference to Equation (6), source- and path-adjusted logarithm of overpressure duration are given by $\boldsymbol{y}_{hij2} - \boldsymbol{Z}_{hi2,j}\hat{\boldsymbol{b}}^{(S)}_{h2,i}$ and $\boldsymbol{y}_{hij2} - \boldsymbol{Z}_{hj2,i}\hat{\boldsymbol{b}}^{(P)}_{h2,j}$, respectively. Observed $\log(d_{a2})$ is adjusted by path effects in the left panel, and by source effects in the right panel, to allow isolation of the indicated clustering for graphical display.

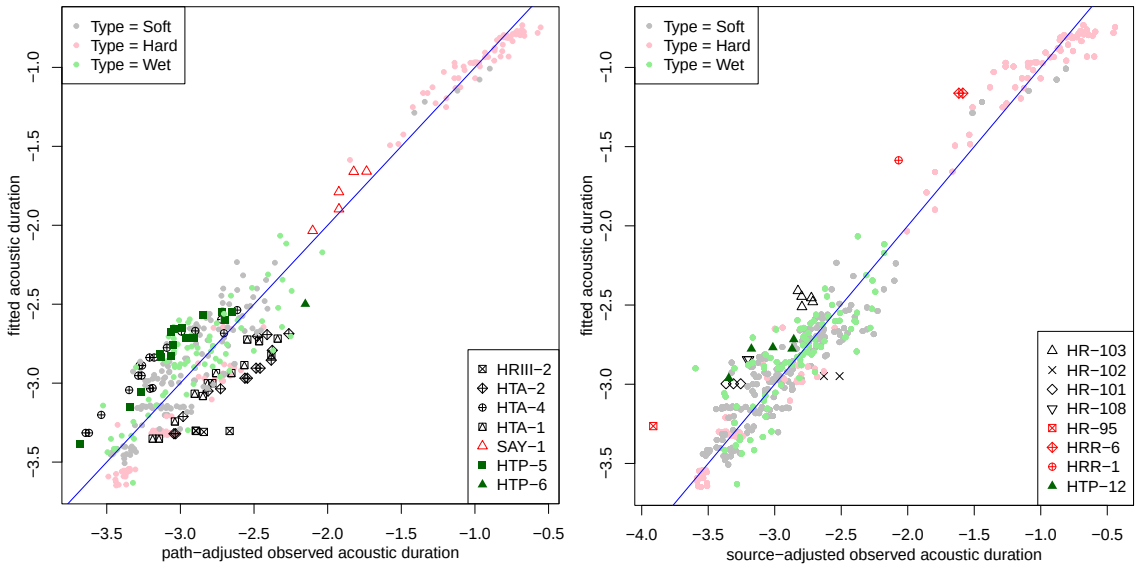

Figure 7: Fitted versus observed $\log(d_{a2})$ isolating larger estimated source effects (left) and path effects (right) shown by emplacement (soft, hard, or wet rock).

For seismic, each of Equations (6)–(9) may be considered for each signature, as with acoustic. Table 6 presents calculated values for the model selection criteria discussed in the previous subsection. For all criteria, selecting Equation (6) with *crossed* paths to model both seismic signatures is optimal, as is also seen with both acoustic signatures.

Table 6: Model selection results for seismic sensor type

Dynamic Bias Terms							
Source		Path					
f_{s1}	f_{s2}	f_{s1}	f_{s2}	AIC	BIC	DIC	PIC
✓	✓	C	C	**1053**	**1163**	**1064**	**1076**
✓	✓	N	N	1315	1438	1325	1336
✓	✓	C		1161	1276	1163	1187
✓	✓	N		1345	1465	1336	1340
✓	✓		C	1214	1328	1201	1196
✓	✓		N	1314	1434	1310	1319
✓	✓			1347	1464	1348	1352
		C	C	1147	1275	1126	1141
✓		C		1194	1336	1175	1188
✓		N		1378	1522	1373	1397
✓				1380	1520	1359	1370
		C		1231	1372	1211	1225
	✓		C	1254	1396	1252	1278
	✓		N	1356	1500	1335	1342
	✓			1389	1529	1383	1392
			C	1314	1455	1296	1311
				1426	1572	1405	1417

The "new event" for which the quantities of interest $\boldsymbol{\theta}_0 = (w, h)$ are inferred is the above-ground nuclear test Sugar, conducted at the Nevada Test Site in 1951. 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 $[0, 160]$ — roughly the envelope of values (in meters) represented in the benchmark data across all sensor types with an additional restriction to above-ground based on independently acquired knowledge. The seismic and acoustic forward models are only appropriate in the *strong shock* regime of 20 to 1,000 meters range scaled by cube-root of explosive yield. In the Bayesian analyses, the prior specification for $\boldsymbol{\theta}_0$ assigned an improper uniform distribution on the real line to w and a proper uniform distribution on the restricted HOB range to h.

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. In application, the chemical-to-nuclear equivalency will be uncertain and may depend on many factors

such as device yield, emplacement condition and sensor type. The corresponding relative measurement uncertainty for the errors-in-variables yield estimation is arbitrarily assumed to be 30% (10% relative standard deviation). 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 from a target distribution that directly depends solely on the new event data. The forward and error model parameters are *imputed* from their posterior distribution, computed using only the benchmark data and sampled via MCMC. In this application, 1,000 imputed values of the forward and error model parameters are obtained. For each imputed sample, 4,000 post-burnin samples of $\boldsymbol{\theta}_0$ are generated via MCMC and decorrelated by thinning out the chain every 200-th sample, resulting in a total of 20,000 (approximate) posterior samples of $\boldsymbol{\theta}_0$. Table 7 presents ML and Bayesian inference for $\boldsymbol{\theta}_0$ based on the benchmark and new event data from each individual sensor type.

The ML and Bayesian results are generally consistent with the true values of w and h 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 7 together with the corresponding yield estimates. For example, the ML acoustic yield estimate in Table 7 has a corresponding uncertainty factor of 4, or roughly a 300% relative error as based on the available data. For HOB, 2σ uncertainty is provided parenthetically, and correlations between w and h (or their MLEs) are also indicated.

The seismic, acoustic, and optical ML estimators of (w, h) have extremely large uncertainty. The forward models used herein simply can't navigate the yield-HOB tradeoff given the degree of uncertainty in the observed signatures: a lower-yield event at ground level gives similar seismic and optical signatures as does a higher-yield event detonated well above the ground, phenomenologically similar to the way that firing a device in a large underground cavity can depress a seismic magnitude relative to a fully-coupled event. The yield-HOB tradeoff is reversed in the vicinity of the ground for the acoustic sensor type. The posterior means of h for each sensor type are larger than the corresponding ML estimates. This is again a consequence of the yield-HOB tradeoff: The profile log-likelihood surfaces are nearly flat in the HOB dimension (see Figure 9 for acoustic), which — when coupled with the flat prior for HOB — carries through to the marginal posterior distribution of h.

Figure 8 shows the fitted versus observed data for the logarithm of overpressure duration. Fitted signatures are obtained by replacing unknown parameters with their ML estimate in the appropriate acoustic forward model and calculating the model results. These plots show substantial agreement between these data, particularly when the observed data are adjusted by both source and path effects ($\boldsymbol{y}_{hij2} - \boldsymbol{Z}_{hi2,j}\hat{\boldsymbol{b}}^{(S)}_{h2,i} - \boldsymbol{Z}_{hj2,i}\hat{\boldsymbol{b}}^{(P)}_{h2,j}$ from Equation (6)).

The benchmark data used to train the forward model parameters combined with the ML estimate of (w, h) for the new event predict the new event data well. However, the forward models do not allow h to be estimated precisely. This is shown in Figure 9, where the profile

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

Method	Sensor Type	Yield [kt] (RE)	HOB [m]	Correlation
ML	Acoustic	0.91 (300%)	7.4 (65)	−0.70
ML	Seismic	0.67 (620%)	0.3 (390)	0.84
ML	Optical	1.42 (890%)	40 (35)	0.74
ML	Crater	1.12 (130%)		
Bayes	Acoustic	0.74 (220%)	65 (90)	−0.12
Bayes	Seismic	0.95 (260%)	75 (95)	0.62
Bayes	Optical	1.78 (990%)	60 (80)	0.14
Bayes	Crater	1.14 (310%)		

log-likelihood in (yield, HOB) is plotted. The nearly vertical contours suggest that for any given yield, the maximized log-likelihood is nearly constant in h.

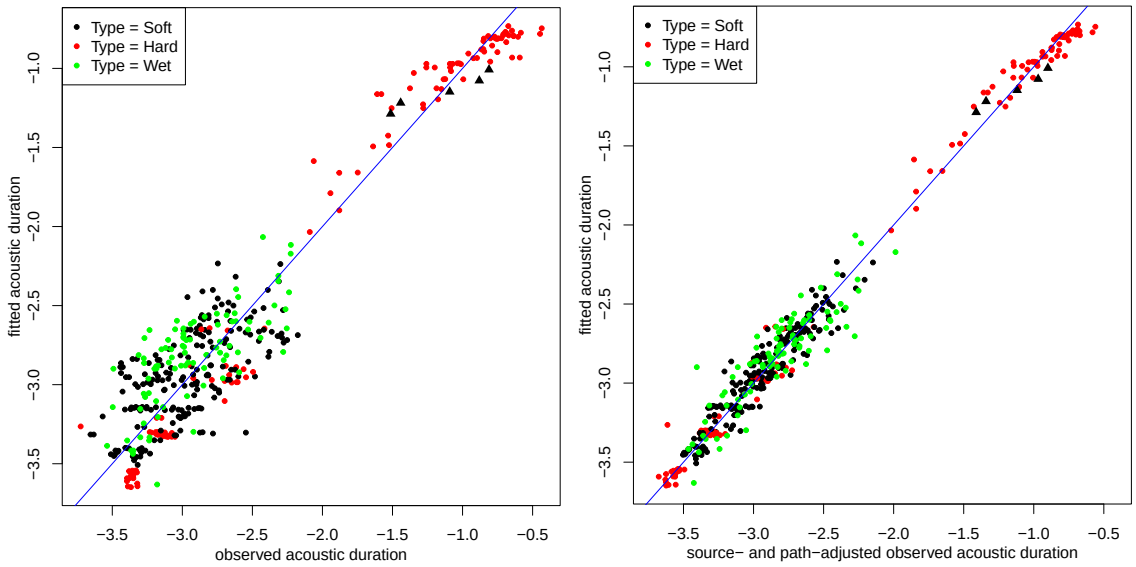

Figure 8: Fitted versus observed logarithm of overpressure duration unadjusted (left) and adjusted for source and path bias (right). Benchmark data (circles) and new event data (triangles) are shown by emplacement (soft, hard, or wet rock).

Table 8 presents maximum-likelihood (ML) and Bayesian inference for $\boldsymbol{\theta}_0$ based on the observed benchmark and new event data integrated across all four sensor types (MultiPEM). Relative uncertainty (roughly 70%) 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, highlighting the potential of integrated analysis to meaningfully reduce uncertainty in the assessment of critical new event parameters.

Figure 10 shows the marginal posterior distributions of yield and HOB for each individual sensor type and across all four sensor types. Although all the marginal distributions corresponding to the seismic, acoustic, and optical sensor types are probabilistically consistent with the reported Sugar values (gold vertical dashed lines), their modes are all biased low for yield ($\exp(w)$) and high for h, as is the surface effects distribution for yield. The highest

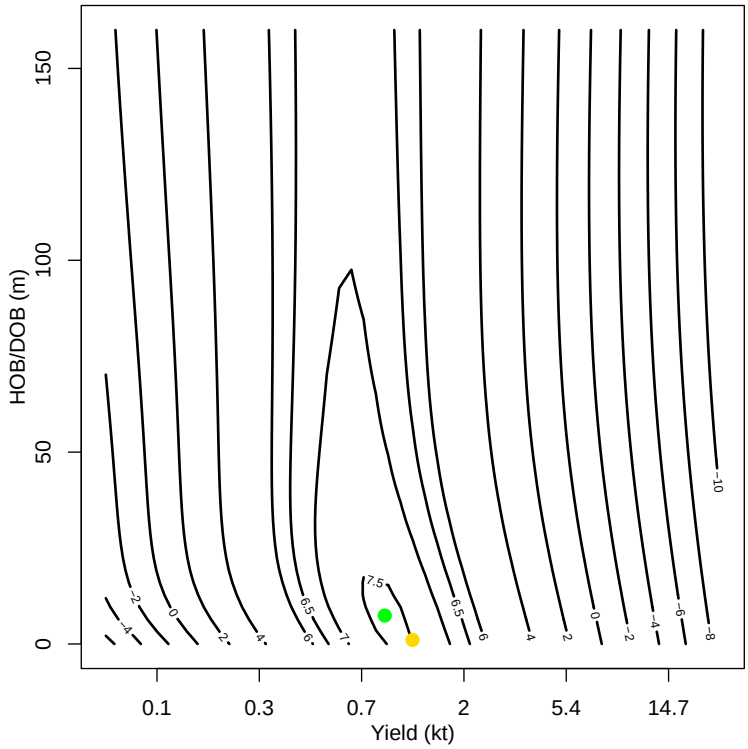

Figure 9: Profile log-likelihood in (yield, HOB) for acoustic signatures. ML estimate (green dot) and true value (gold dot) are plotted for reference.

Table 8: Multiple sensor type ML and Bayesian inference for new event parameters.

Method	Yield [kt] (RE)	HOB [m]	Correlation
ML	0.94 (70%)	1.1 (1.2)	0.23
Bayes	0.92 (100%)	45 (70)	0

posterior density intervals of the acoustic marginal distribution of yield is aligned less closely with the Sugar value. 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)-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 11 shows joint posterior samples of yield and HOB for the seismic, acoustic, and optical sensor types and across all four sensor types. The "sweet spot" described in the previous paragraph is clearly indicated by the MultiPEM yield posterior samples, which are seen to be focused on the region of yield space containing the Sugar values (gold dot). Neither the individual sensor type nor the MultiPEM solutions are particularly informative about the values of h consistent with the observed signatures, as could be expected by examination of the representative profile log-likelihood function in Figure 9.

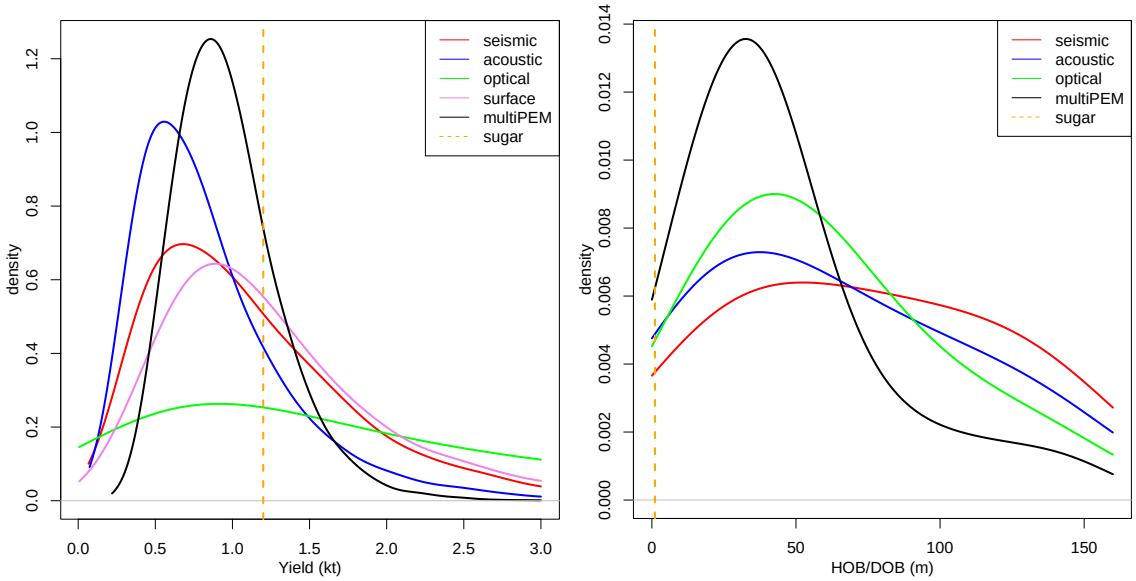

Figure 10: Marginal posterior distributions of yield and HOB for each sensor type and across all four sensor types.

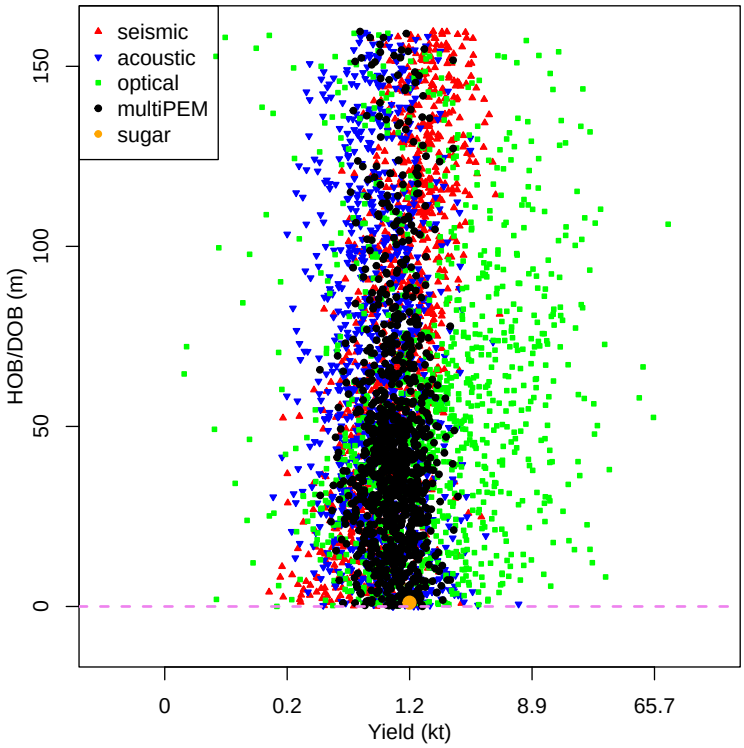

Figure 11: Joint posterior samples of yield and HOB for each sensor type and across all four sensor types.

The errors-in-variables approach of Section 2 was implemented for the log yields of the optical and surface effects benchmark events by assuming their reported yields were correct up to a "three sigma" value of 30%. That is, the value of σ_s in Equation (3) was set to

$0.1 = 0.3/3$. For Bayesian analysis, the unknown yield covariate values pertaining to each unique benchmark 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 9 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 8, with slight reduction in the bias and uncertainty of the yield estimate for both ML and Bayesian inference, and in the HOB estimate for Bayesian inference, when uncertainty in benchmark yields is accounted for.

Table 9: Multiple sensor type ML and Bayesian inference for new event parameters incorporating errors-in-variables for optical and surface effects benchmark event yields.

Method	Yield [kt] (RE)	HOB [m]	Correlation
ML	0.96 (66%)	1.1 (1.2)	0.20
Bayes	0.95 (75%)	40 (45)	0.08

6 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 benchmark data base as is practical in order to reduce turnaround times for data analyses by thorough analysis of benchmark data in advance of the new events.

7 Acknowledgments

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.

Vaněk, J. & Kondorskaya, N. V. (1974), Determination of magnitude station corrections for a continental network of seismic stations, *Studia geoph. et geod.* **18**, 69-78.

Vihola, M. (2012), Robust adaptive Metropolis algorithm with coerced acceptance rate, *Stat. Comput.* **22**, 997-1008.

Whitaker, R. & Symbalisty, E. (2009). Surface to Free Air Optical Scaling Law, Los Alamos National Laboratory Technical Report LA-UR-09-00514.

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 benchmark 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. This framework also applies to unknown calibration parameters at the benchmarking stage (such as chemical-to-nuclear equivalency). The related material to follow is, necessarily, heavily mathematical and notationally intensive.

A.1 Statistical Modeling of Benchmark Event Data

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

$$\boldsymbol{y}_{hgijr} = \boldsymbol{f}_{hr}((\boldsymbol{\beta}_{h0r}, \boldsymbol{\beta}_{ht_{hg}r}), \boldsymbol{\theta}_c, \boldsymbol{v}_{hgij}) + \boldsymbol{Z}_{hgir,j}\boldsymbol{b}_{hr}^{(S)} + \boldsymbol{Z}_{hjr,gi}\boldsymbol{b}_{hr}^{(P)} + \boldsymbol{\varepsilon}_{hgijr}\,, \tag{10}$$

for $h = 1, \ldots, M$, $g = 1, \ldots, M_h$, $i = 1, \ldots, M_{hg}$, $j = 1, \ldots, M_{hgi}$, $r = 1, \ldots, R_h$, and $t_{hg} \in \{1, \ldots, T_h\}$. The unknown model parameters $\boldsymbol{\beta}_{h0r}$ are common to every source observed by sensor type h, while the parameters $\boldsymbol{\beta}_{htr}$ are common to the set of such sources having emplacement condition t. The calibration parameter vector $\boldsymbol{\theta}_c$ of length $\tilde{q}$ collects all quantities of interest to benchmark events that are common across sensor measurement types and present for at least one sensor type (such as chemical-to-nuclear equivalency, possibly of interest in benchmark analysis of chemical explosions used to help infer nuclear event device parameters). If there are no such calibration parameters, this vector is taken to be empty and subsequent calculations involving it are ignored. The vector $\boldsymbol{v}_{hgij}$ collects the known/measured covariates utilized in the model for source i and path j within group g, such as origin time, location, HOB/DOB, and yield (which are independent of source-to-sensor path). The dataset may contain any level of source overlap across sensor types (including none).

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'. There are assumed to be T_h disjoint sets of sources $\mathcal{S}_{ht}$, where $\mathcal{S}_{h1} \cup \cdots \cup \mathcal{S}_{hT_h}$ partitions the $\sum_{g=1}^{M_h} M_{hg}$ 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 $\boldsymbol{\beta}_{ht_{hg}r}$ are eliminated from (10). If $T_h > 1$ and there are no common parameters across all sources, then the parameter $\boldsymbol{\beta}_{h0r}$ is eliminated from (10).

The index set $\mathcal{S}_{ht}$ is partitioned further into source groups $\mathcal{G}_{ht}$. Let $\mathcal{P}_{hg}$ designate the set of unique paths observed by the sources in group $g \in \mathcal{G}_{ht}$. A *source group* g is defined as a collection of sources such that (a) if source $i_1 \in g$, then there exists at least one distinct source $i_2 \in g$ having at least one path in common with i_1, and (b) if path $j \in \mathcal{P}_{hg}$, then $j \notin \mathcal{P}_{h\tilde{g}}$

for $\tilde{g} \in \mathcal{G}_{ht}$ and $\tilde{g} \neq g$. For *crossed* paths as defined in Section 2, this formulation allows cross-correlation due to common paths across sources to be accounted for straightforwardly. For *nested* paths, the source groups align with the individual sources.

Per the main text, random effects are treated as having Gaussian distributions,

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

where $\boldsymbol{\Sigma}_{hgijr}^{(\varepsilon)} = \sigma_{hrr}\mathbf{I}_{n_{hgijr}}$. 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 (10) aggregated across measurement types r for sensor type h.

The $q_{S,hr} \times 1$ source random effect vectors $\boldsymbol{b}_{hr}^{(S)}$ are mutually independent, as are the $q_{P,hr} \times 1$ path random effect vectors $\boldsymbol{b}_{hr}^{(P)}$. Furthermore, the source and path random effect vectors are mutually independent. The covariate matrices $\boldsymbol{Z}_{hgir,j}$ and $\boldsymbol{Z}_{hjr,gi}$ are $n_{hgijr} \times q_{S,hr}$ and $n_{hgijr} \times q_{P,hr}$, respectively. Finally, the elements of each $n_{hgijr} \times 1$ error vector $\boldsymbol{\varepsilon}_{hgijr}$ are mutually independent and also independent of the source and path random effect vectors.

The statistical model (10) assumes the presence of source and path random effects for all sensor measurement types. For each measurement type r, reduced random effect models are also considered, obtained by replacing the term $\boldsymbol{Z}_{hgir,j}\boldsymbol{b}_{hr}^{(S)} + \boldsymbol{Z}_{hjr,gi}\boldsymbol{b}_{hr}^{(P)}$ by (i) only source random effects $\boldsymbol{Z}_{hgir,j}\boldsymbol{b}_{hr}^{(S)}$, (ii) only path random effects $\boldsymbol{Z}_{hjr,gi}\boldsymbol{b}_{hr}^{(P)}$, or (iii) 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 source and path random effect terms in their statistical models. Note that $\mathcal{R}_h^P \subset \mathcal{R}_h^S$ for *nested* paths, and one or both index sets may be empty.

The general formulation to follow assumes *crossed* paths, with caveats for *nested* paths stated as necessary. To facilitate construction of the log-likelihood function in Equation (14), the observed signatures from the benchmark data are organized according to source groups. If $r \in \mathcal{R}_h^P$, each unique path $j \in \mathcal{P}_{hg}$ observed by the sources in group g is assigned a path random effect,

$$\boldsymbol{b}_{hgr}^P = \left[\left(\boldsymbol{b}_{hg1r}^P\right)^\top \quad \left(\boldsymbol{b}_{hg2r}^P\right)^\top \quad \cdots \quad \left(\boldsymbol{b}_{hg|\mathcal{P}_{hg}|r}^P\right)^\top \right]^\top , \tag{11}$$

where the notation "$\top$" denotes vector/matrix transposition, $|\mathcal{P}_{hg}|$ denotes the cardinality of $\mathcal{P}_{hg}$, and the vector elements $\{\boldsymbol{b}_{hgjr}^P\}_j$ of $\boldsymbol{b}_{hgr}^P$ are mutually independently distributed copies of $\boldsymbol{b}_{hr}^{(P)}$.

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

$$\begin{aligned}
\boldsymbol{y}_{hgir} &= \left[\boldsymbol{y}_{hgi1r}^\top \quad \boldsymbol{y}_{hgi2r}^\top \quad \cdots \quad \boldsymbol{y}_{hgiM_{hgi}r}^\top\right]^\top \\
\boldsymbol{f}_{hgir}(\boldsymbol{\beta}_{hr}^{(t_{hg})}, \boldsymbol{\theta}_c) &= \left[\boldsymbol{f}_{hr}^\top(\boldsymbol{\beta}_{hr}^{(t_{hg})}, \boldsymbol{\theta}_c, \boldsymbol{v}_{hgi1}) \quad \boldsymbol{f}_{hr}^\top(\boldsymbol{\beta}_{hr}^{(t_{hg})}, \boldsymbol{\theta}_c, \boldsymbol{v}_{hgi2}) \quad \cdots \quad \boldsymbol{f}_{hr}^\top(\boldsymbol{\beta}_{hr}^{(t_{hg})}, \boldsymbol{\theta}_c, \boldsymbol{v}_{hgiM_{hgi}})\right]^\top
\end{aligned}$$

$$\boldsymbol{\varepsilon}_{hgir} = \begin{bmatrix} \boldsymbol{\varepsilon}_{hgi1r}^\top & \boldsymbol{\varepsilon}_{hgi2r}^\top & \cdots & \boldsymbol{\varepsilon}_{hgiM_{hgi}r}^\top \end{bmatrix}^\top .$$

Let $\boldsymbol{P}_{hgir}$ denote the $(M_{hgi} \cdot q_{P,hr}) \times (|\mathcal{P}_{hg}| \cdot q_{P,hr})$ matrix whose rows extract the random effects vectors $\boldsymbol{b}_{hgjr}^P$ from $\boldsymbol{b}_{hgr}^P$ in Equation (11) corresponding to the $j = 1, \ldots, M_{hgi}$ paths observing source i within group g. Next, define the random effects coefficient matrices

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

$$\boldsymbol{Z}_{hgir}^P = \begin{cases} \text{Diag}\left(\boldsymbol{Z}_{h1r,gi}, \boldsymbol{Z}_{h2r,gi}, \ldots, \boldsymbol{Z}_{hM_{hgi}r,gi}\right) \boldsymbol{P}_{hgir} , & r \in \mathcal{R}_h^P \\ \mathbf{0}_{n_{hgir}, |\mathcal{P}_{hg}| \cdot q_{P,hr}} , & r \notin \mathcal{R}_h^P \end{cases}$$

for $n_{hgir} = \sum_{j=1}^{M_{hgi}} n_{hgijr}$. Concatenating model (10) over source-to-sensor paths gives,

$$\boldsymbol{y}_{hgir} = \boldsymbol{f}_{hgir}(\boldsymbol{\beta}_{hr}^{(t_{hg})}, \boldsymbol{\theta}_c) + \boldsymbol{Z}_{hgir}^S \boldsymbol{b}_{hgir}^S + \boldsymbol{Z}_{hgir}^P \boldsymbol{b}_{hgr}^P + \boldsymbol{\varepsilon}_{hgir} , \tag{12}$$

where $\boldsymbol{b}_{hgir}^S$ is an independent copy of $\boldsymbol{b}_{hr}^{(S)}$. The vector elements $\{\boldsymbol{\varepsilon}_{hgijr}\}_j$ of $\boldsymbol{\varepsilon}_{hgir}$ are mutually independently distributed. If only common model parameters are present in (12), $\boldsymbol{\beta}_{hr}^{(t)} = \boldsymbol{\beta}_{h0r}$, 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 (12) in terms of individual source components $\boldsymbol{y}_{hgi}$,

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

with vectors

$$\begin{aligned}
\boldsymbol{y}_{hgi} &= \begin{bmatrix} \boldsymbol{y}_{hgi1}^\top & \boldsymbol{y}_{hgi2}^\top & \cdots & \boldsymbol{y}_{hgiR_h}^\top \end{bmatrix}^\top \\
\boldsymbol{f}_{hgi}(\boldsymbol{\beta}_h^{(t_{hg})}, \boldsymbol{\theta}_c) &= \begin{bmatrix} \boldsymbol{f}_{hgi1}^\top(\boldsymbol{\beta}_{h1}^{(t_{hg})}, \boldsymbol{\theta}_c) & \boldsymbol{f}_{hgi2}^\top(\boldsymbol{\beta}_{h2}^{(t_{hg})}, \boldsymbol{\theta}_c) & \cdots & \boldsymbol{f}_{hgiR_h}^\top(\boldsymbol{\beta}_{hR_h}^{(t_{hg})}, \boldsymbol{\theta}_c) \end{bmatrix}^\top \\
\boldsymbol{b}_{hgi}^S &= \begin{bmatrix} \left(\boldsymbol{b}_{hgir_{S,1}}^S\right)^\top & \left(\boldsymbol{b}_{hgir_{S,2}}^S\right)^\top & \cdots & \left(\boldsymbol{b}_{hgir_{S,s_{S,h}}}^S\right)^\top \end{bmatrix}^\top \\
\boldsymbol{b}_{hg}^P &= \begin{bmatrix} \left(\boldsymbol{b}_{hgr_{P,1}}^P\right)^\top & \left(\boldsymbol{b}_{hgr_{P,2}}^P\right)^\top & \cdots & \left(\boldsymbol{b}_{hgr_{P,s_{P,h}}}^P\right)^\top \end{bmatrix}^\top \\
\boldsymbol{\varepsilon}_{hgi} &= \begin{bmatrix} \boldsymbol{\varepsilon}_{hgi1}^\top & \boldsymbol{\varepsilon}_{hgi2}^\top & \cdots & \boldsymbol{\varepsilon}_{hgiR_h}^\top \end{bmatrix}^\top .
\end{aligned}$$

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

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

where $n_{hgi} = \sum_{r=1}^{R_h} n_{hgir}$ and $\boldsymbol{\Sigma}_{hgi}$ is composed of $R_h \times R_h$ blocks. The (r_1, r_2) block matrix $[\boldsymbol{\Sigma}_{hgi}]_{r_1,r_2}$ has size $n_{hgir_1} \times n_{hgir_2}$ and is constructed as follows: Let $\mathcal{J}_{r_1,r_2}^{hgi}$ denote the index set containing pairs of indices (k_1, k_2) specifying all joint observations of measurement types r_1 and r_2 for source i within group g by sensor type h. If $r_1 = r_2 = r$ this index set contains every measurement type r paired with itself. Let $\boldsymbol{e}_{hgir,k}$ denote the $n_{hgir} \times 1$ unit vector containing a 1 in the k-th position. Then

$$[\boldsymbol{\Sigma}_{hgi}]_{r_1,r_2} = \begin{cases} \sum_{(k_1,k_2)\in\mathcal{J}_{r_1,r_2}^{hgi}} \sigma_{hr_1r_2}\, \boldsymbol{e}_{hgir_1,k_1} \boldsymbol{e}_{hgir_2,k_2}^\top\,, & r_2 \geq r_1 \\ [\boldsymbol{\Sigma}_{hgi}]_{r_2,r_1}^\top & r_2 < r_1 \end{cases}.$$

Finally, the source-level observations are concatenated over source group to form independent components $\boldsymbol{y}_{hg}$ (assuming conditional independence across sensor types for common sources),

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

with vectors

$$\begin{aligned} \boldsymbol{y}_{hg} &= \begin{bmatrix} \boldsymbol{y}_{hg1}^\top & \boldsymbol{y}_{hg2}^\top & \cdots & \boldsymbol{y}_{hgM_{hg}}^\top \end{bmatrix}^\top \\ \boldsymbol{f}_{hg}(\boldsymbol{\beta}_h^{(t_{hg})}, \boldsymbol{\theta}_c) &= \begin{bmatrix} \boldsymbol{f}_{hg1}^\top(\boldsymbol{\beta}_h^{(t_{hg})}, \boldsymbol{\theta}_c) & \boldsymbol{f}_{hg2}^\top(\boldsymbol{\beta}_h^{(t_{hg})}, \boldsymbol{\theta}_c) & \cdots & \boldsymbol{f}_{hgM_{hg}}^\top(\boldsymbol{\beta}_h^{(t_{hg})}, \boldsymbol{\theta}_c) \end{bmatrix}^\top \\ \boldsymbol{b}_{hg}^S &= \begin{bmatrix} \left(\boldsymbol{b}_{hg1}^S\right)^\top & \left(\boldsymbol{b}_{hg2}^S\right)^\top & \cdots & \left(\boldsymbol{b}_{hgM_{hg}}^S\right)^\top \end{bmatrix}^\top \\ \boldsymbol{\varepsilon}_{hg} &= \begin{bmatrix} \boldsymbol{\varepsilon}_{hg1}^\top & \boldsymbol{\varepsilon}_{hg2}^\top & \cdots & \boldsymbol{\varepsilon}_{hgM_{hg}}^\top \end{bmatrix}^\top, \end{aligned}$$

where the vector elements $\{\boldsymbol{b}_{hgi}^S\}_i$ of $\boldsymbol{b}_{hg}^S$ are mutually independently distributed. The random effects coefficient matrices are given by

$$\begin{aligned} \boldsymbol{Z}_{hg}^S &= \mathrm{Diag}\left(\boldsymbol{Z}_{hg1}^S, \boldsymbol{Z}_{hg2}^S, \ldots, \boldsymbol{Z}_{hgM_{hg}}^S\right) \\ \boldsymbol{Z}_{hg}^P &= \begin{bmatrix} \left(\boldsymbol{Z}_{hg1}^P\right)^\top & \left(\boldsymbol{Z}_{hg2}^P\right)^\top & \cdots & \left(\boldsymbol{Z}_{hgM_{hg}}^P\right)^\top \end{bmatrix}^\top. \end{aligned}$$

This source-level concatenated model is

$$\boldsymbol{y}_{hg} = \boldsymbol{f}_{hg}(\boldsymbol{\beta}_h^{(t_{hg})}, \boldsymbol{\theta}_c) + \boldsymbol{\epsilon}_{hg}\,, \tag{13}$$

where

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

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

$$\boldsymbol{\Xi}_{hg}^{(S)} = \mathrm{Diag}\left(\boldsymbol{Z}_{hg1}^S \boldsymbol{\Sigma}_h^S \left(\boldsymbol{Z}_{hg1}^S\right)^\top, \boldsymbol{Z}_{hg2}^S \boldsymbol{\Sigma}_h^S \left(\boldsymbol{Z}_{hg2}^S\right)^\top, \ldots, \boldsymbol{Z}_{hgM_{hg}}^S \boldsymbol{\Sigma}_h^S \left(\boldsymbol{Z}_{hgM_{hg}}^S\right)^\top\right)$$

for

$$\boldsymbol{\Sigma}_h^S = \mathrm{Diag}\left(\boldsymbol{\Sigma}_{hr_{S,1}}^{(S)}, \boldsymbol{\Sigma}_{hr_{S,2}}^{(S)}, \ldots, \boldsymbol{\Sigma}_{hr_{S,s_{S,h}}}^{(S)}\right).$$

The covariance matrix of $\boldsymbol{Z}_{hg}^P \boldsymbol{b}_{hg}^P$ is given by

$$\boldsymbol{\Xi}_{hg}^{(P)} = \boldsymbol{Z}_{hg}^P \operatorname{Diag}\left(\left(\boldsymbol{I}_{|\mathcal{P}_{hg}|} \otimes \boldsymbol{\Sigma}_{hr_{P,1}}^{(P)}\right), \ldots, \left(\boldsymbol{I}_{|\mathcal{P}_{hg}|} \otimes \boldsymbol{\Sigma}_{hr_{P,s_{P,h}}}^{(P)}\right)\right) \left(\boldsymbol{Z}_{hg}^P\right)^\top .$$

Finally, the covariance matrix of $\boldsymbol{\varepsilon}_{hg}$ is given by

$$\boldsymbol{\Sigma}_{hg} = \operatorname{Diag}\left(\boldsymbol{\Sigma}_{hg1}, \boldsymbol{\Sigma}_{hg2}, \ldots, \boldsymbol{\Sigma}_{hgM_{hg}}\right) .$$

Any measurement type r that is not observed by sensor type h for source i within group g is excluded from the vectors and matrices above. Any measurement type r that is observed by sensor type h for source i within group g, but on a reduced set of $1 < M_{hgir} < M_{hgi}$ paths, has M_{hgi} replaced by M_{hgir} in the above and subsequent development.

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

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

for $n_{hg} = \sum_{i=1}^{M_{hg}} n_{hgi}$, and is therefore distributed as

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

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

$$\begin{aligned}
&\mathcal{L}_c(\boldsymbol{\beta}, \boldsymbol{\theta}_c, \boldsymbol{\varsigma} | \boldsymbol{y}) = \\
&\quad -\frac{1}{2} \sum_{h=1}^{M} \sum_{g=1}^{M_h} \log\det\left(\boldsymbol{\Omega}_{hg}\right) \\
&\quad -\frac{1}{2} \sum_{h=1}^{M} \sum_{t=1}^{T_h} \sum_{g \in \mathcal{G}_{ht}} \left(\boldsymbol{y}_{hg} - \boldsymbol{f}_{hg}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c)\right)^\top \boldsymbol{\Omega}_{hg}^{-1} \left(\boldsymbol{y}_{hg} - \boldsymbol{f}_{hg}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c)\right) . \qquad (14)
\end{aligned}$$

Let

$$\boldsymbol{r}'_{hgt} = \boldsymbol{y}_{hg} - \boldsymbol{f}_{hg}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c) .$$

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

$$\frac{\partial \mathcal{L}_c}{\partial \boldsymbol{\beta}_{h0}} = \sum_{t=1}^{T_h} \sum_{g \in \mathcal{G}_{ht}} \boldsymbol{J}_{\boldsymbol{\beta}_{h0}, hg}^\top (\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c) \boldsymbol{\Omega}_{hg}^{-1} \boldsymbol{r}'_{hgt}$$

$$\frac{\partial \mathcal{L}_c}{\partial \boldsymbol{\beta}_{ht}} = \sum_{g \in \mathcal{G}_{ht}} \boldsymbol{J}^{\top}_{\boldsymbol{\beta}_{ht},hg}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c)\boldsymbol{\Omega}^{-1}_{hg}\boldsymbol{r}'_{hgt}\,,\; t = 1, \ldots, T_h$$

$$\frac{\partial \mathcal{L}_c}{\partial \boldsymbol{\theta}_c} = \sum_{h=1}^{M}\sum_{t=1}^{T_h}\sum_{g \in \mathcal{G}_{ht}} \boldsymbol{J}^{\top}_{\boldsymbol{\theta}_c,hg}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c)\boldsymbol{\Omega}^{-1}_{hg}\boldsymbol{r}'_{hgt}$$

$$\frac{\partial \mathcal{L}_c}{\partial \varsigma_{hk}} = -\frac{1}{2}\sum_{t=1}^{T_h}\sum_{g \in \mathcal{G}_{ht}} \mathrm{tr}\left\{\left[\boldsymbol{\Omega}^{-1}_{hg} - \boldsymbol{\Omega}^{-1}_{hg}\boldsymbol{r}'_{hgt}\left(\boldsymbol{r}'_{hgt}\right)^{\top}\boldsymbol{\Omega}^{-1}_{hg}\right]\frac{\partial \boldsymbol{\Omega}_{hg}}{\partial \varsigma_{hk}}\right\}$$

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

$$\boldsymbol{J}_{hg}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c) = \left[\boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c) \quad \boldsymbol{J}_{\boldsymbol{\beta}_{ht},hg}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c) \quad \boldsymbol{J}_{\boldsymbol{\theta}_c,hg}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c)\right] \tag{15}$$

for

$$\boldsymbol{J}^{\top}_{\boldsymbol{\beta}_{hu},hg}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c) = \left[\nabla_{\boldsymbol{\beta}_{hu}} f_{hg,1}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c) \quad \nabla_{\boldsymbol{\beta}_{hu}} f_{hg,2}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c) \quad \cdots \quad \nabla_{\boldsymbol{\beta}_{hu}} f_{hg,n_{hg}}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c)\right]$$

$$\boldsymbol{J}^{\top}_{\boldsymbol{\theta}_c,hg}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c) = \left[\nabla_{\boldsymbol{\theta}_c} f_{hg,1}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c) \quad \nabla_{\boldsymbol{\theta}_c} f_{hg,2}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c) \quad \cdots \quad \nabla_{\boldsymbol{\theta}_c} f_{hg,n_{hg}}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c)\right].$$

Here $\nabla_{\boldsymbol{\eta}} f_{hg,k}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c)$ is the gradient vector of the k-th element of $\boldsymbol{f}_{hg}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c)$ with respect to $\boldsymbol{\eta}$, $k = 1, \ldots, n_{hg}$, for $\boldsymbol{\eta}$ one of $\boldsymbol{\beta}_{hu}$ for $u \in \{0, t\}$ or $\boldsymbol{\theta}_c$. 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 $\boldsymbol{\beta}^{(t)}_{hr} = \boldsymbol{\beta}_{h0r}$ or $\boldsymbol{\beta}^{(t)}_{hr} = \boldsymbol{\beta}_{htr}$ pertains.

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

$$\boldsymbol{J}^{\top}_{\boldsymbol{\beta}_{hur},hgir}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c) = \left[\boldsymbol{J}^{\top}_{\boldsymbol{\beta}_{hur},hgi1r}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c) \quad \boldsymbol{J}^{\top}_{\boldsymbol{\beta}_{hur},hgi2r}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c) \quad \cdots \quad \boldsymbol{J}^{\top}_{\boldsymbol{\beta}_{hur},hgiM_{hgi}r}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c)\right]$$

$$\boldsymbol{J}^{\top}_{\boldsymbol{\theta}_c,hgir}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c) = \left[\boldsymbol{J}^{\top}_{\boldsymbol{\theta}_c,hgi1r}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c) \quad \boldsymbol{J}^{\top}_{\boldsymbol{\theta}_c,hgi2r}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c) \quad \cdots \quad \boldsymbol{J}^{\top}_{\boldsymbol{\theta}_c,hgiM_{hgi}r}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c)\right],$$

where $\boldsymbol{J}_{\boldsymbol{\eta},hgijr}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c)$ is the Jacobian matrix of $\boldsymbol{f}_{hr}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c, \boldsymbol{v}_{hgij})$ with respect to $\boldsymbol{\eta}$ for $\boldsymbol{\eta}$ one of $\boldsymbol{\beta}_{hur}$ with $u \in \{0, t\}$ or $\boldsymbol{\theta}_c$. The vector $\boldsymbol{f}_{hr}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c, \boldsymbol{v}_{hgij})$ contains n_{hgijr} copies of the scalar function $f_{hr}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c, \boldsymbol{v}_{hgij})$, so that each column of the matrix $\boldsymbol{J}^{\top}_{\boldsymbol{\eta},hgijr}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c)$ contains $\nabla_{\boldsymbol{\eta}} f_{hr}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c, \boldsymbol{v}_{hgij})$ with respect to $\boldsymbol{\eta}$, for $\boldsymbol{\eta}$ one of $\boldsymbol{\beta}_{hur}$ with $u \in \{0, t\}$ or $\boldsymbol{\theta}_c$, i.e.

$$\boldsymbol{J}^{\top}_{\boldsymbol{\eta},hgijr}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c) = \nabla_{\boldsymbol{\eta}} f_{hr}(\boldsymbol{\beta}^{(t)}_{hr}, \boldsymbol{\theta}_c, \boldsymbol{v}_{hgij})\mathbf{1}^{\top}_{n_{hgijr}}$$

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

Collecting Jacobian matrices across sensor measurement types,

$$\boldsymbol{J}_{\boldsymbol{\beta}_{hu},hgi}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c) = \mathrm{Diag}\left(\boldsymbol{J}_{\boldsymbol{\beta}_{hu1},hgi1}(\boldsymbol{\beta}^{(t)}_{h1}, \boldsymbol{\theta}_c), \boldsymbol{J}_{\boldsymbol{\beta}_{hu2},hgi2}(\boldsymbol{\beta}^{(t)}_{h2}, \boldsymbol{\theta}_c), \ldots, \boldsymbol{J}_{\boldsymbol{\beta}_{huR_h},hgiR_h}(\boldsymbol{\beta}^{(t)}_{hR_h}, \boldsymbol{\theta}_c)\right)$$

$$\boldsymbol{J}^{\top}_{\boldsymbol{\theta}_c,hgi}(\boldsymbol{\beta}^{(t)}_h, \boldsymbol{\theta}_c) = \left[\boldsymbol{J}^{\top}_{\boldsymbol{\theta}_c,hgi1}(\boldsymbol{\beta}^{(t)}_{h1}, \boldsymbol{\theta}_c) \quad \boldsymbol{J}^{\top}_{\boldsymbol{\theta}_c,hgi2}(\boldsymbol{\beta}^{(t)}_{h2}, \boldsymbol{\theta}_c) \quad \cdots \quad \boldsymbol{J}^{\top}_{\boldsymbol{\theta}_c,hgiR_h}(\boldsymbol{\beta}^{(t)}_{hR_h}, \boldsymbol{\theta}_c)\right],$$

results in expressing the elements of (15) as

$$\boldsymbol{J}_{\boldsymbol{\eta},hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c) = \left[\boldsymbol{J}_{\boldsymbol{\eta},hg1}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c) \quad \boldsymbol{J}_{\boldsymbol{\eta},hg2}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c) \quad \cdots \quad \boldsymbol{J}_{\boldsymbol{\eta},hgM_{hg}}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)\right]$$

for $\boldsymbol{\eta}$ one of $\boldsymbol{\beta}_{hu}$ with $u \in \{0,t\}$ or $\boldsymbol{\theta}_c$. Let

$$\boldsymbol{K}_{h0}(\boldsymbol{\beta}_h,\boldsymbol{\theta}_c) = \sum_{t=1}^{T_h}\sum_{g\in\mathcal{G}_{ht}}\left[\frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)}{\partial\beta_{h01}}\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{r}'_{hgt} \;\cdots\; \frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)}{\partial\beta_{h0p_{h0}}}\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{r}'_{hgt}\right]$$

$$\boldsymbol{K}_{h0t}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c) = \sum_{g\in\mathcal{G}_{ht}}\left[\frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)}{\partial\beta_{ht1}}\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{r}'_{hgt} \;\cdots\; \frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)}{\partial\beta_{htp_{ht}}}\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{r}'_{hgt}\right],$$
$$t = 1,\ldots,T_h\,,$$

$$\boldsymbol{K}_{ht}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c) = \sum_{g\in\mathcal{G}_{ht}}\left[\frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{ht},hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)}{\partial\beta_{ht1}}\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{r}'_{hgt} \;\cdots\; \frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{ht},hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)}{\partial\beta_{htp_{ht}}}\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{r}'_{hgt}\right],$$
$$t = 1,\ldots,T_h\,,$$

$$\boldsymbol{K}_{(\boldsymbol{\theta}_c,\boldsymbol{\theta}_c)}(\boldsymbol{\beta},\boldsymbol{\theta}_c) = \sum_{h=1}^{M}\sum_{t=1}^{T_h}\sum_{g\in\mathcal{G}_{ht}}\left[\frac{\partial \boldsymbol{J}_{\boldsymbol{\theta}_c,hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)}{\partial\theta_{c1}}\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{r}'_{hgt} \;\cdots\; \frac{\partial \boldsymbol{J}_{\boldsymbol{\theta}_c,hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)}{\partial\theta_{c\tilde{q}}}\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{r}'_{hgt}\right]$$

$$\boldsymbol{K}_{(\boldsymbol{\beta}_{h0},\boldsymbol{\theta}_c)}(\boldsymbol{\beta}_h,\boldsymbol{\theta}_c) = \sum_{t=1}^{T_h}\sum_{g\in\mathcal{G}_{ht}}\left[\frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)}{\partial\theta_{c1}}\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{r}'_{hgt} \;\cdots\; \frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)}{\partial\theta_{c\tilde{q}}}\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{r}'_{hgt}\right]$$

$$\boldsymbol{K}_{(\boldsymbol{\beta}_{ht},\boldsymbol{\theta}_c)}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c) = \sum_{g\in\mathcal{G}_{ht}}\left[\frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{ht},hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)}{\partial\theta_{c1}}\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{r}'_{hgt} \;\cdots\; \frac{\partial \boldsymbol{J}_{\boldsymbol{\beta}_{ht},hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)}{\partial\theta_{c\tilde{q}}}\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{r}'_{hgt}\right],$$
$$t = 1,\ldots,T_h\,,$$

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

The second partial derivatives of the log-likelihood function are

$$\frac{\partial^2\mathcal{L}_c}{\partial\boldsymbol{\beta}_{h0}\partial\boldsymbol{\beta}_{h0}^{\top}} = -\sum_{t=1}^{T_h}\sum_{g\in\mathcal{G}_{ht}}\boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c) + \boldsymbol{K}_{h0}(\boldsymbol{\beta}_h,\boldsymbol{\theta}_c)$$

$$\frac{\partial^2\mathcal{L}_c}{\partial\boldsymbol{\beta}_{h0}\partial\boldsymbol{\beta}_{ht}^{\top}} = -\sum_{g\in\mathcal{G}_{ht}}\boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{J}_{\boldsymbol{\beta}_{ht},hg}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c) + \boldsymbol{K}_{h0t}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)\,,\ t = 1,\ldots,T_h$$

$$\frac{\partial^2\mathcal{L}_c}{\partial\boldsymbol{\beta}_{ht}\partial\boldsymbol{\beta}_{ht}^{\top}} = -\sum_{g\in\mathcal{G}_{ht}}\boldsymbol{J}_{\boldsymbol{\beta}_{ht},hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{J}_{\boldsymbol{\beta}_{ht},hg}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c) + \boldsymbol{K}_{ht}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)\,,\ t = 1,\ldots,T_h$$

$$\frac{\partial^2\mathcal{L}_c}{\partial\boldsymbol{\theta}_c\partial\boldsymbol{\theta}_c^{\top}} = -\sum_{h=1}^{M}\sum_{t=1}^{T_h}\sum_{g\in\mathcal{G}_{ht}}\boldsymbol{J}_{\boldsymbol{\theta}_c,hg}^{\top}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c)\boldsymbol{\Omega}_{hg}^{-1}\boldsymbol{J}_{\boldsymbol{\theta}_c,hg}(\boldsymbol{\beta}_h^{(t)},\boldsymbol{\theta}_c) + \boldsymbol{K}_{(\boldsymbol{\theta}_c,\boldsymbol{\theta}_c)}(\boldsymbol{\beta},\boldsymbol{\theta}_c)$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial \boldsymbol{\beta}_{h0} \partial \boldsymbol{\theta}_c^\top} = -\sum_{t=1}^{T_h} \sum_{g \in \mathcal{G}_{ht}} \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}^\top(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c) \boldsymbol{\Omega}_{hg}^{-1} \boldsymbol{J}_{\boldsymbol{\theta}_c,hg}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c) + \boldsymbol{K}_{(\boldsymbol{\beta}_{h0},\boldsymbol{\theta}_c)}(\boldsymbol{\beta}_h, \boldsymbol{\theta}_c)$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial \boldsymbol{\beta}_{ht} \partial \boldsymbol{\theta}_c^\top} = -\sum_{g \in \mathcal{G}_{ht}} \boldsymbol{J}_{\boldsymbol{\beta}_{ht},hg}^\top(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c) \boldsymbol{\Omega}_{hg}^{-1} \boldsymbol{J}_{\boldsymbol{\theta}_c,hg}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c) + \boldsymbol{K}_{(\boldsymbol{\beta}_{ht},\boldsymbol{\theta}_c)}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c)\,,\ t = 1, \ldots, T_h$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial \boldsymbol{\beta}_{h0} \partial \varsigma_{hk}} = -\sum_{t=1}^{T_h} \sum_{g \in \mathcal{G}_{ht}} \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}^\top(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c) \boldsymbol{\Omega}_{hg}^{-1} \frac{\partial \boldsymbol{\Omega}_{hg}}{\partial \varsigma_{hk}} \boldsymbol{\Omega}_{hg}^{-1} \boldsymbol{r}'_{hgt}$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial \boldsymbol{\beta}_{ht} \partial \varsigma_{hk}} = -\sum_{g \in \mathcal{G}_{ht}} \boldsymbol{J}_{\boldsymbol{\beta}_{ht},hg}^\top(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c) \boldsymbol{\Omega}_{hg}^{-1} \frac{\partial \boldsymbol{\Omega}_{hg}}{\partial \varsigma_{hk}} \boldsymbol{\Omega}_{hg}^{-1} \boldsymbol{r}'_{hgt}\,,\ t = 1, \ldots, T_h$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial \boldsymbol{\theta}_c \partial \varsigma_{hk}} = -\sum_{t=1}^{T_h} \sum_{g \in \mathcal{G}_{ht}} \boldsymbol{J}_{\boldsymbol{\theta}_c,hg}^\top(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c) \boldsymbol{\Omega}_{hg}^{-1} \frac{\partial \boldsymbol{\Omega}_{hg}}{\partial \varsigma_{hk}} \boldsymbol{\Omega}_{hg}^{-1} \boldsymbol{r}'_{hgt}$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial \varsigma_{hk} \partial \varsigma_{hl}} = -\frac{1}{2} \sum_{t=1}^{T_h} \sum_{g \in \mathcal{G}_{ht}} \operatorname{tr}\left\{ \boldsymbol{\Omega}_{hg}^{-1} \frac{\partial \boldsymbol{\Omega}_{hg}}{\partial \varsigma_{hk}} \boldsymbol{S}_{hgt} \frac{\partial \boldsymbol{\Omega}_{hg}}{\partial \varsigma_{hl}} \right\} -$$

$$\frac{1}{2} \sum_{t=1}^{T_h} \sum_{g \in \mathcal{G}_{ht}} \operatorname{tr}\left\{ \left[\boldsymbol{\Omega}_{hg}^{-1} - \boldsymbol{\Omega}_{hg}^{-1} \boldsymbol{r}'_{hgt} \left(\boldsymbol{r}'_{hgt} \right)^\top \boldsymbol{\Omega}_{hg}^{-1} \right] \frac{\partial^2 \boldsymbol{\Omega}_{hg}}{\partial \varsigma_{hk} \partial \varsigma_{hl}} \right\}.$$

Noting that

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

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

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

$$\mathcal{I}^c_{\boldsymbol{\beta}_{h0},\boldsymbol{\beta}_{h0}} = \sum_{t=1}^{T_h} \sum_{g \in \mathcal{G}_{ht}} \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}^\top(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c) \boldsymbol{\Omega}_{hg}^{-1} \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c)$$

$$\mathcal{I}^c_{\boldsymbol{\beta}_{h0},\boldsymbol{\beta}_{ht}} = \sum_{g \in \mathcal{G}_{ht}} \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}^\top(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c) \boldsymbol{\Omega}_{hg}^{-1} \boldsymbol{J}_{\boldsymbol{\beta}_{ht},hg}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c)\,,\ t = 1, \ldots, T_h$$

$$\mathcal{I}^c_{\boldsymbol{\beta}_{ht},\boldsymbol{\beta}_{ht}} = \sum_{g \in \mathcal{G}_{ht}} \boldsymbol{J}_{\boldsymbol{\beta}_{ht},hg}^\top(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c) \boldsymbol{\Omega}_{hg}^{-1} \boldsymbol{J}_{\boldsymbol{\beta}_{ht},hg}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c)\,,\ t = 1, \ldots, T_h$$

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

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

$$\mathcal{I}^c_{\boldsymbol{\beta}_{h0},\boldsymbol{\theta}_c} = \sum_{t=1}^{T_h} \sum_{g \in \mathcal{G}_{ht}} \boldsymbol{J}_{\boldsymbol{\beta}_{h0},hg}^\top(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c) \boldsymbol{\Omega}_{hg}^{-1} \boldsymbol{J}_{\boldsymbol{\theta}_c,hg}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_c)$$

$$\mathcal{I}^c_{\boldsymbol{\beta}_{ht},\boldsymbol{\theta}_c} = \sum_{g\in\mathcal{G}_{ht}} \boldsymbol{J}^\top_{\boldsymbol{\beta}_{ht},hg}(\boldsymbol{\beta}^{(t)}_h,\boldsymbol{\theta}_c)\boldsymbol{\Omega}^{-1}_{hg}\boldsymbol{J}_{\boldsymbol{\theta}_c,hg}(\boldsymbol{\beta}^{(t)}_h,\boldsymbol{\theta}_c)\,,\; t=1,\ldots,T_h\,.$$

Remaining elements of the information matrix for sensor type h are given by

$$\begin{aligned}
\mathcal{I}^c_{\boldsymbol{\beta}_h,\boldsymbol{\varsigma}_h} &= \mathbf{0}_{p_h,v_h}\\
\mathcal{I}^c_{\boldsymbol{\theta}_c,\boldsymbol{\varsigma}_h} &= \mathbf{0}_{\tilde{q},v_h}\\
\mathcal{I}^c_{\varsigma_{hk},\varsigma_{hl}} &= \frac{1}{2}\sum_{g=1}^{M_h}\mathrm{tr}\left\{\boldsymbol{\Omega}^{-1}_{hg}\frac{\partial\boldsymbol{\Omega}_{hg}}{\partial\varsigma_{hk}}\boldsymbol{\Omega}^{-1}_{hg}\frac{\partial\boldsymbol{\Omega}_{hg}}{\partial\varsigma_{hl}}\right\}\,.
\end{aligned}$$

Define the following quantities:

$$\begin{aligned}
\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\beta}} &= \mathrm{Diag}\left(\mathcal{I}^c_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_1},\mathcal{I}^c_{\boldsymbol{\beta}_2,\boldsymbol{\beta}_2},\ldots,\mathcal{I}^c_{\boldsymbol{\beta}_M,\boldsymbol{\beta}_M}\right)\\
\mathcal{I}^c_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_c} &= \sum_{h=1}^{M}\sum_{t=1}^{T_h}\sum_{g\in\mathcal{G}_{ht}}\boldsymbol{J}^\top_{\boldsymbol{\theta}_c,hg}(\boldsymbol{\beta}^{(t)}_h,\boldsymbol{\theta}_c)\boldsymbol{\Omega}^{-1}_{hg}\boldsymbol{J}_{\boldsymbol{\theta}_c,hg}(\boldsymbol{\beta}^{(t)}_h,\boldsymbol{\theta}_c)\\
\left(\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\theta}_c}\right)^\top &= \left[\left(\mathcal{I}^c_{\boldsymbol{\beta}_1,\boldsymbol{\theta}_c}\right)^\top\quad\left(\mathcal{I}^c_{\boldsymbol{\beta}_2,\boldsymbol{\theta}_c}\right)^\top\quad\cdots\quad\left(\mathcal{I}^c_{\boldsymbol{\beta}_M,\boldsymbol{\theta}_c}\right)^\top\right]\\
\mathcal{I}^c_{\boldsymbol{\varsigma},\boldsymbol{\varsigma}} &= \mathrm{Diag}\left(\mathcal{I}^c_{\boldsymbol{\varsigma}_1,\boldsymbol{\varsigma}_1},\mathcal{I}^c_{\boldsymbol{\varsigma}_2,\boldsymbol{\varsigma}_2},\ldots,\mathcal{I}^c_{\boldsymbol{\varsigma}_M,\boldsymbol{\varsigma}_M}\right)\,.
\end{aligned}$$

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

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

Let

$$\begin{aligned}
\mathcal{I}^c_{\boldsymbol{\beta}|\boldsymbol{\theta}_c} &= \mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\beta}} - \mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\theta}_c}\left(\mathcal{I}^c_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_c}\right)^{-1}\left(\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\theta}_c}\right)^\top\\
\mathcal{I}^c_{\boldsymbol{\theta}_c|\boldsymbol{\beta}} &= \mathcal{I}^c_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_c} - \left(\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\theta}_c}\right)^\top\left(\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1}\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\theta}_c}\\
&= \mathcal{I}^c_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_c} - \sum_{h=1}^{M}\left(\mathcal{I}^c_{\boldsymbol{\beta}_h,\boldsymbol{\theta}_c}\right)^\top\left(\mathcal{I}^c_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right)^{-1}\mathcal{I}^c_{\boldsymbol{\beta}_h,\boldsymbol{\theta}_c}\,.
\end{aligned}$$

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

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

The corresponding distribution of the maximum likelihood estimators is

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

A.2 Statistical Modeling of New Event Data

Now consider a *new* event. Assuming conditional independence of the new event and benchmark signatures, the new event belongs to its own source group labeled '0', consisting of a single source labeled '1' ($M_{h0} = 1$). For sensor type h, (13) becomes

$$\boldsymbol{y}_{h0} = \boldsymbol{f}_{h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0) + \boldsymbol{\epsilon}_{h0} \,, \tag{16}$$

where

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

for

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

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 benchmark data discussed earlier (i.e., contained within the covariate vectors $\boldsymbol{v}_{hgij}$), 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 (13). 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, (16) becomes

$$\boldsymbol{y}_0 = \boldsymbol{f}_0(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0) + \boldsymbol{\epsilon}_0 \tag{17}$$

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 (17) is given by

$$\mathcal{L}_0(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \boldsymbol{\varsigma} | \boldsymbol{y}_0) =$$

$$
\begin{aligned}
&- \frac{1}{2}\sum_{h=1}^{M} \log\det\left(\boldsymbol{\Omega}_{h0}\right) \\
&- \frac{1}{2}\sum_{h=1}^{M} \left(\boldsymbol{y}_{h0} - \boldsymbol{f}_{h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0)\right)^\top \boldsymbol{\Omega}_{h0}^{-1} \left(\boldsymbol{y}_{h0} - \boldsymbol{f}_{h0}(\boldsymbol{\beta}_h^{(t_{h0})}, \boldsymbol{\theta}_0)\right) . \qquad (18)
\end{aligned}
$$

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

$$
\boldsymbol{J}_{h0}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_0) = \left[\boldsymbol{J}_{\boldsymbol{\beta}_{h0},h0}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_0) \quad \boldsymbol{J}_{\boldsymbol{\beta}_{ht},h0}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_0) \quad \boldsymbol{J}_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t)}, \boldsymbol{\theta}_0)\right] \qquad (19)
$$

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 $\boldsymbol{\beta}_{hu}$ with $u \in \{0, t\}$ or $\boldsymbol{\theta}_0$.

For the combination $(h, 0, 1, r)$, the Jacobian matrices of $\boldsymbol{f}_{h01r}(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0)$ with respect to $\boldsymbol{\beta}_{hur}$ for $u \in \{0, t\}$ and $\boldsymbol{\theta}_0$ are given by

$$
\begin{aligned}
\boldsymbol{J}_{\boldsymbol{\beta}_{hur},h01r}^\top(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0) &= \left[\boldsymbol{J}_{\boldsymbol{\beta}_{hur},h011r}^\top(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0) \quad \boldsymbol{J}_{\boldsymbol{\beta}_{hur},h012r}^\top(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0) \quad \cdots \quad \boldsymbol{J}_{\boldsymbol{\beta}_{hur},h01M_{h01}r}^\top(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0)\right] \\
\boldsymbol{J}_{\boldsymbol{\theta}_0,h01r}^\top(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0) &= \left[\boldsymbol{J}_{\boldsymbol{\theta}_0,h011r}^\top(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0) \quad \boldsymbol{J}_{\boldsymbol{\theta}_0,h012r}^\top(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0) \quad \cdots \quad \boldsymbol{J}_{\boldsymbol{\theta}_0,h01M_{h01}r}^\top(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0)\right] ,
\end{aligned}
$$

where $\boldsymbol{J}_{\boldsymbol{\eta},h01jr}(\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0)$ is the Jacobian matrix of $\boldsymbol{f}_{hr}((\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0), \boldsymbol{v}'_{h01j})$ 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 $\boldsymbol{v}'_{h01j}$ is the subset of $\boldsymbol{v}_{h01j}$ 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 benchmark data but may be unknown for a new event) — in other words, for the new event $\boldsymbol{v}_{h01j} = (\boldsymbol{\theta}_0, \boldsymbol{v}'_{h01j})$. The vector $\boldsymbol{f}_{hr}((\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0), \boldsymbol{v}'_{h01j})$ contains n_{h01jr} copies of the scalar function $f_{hr}((\boldsymbol{\beta}_{hr}^{(t)}, \boldsymbol{\theta}_0), \boldsymbol{v}'_{h01j})$,

so that each column of the matrix $\boldsymbol{J}^{\top}_{\boldsymbol{\eta},h01jr}(\boldsymbol{\beta}^{(t)}_{hr},\boldsymbol{\theta}_0)$ contains $\nabla_{\boldsymbol{\eta}} f_{hr}((\boldsymbol{\beta}^{(t)}_{hr},\boldsymbol{\theta}_0),\boldsymbol{v}'_{h01j})$ with respect to $\boldsymbol{\eta}$, for $\boldsymbol{\eta}$ one of $\boldsymbol{\beta}_{hur}$ with $u \in \{0,t\}$ or $\boldsymbol{\theta}_0$, i.e.

$$\boldsymbol{J}^{\top}_{\boldsymbol{\eta},h01jr}(\boldsymbol{\beta}^{(t)}_{hr},\boldsymbol{\theta}_0) = \nabla_{\boldsymbol{\eta}} f_{hr}((\boldsymbol{\beta}^{(t)}_{hr},\boldsymbol{\theta}_0),\boldsymbol{v}'_{h01j})\mathbf{1}^{\top}_{n_{h01jr}}$$

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

Collecting Jacobian matrices across sensor measurement types,

$$\boldsymbol{J}_{\boldsymbol{\beta}_{hu},h01}(\boldsymbol{\beta}^{(t)}_{h},\boldsymbol{\theta}_0) = \text{Diag}\left(\boldsymbol{J}_{\boldsymbol{\beta}_{hu1},h011}(\boldsymbol{\beta}^{(t)}_{h1},\boldsymbol{\theta}_0),\boldsymbol{J}_{\boldsymbol{\beta}_{hu2},h012}(\boldsymbol{\beta}^{(t)}_{h2},\boldsymbol{\theta}_0),\ldots,\boldsymbol{J}_{\boldsymbol{\beta}_{huR_h},h01R_h}(\boldsymbol{\beta}^{(t)}_{hR_h},\boldsymbol{\theta}_0)\right)$$

$$\boldsymbol{J}^{\top}_{\boldsymbol{\theta}_0,h01}(\boldsymbol{\beta}^{(t)}_{h},\boldsymbol{\theta}_0) = \left[\boldsymbol{J}^{\top}_{\boldsymbol{\theta}_0,h011}(\boldsymbol{\beta}^{(t)}_{h1},\boldsymbol{\theta}_0) \quad \boldsymbol{J}^{\top}_{\boldsymbol{\theta}_0,h012}(\boldsymbol{\beta}^{(t)}_{h2},\boldsymbol{\theta}_0) \quad \cdots \quad \boldsymbol{J}^{\top}_{\boldsymbol{\theta}_0,h01R_h}(\boldsymbol{\beta}^{(t)}_{hR_h},\boldsymbol{\theta}_0)\right],$$

results in expressing the elements of (19) as $\boldsymbol{J}_{\boldsymbol{\beta}_{hu},h0}(\boldsymbol{\beta}^{(t)}_{h},\boldsymbol{\theta}_0) = \boldsymbol{J}_{\boldsymbol{\beta}_{hu},h01}(\boldsymbol{\beta}^{(t)}_{h},\boldsymbol{\theta}_0)$ for $u \in \{0,t\}$ and $\boldsymbol{J}_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}^{(t)}_{h},\boldsymbol{\theta}_0) = \boldsymbol{J}_{\boldsymbol{\theta}_0,h01}(\boldsymbol{\beta}^{(t)}_{h},\boldsymbol{\theta}_0)$ due to the presence of a single source. Let

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

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

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 benchmark data (replacing index g with '0'), with new elements relating to $\boldsymbol{\theta}_0$:

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

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

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

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

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

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

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

where

$$\mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\beta}} = \text{Diag}\left(\mathcal{I}^0_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_1}, \mathcal{I}^0_{\boldsymbol{\beta}_2,\boldsymbol{\beta}_2}, \ldots, \mathcal{I}^0_{\boldsymbol{\beta}_M,\boldsymbol{\beta}_M}\right)$$

$$\left[\mathcal{I}^0_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right]_{t_1t_2} = \begin{cases} \mathcal{I}^0_{\boldsymbol{\beta}_{ht_1},\boldsymbol{\beta}_{ht_2}}\,, & t_1,t_2 \in \{0,t_{h0}\} \\ \mathbf{0}_{p_{ht_1},p_{ht_2}}\,, & (t_1,t_2) \notin \{0,t_{h0}\}\times\{0,t_{h0}\} \end{cases}$$

$$\mathcal{I}^0_{\boldsymbol{\beta}_{ht_1},\boldsymbol{\beta}_{ht_2}} = \boldsymbol{J}^\top_{\boldsymbol{\beta}_{ht_1},h0}(\boldsymbol{\beta}_h^{(t_{h0})},\boldsymbol{\theta}_0)\boldsymbol{\Omega}^{-1}_{h0}\boldsymbol{J}_{\boldsymbol{\beta}_{ht_2},h0}(\boldsymbol{\beta}_h^{(t_{h0})},\boldsymbol{\theta}_0)\,,\ t_1,t_2 \in \{0,t_{h0}\}$$

$$\left(\mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^\top = \left[\left(\mathcal{I}^0_{\boldsymbol{\beta}_1,\boldsymbol{\theta}_0}\right)^\top \quad \left(\mathcal{I}^0_{\boldsymbol{\beta}_2,\boldsymbol{\theta}_0}\right)^\top \quad \cdots \quad \left(\mathcal{I}^0_{\boldsymbol{\beta}_M,\boldsymbol{\theta}_0}\right)^\top\right]$$

$$\left[\mathcal{I}^0_{\boldsymbol{\beta}_h,\boldsymbol{\theta}_0}\right]_{t1} = \begin{cases} \boldsymbol{J}^\top_{\boldsymbol{\beta}_{ht},h0}(\boldsymbol{\beta}_h^{(t_{h0})},\boldsymbol{\theta}_0)\boldsymbol{\Omega}^{-1}_{h0}\boldsymbol{J}_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t_{h0})},\boldsymbol{\theta}_0)\,, & t \in \{0,t_{h0}\} \\ \mathbf{0}_{p_{ht},q}\,, & t \notin \{0,t_{h0}\} \end{cases}$$

$$\mathcal{I}^0_{\boldsymbol{\theta}_0,\boldsymbol{\theta}_0} = \sum_{h=1}^M \boldsymbol{J}^\top_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t_{h0})},\boldsymbol{\theta}_0)\boldsymbol{\Omega}^{-1}_{h0}\boldsymbol{J}_{\boldsymbol{\theta}_0,h0}(\boldsymbol{\beta}_h^{(t_{h0})},\boldsymbol{\theta}_0)$$

$$\mathcal{I}^0_{\boldsymbol{\varsigma},\boldsymbol{\varsigma}} = \text{Diag}\left(\mathcal{I}^0_{\boldsymbol{\varsigma}_1,\boldsymbol{\varsigma}_1}, \mathcal{I}^0_{\boldsymbol{\varsigma}_2,\boldsymbol{\varsigma}_2}, \ldots, \mathcal{I}^0_{\boldsymbol{\varsigma}_M,\boldsymbol{\varsigma}_M}\right)$$

$$\mathcal{I}^0_{\varsigma_{hk},\varsigma_{hl}} = \frac{1}{2}\text{tr}\left\{\boldsymbol{\Omega}^{-1}_{h0}\frac{\partial\boldsymbol{\Omega}_{h0}}{\partial\varsigma_{hk}}\boldsymbol{\Omega}^{-1}_{h0}\frac{\partial\boldsymbol{\Omega}_{h0}}{\partial\varsigma_{hl}}\right\}.$$

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 $0,1,\ldots,T_h$. Let

$$\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}} = \mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\beta}} + \mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\beta}}$$

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

$$\begin{aligned}\mathcal{I}^{c,0}_{\boldsymbol{\theta}_c|\boldsymbol{\beta}} &= \mathcal{I}^c_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_c} - \left(\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\theta}_c}\right)^\top\left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1}\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\theta}_c} \\ &= \mathcal{I}^c_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_c} - \sum_{h=1}^M\left(\mathcal{I}^c_{\boldsymbol{\beta}_h,\boldsymbol{\theta}_c}\right)^\top\left(\mathcal{I}^{c,0}_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h}\right)^{-1}\mathcal{I}^c_{\boldsymbol{\beta}_h,\boldsymbol{\theta}_c}\end{aligned}$$

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

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

$$\begin{aligned}\mathcal{I}^{c,0}_{\boldsymbol{\theta}_c|\boldsymbol{\beta},\boldsymbol{\theta}_0} &= \mathcal{I}^{c,0}_{\boldsymbol{\theta}_c|\boldsymbol{\beta}} - \left(\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\theta}_c}\right)^\top\left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1}\mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\theta}_0}\left(\mathcal{I}^{c,0}_{\boldsymbol{\theta}_0|\boldsymbol{\beta}}\right)^{-1}\left(\mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^\top\left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1}\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\theta}_c} \qquad (21) \\ &= \mathcal{I}^c_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_c} - \left(\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\theta}_c}\right)^\top\left(\mathcal{I}^{c,0}_{\boldsymbol{\beta}|\boldsymbol{\theta}_0}\right)^{-1}\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\theta}_c}\end{aligned}$$

$$\mathcal{I}^{c,0}_{\boldsymbol{\theta}_0|\boldsymbol{\beta},\boldsymbol{\theta}_c} = \mathcal{I}^{c,0}_{\boldsymbol{\theta}_0|\boldsymbol{\beta}} - \left(\mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^\top\left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1}\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\theta}_c}\left(\mathcal{I}^{c,0}_{\boldsymbol{\theta}_c|\boldsymbol{\beta}}\right)^{-1}\left(\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{\theta}_c}\right)^\top\left(\mathcal{I}^{c,0}_{\boldsymbol{\beta},\boldsymbol{\beta}}\right)^{-1}\mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\theta}_0} \qquad (22)$$

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

$$\mathcal{I}^{c,0}_{\boldsymbol{\varsigma},\boldsymbol{\varsigma}} = \mathcal{I}^c_{\boldsymbol{\varsigma},\boldsymbol{\varsigma}} + \mathcal{I}^0_{\boldsymbol{\varsigma},\boldsymbol{\varsigma}}\,.$$

Computation of (21) and (22) can be simplified by noting that

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

The approximate covariance matrix of the maximum likelihood estimators $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_c, \widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\varsigma}})$ is given by $\left(\mathcal{I}^{c,0}\right)^{-1}$, resulting in a corresponding Gaussian distribution,

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\theta}}_c \\ \widehat{\boldsymbol{\theta}}_0 \\ \widehat{\boldsymbol{\varsigma}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\theta}_c \\ \boldsymbol{\theta}_0 \\ \boldsymbol{\varsigma} \end{pmatrix}, \left(\mathcal{I}^{c,0}\right)^{-1} \right)\,. \tag{23}$$

The marginal distributions of $\widehat{\boldsymbol{\theta}}_c$ and $\widehat{\boldsymbol{\theta}}_0$ are of greatest interest,

$$\widehat{\boldsymbol{\theta}}_c \sim \mathcal{N}\left(\boldsymbol{\theta}_c\,,\, \left(\mathcal{I}^{c,0}_{\boldsymbol{\theta}_c|\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^{-1}\right)$$

$$\widehat{\boldsymbol{\theta}}_0 \sim \mathcal{N}\left(\boldsymbol{\theta}_0\,,\, \left(\mathcal{I}^{c,0}_{\boldsymbol{\theta}_0|\boldsymbol{\beta},\boldsymbol{\theta}_c}\right)^{-1}\right)\,.$$

For *crossed* paths, the conditional independence assumption of the new event and benchmark data may be relaxed by incorporating the new event data into the development of Appendix A.1 as an additional source. This allows common paths observed among the new event data and any of the benchmark events to be correlated, and for the new event data to be included in estimation of the forward model parameters, at the expense of considerable computational time for simultaneous statistical inference of all unknown parameters in one stage.

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_{01}:

$$\boldsymbol{f}_{hgi}(\boldsymbol{\beta}_h) = \begin{bmatrix} \boldsymbol{X}_{hgi} & \mathbf{1}_{n_{hgi}} W_{gi} \end{bmatrix} \boldsymbol{\beta}_h$$

$$\boldsymbol{f}_{h01}(\boldsymbol{\beta}_h, W_{01}) = \begin{bmatrix} \boldsymbol{X}_{h01} & \mathbf{1}_{n_{h01}} W_{01} \end{bmatrix} \boldsymbol{\beta}_h\,.$$

Here $T_h = 0$. Simplifying to the case where each benchmark source group g is populated with a single source ($M_{hg} = 1$), it is straightforward to calculate

$$\boldsymbol{J}_{hg}(\boldsymbol{\beta}_h) = \begin{bmatrix} \boldsymbol{X}_{hg1} & \mathbf{1}_{n_{hg1}} W_{g1} \end{bmatrix}$$

$$\boldsymbol{J}_{\boldsymbol{\beta}_h,h0}(\boldsymbol{\beta}_h, W_{01}) = \begin{bmatrix}\boldsymbol{X}_{h01} & \mathbf{1}_{n_{h01}} W_{01}\end{bmatrix}$$
$$\boldsymbol{J}_{W_{01},h0}(\boldsymbol{\beta}_h, W_{01}) = \mathbf{1}_{n_{h01}} \omega_h$$

resulting in

$$\begin{aligned}\mathcal{I}^{c,0}_{\boldsymbol{\beta}_h,\boldsymbol{\beta}_h} &= \sum_{g=0}^{M_h} \begin{bmatrix} \boldsymbol{X}_{hg1}^\top \boldsymbol{\Omega}_{hg}^{-1} \boldsymbol{X}_{hg1} & W_{g1} \boldsymbol{X}_{hg1}^\top \boldsymbol{\Omega}_{hg}^{-1} \mathbf{1}_{n_{hg1}} \\ W_{g1} \mathbf{1}_{n_{hg1}}^\top \boldsymbol{\Omega}_{hg}^{-1} \boldsymbol{X}_{hg1} & W_{g1}^2 \mathbf{1}_{n_{hg1}}^\top \boldsymbol{\Omega}_{hg}^{-1} \mathbf{1}_{n_{hg1}} \end{bmatrix} \\ &= \begin{bmatrix} \mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\boldsymbol{\mu}_h} & \mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\omega_h} \\ \left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\omega_h}\right)^\top & \mathcal{I}^{c,0}_{\omega_h,\omega_h} \end{bmatrix},\end{aligned}$$

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

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

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

Utilizing the following definitions,

$$\begin{aligned}\boldsymbol{X}_{01} &= \text{Diag}\left(\boldsymbol{X}_{101}, \boldsymbol{X}_{201}, \ldots, \boldsymbol{X}_{M01}\right) \\ \boldsymbol{U}_{01} &= \text{Diag}\left(\mathbf{1}_{n_{101}}, \mathbf{1}_{n_{201}}, \ldots, \mathbf{1}_{n_{M01}}\right) \\ \overline{\boldsymbol{X}}_{01} &= \boldsymbol{U}_{01}^\top \boldsymbol{\Omega}_0^{-1} \boldsymbol{X}_{01} \\ \mathcal{I}^{c,0}_{\boldsymbol{\mu},\boldsymbol{\mu}} &= \text{Diag}\left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_1}, \mathcal{I}^{c,0}_{\boldsymbol{\mu}_2,\boldsymbol{\mu}_2}, \ldots, \mathcal{I}^{c,0}_{\boldsymbol{\mu}_M,\boldsymbol{\mu}_M}\right) \\ \mathcal{I}^{c,0}_{\boldsymbol{\mu},\boldsymbol{\omega}} &= \text{Diag}\left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_1,\omega_1}, \mathcal{I}^{c,0}_{\boldsymbol{\mu}_2,\omega_2}, \ldots, \mathcal{I}^{c,0}_{\boldsymbol{\mu}_M,\omega_M}\right) \\ \widetilde{\boldsymbol{X}}_{01} &= \overline{\boldsymbol{X}}_{01} \left(\mathcal{I}^{c,0}_{\boldsymbol{\mu},\boldsymbol{\mu}}\right)^{-1} \mathcal{I}^{c,0}_{\boldsymbol{\mu},\boldsymbol{\omega}} \\ \mathcal{I}^{c,0}_{\boldsymbol{\omega}|\boldsymbol{\mu}} &= \text{Diag}\left(\mathcal{I}^{c,0}_{\omega_1|\boldsymbol{\mu}_1}, \mathcal{I}^{c,0}_{\omega_2|\boldsymbol{\mu}_2}, \ldots, \mathcal{I}^{c,0}_{\omega_M|\boldsymbol{\mu}_M}\right) \\ \boldsymbol{\omega} &= \left(\omega_1, \omega_2, \ldots, \omega_M\right),\end{aligned}$$

the maximum likelihood estimator (MLE) $\widehat{W}_{01}$ of W_{01} has asymptotic variance $V\left[\widehat{W}_{01}\right]$ given by

$$\begin{aligned}V^{-1}\left[\widehat{W}_{01}\right] &= \boldsymbol{\omega}^\top \left[\boldsymbol{U}_{01}^\top \boldsymbol{\Omega}_0^{-1} \boldsymbol{U}_{01} - \overline{\boldsymbol{X}}_{01} \left(\mathcal{I}^{c,0}_{\boldsymbol{\mu},\boldsymbol{\mu}}\right)^{-1} \overline{\boldsymbol{X}}_{01}^\top\right] \boldsymbol{\omega} \\ &\quad - \boldsymbol{\omega}^\top \left(\widetilde{\boldsymbol{X}}_{01} - W_{01} \boldsymbol{U}_{01}^\top \boldsymbol{\Omega}_0^{-1} \boldsymbol{U}_{01}\right) \left(\mathcal{I}^{c,0}_{\boldsymbol{\omega}|\boldsymbol{\mu}}\right)^{-1} \left(\widetilde{\boldsymbol{X}}_{01}^\top - W_{01} \boldsymbol{U}_{01}^\top \boldsymbol{\Omega}_0^{-1} \boldsymbol{U}_{01}\right) \boldsymbol{\omega},\end{aligned}$$

where MLEs replace parameters in this expression to obtain the estimate $\hat{V}\left[\widehat{W}_{01}\right]$.

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

The above framework, similar to traditional modeling of high-yield nuclear events of the TTBT era, assumes yields for sources in the benchmark dataset(s) are known with certainty (and thus included in the covariate vectors $\boldsymbol{v}_{hgij}$). For lower-yield events, uncertainties may not be negligible compared with other sources of variability. To accommodate this situation for nuclear yield data, let $\mathcal{S}$ denote the index set of unique sources (across sensor types) contained in the benchmark 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 $\mathcal{L}_c(\boldsymbol{\beta}, \boldsymbol{\theta}_c, \boldsymbol{\varsigma})$ for the benchmark data. That is, log-yield is removed from the covariate vectors $\boldsymbol{v}_{hgij}$ and the benchmark log-likelihood function is written $\mathcal{L}_c(\boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{\theta}_c, \boldsymbol{\varsigma})$ to represent its additional dependence on the unknown log-yields $\boldsymbol{w}$ for each benchmark source.

For the high explosive (HE) events, quantifying uncertainty in the chemical-to-nuclear (C2N) factor used to convert conventional HE yields to their equivalent nuclear yields is a highly problematic exercise. Ideally, a yield model structure for equivalent nuclear yields that is similar in spirit to Equation (3) would be available from data sources in the open literature, based on formally comparing sensor signatures across nuclear and HE explosions for several scenarios of interest. That is, nuclear and HE benchmark data would consist of experimental results for several seismic and acoustic signatures obtained from several emplacement conditions such as above/below ground and in different geology types. This would allow for the apples-to-apples signature-specific comparisons required to derive C2N factors for each scenario of interest.

Unfortunately, we are unaware of any such data sources. Indeed, because U.S. nuclear testing has ceased since the early 1990s, obtaining "new" nuclear data for this purpose is not realistic. Rather than inventing scenario-specific C2N factors out of thin air, those factors are treated as being known exactly for the results herein, where the one-size-fits-all value C2N=2 is used across all scenarios for simplicity.

Importantly, the MultiPEM Toolbox software is capable of incorporating new C2N-related information, perhaps based on computer simulation, were it to exist. Estimation and uncertainty quantification of empirical parameters in a C2N model would be accomplished through the calibration parameter $\boldsymbol{\theta}_c$. The benchmark data set would contain relevant signatures from both chemical and nuclear explosions, allowing the C2N empirical parameters $\boldsymbol{\theta}_c$ to be inferred with uncertainties that would then be propagated to the estimation and uncertainty quantification of new event quantities of interest $\boldsymbol{\theta}_0$.

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

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

assuming conditional independence between the new event and benchmark data, so that the log-likelihood function from the previous development is augmented to accommodate a term informing on the actual device yields,

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{\theta}_c, \boldsymbol{\theta}_0, \boldsymbol{\varsigma}) = \mathcal{L}_c(\boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{\theta}_c, \boldsymbol{\varsigma}) + \mathcal{L}_0(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0, \boldsymbol{\varsigma}) + \mathcal{L}_w(\boldsymbol{w})\,. \tag{24}$$

The information matrices required for estimating the asymptotic covariance matrices of the MLEs for $\boldsymbol{\theta}_c$ and $\boldsymbol{\theta}_0$ 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 $s \in \mathcal{S}$. The index "hg_s" represents the source group $g \in \{1, 2, \ldots, M_h\}$ containing source s for sensor type h, and the index "hg_si_s" represents the source $i \in \{1, 2, \ldots, M_{hg_s}\}$ aligned with source s in group g_s for sensor type h. Let

$$\boldsymbol{g}_{hgi}(\boldsymbol{\beta}_h^{(t_{hg})}, w, \boldsymbol{\theta}_c) = \frac{\partial \boldsymbol{f}_{hgi}(\boldsymbol{\beta}_h^{(t_{hg})}, w, \boldsymbol{\theta}_c)}{\partial w}$$

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

$$g_{hgir}(\boldsymbol{\beta}_{hr}^{(t_{hg})}, w, \boldsymbol{\theta}_c, \boldsymbol{v}''_{hgij}) = \frac{\partial f_{hgir}(\boldsymbol{\beta}_{hr}^{(t_{hg})}, w, \boldsymbol{\theta}_c, \boldsymbol{v}''_{hgij})}{\partial w}\,.$$

In similar fashion, the Jacobian matrices of $\boldsymbol{f}_{hgi}(\boldsymbol{\beta}_h^{(t_{hg})}, w, \boldsymbol{\theta}_c)$ with respect to $\boldsymbol{\beta}_{hu}$ for $u \in \{0, t_{hg}\}$ and $\boldsymbol{\theta}_c$ are given by

$$\boldsymbol{J}_{\boldsymbol{\beta}_{hu},hgi}(\boldsymbol{\beta}_h^{(t_{hg})}, w, \boldsymbol{\theta}_c) = \left[\frac{\partial \boldsymbol{f}_{hgi}(\boldsymbol{\beta}_h^{(t_{hg})}, w, \boldsymbol{\theta}_c)}{\partial \beta_{hu1}} \quad \frac{\partial \boldsymbol{f}_{hgi}(\boldsymbol{\beta}_h^{(t_{hg})}, w, \boldsymbol{\theta}_c)}{\partial \beta_{hu2}} \quad \cdots \quad \frac{\partial \boldsymbol{f}_{hgi}(\boldsymbol{\beta}_h^{(t_{hg})}, w, \boldsymbol{\theta}_c)}{\partial \beta_{hup_{hu}}}\right]$$

$$\boldsymbol{J}_{\boldsymbol{\theta}_c,hgi}(\boldsymbol{\beta}_h^{(t_{hg})}, w, \boldsymbol{\theta}_c) = \left[\frac{\partial \boldsymbol{f}_{hgi}(\boldsymbol{\beta}_h^{(t_{hg})}, w, \boldsymbol{\theta}_c)}{\partial \theta_{c1}} \quad \frac{\partial \boldsymbol{f}_{hgi}(\boldsymbol{\beta}_h^{(t_{hg})}, w, \boldsymbol{\theta}_c)}{\partial \theta_{c2}} \quad \cdots \quad \frac{\partial \boldsymbol{f}_{hgi}(\boldsymbol{\beta}_h^{(t_{hg})}, w, \boldsymbol{\theta}_c)}{\partial \theta_{c\tilde{q}}}\right].$$

Let $\boldsymbol{g}_{hg_s}(\boldsymbol{\beta}_h^{(t_{hg_s})}, w, \boldsymbol{\theta}_c)$ be a $n_{hg_s} \times 1$ vector of zeroes, except for the elements corresponding to source hg_si_s, which are filled with the vector $\boldsymbol{g}_{hg_si_s}(\boldsymbol{\beta}_h^{(t_{hg_s})}, w, \boldsymbol{\theta}_c)$, $\boldsymbol{J}_{\boldsymbol{\beta}_{hu},hg_s}(\boldsymbol{\beta}_h^{(t_{hg_s})}, \boldsymbol{w}_{hg_s}, \boldsymbol{\theta}_c)$ be a $n_{hg_s} \times p_{hu}$ matrix having row blocks filled with the matrices $\boldsymbol{J}_{\boldsymbol{\beta}_{hu},hg_si}(\boldsymbol{\beta}_h^{(t_{hg_s})}, w_{hg_si}, \boldsymbol{\theta}_c)$, and $\boldsymbol{J}_{\boldsymbol{\theta}_c,hg_s}(\boldsymbol{\beta}_h^{(t_{hg_s})}, \boldsymbol{w}_{hg_s}, \boldsymbol{\theta}_c)$ be a $n_{hg_s} \times \tilde{q}$ matrix having row blocks filled with the matrices $\boldsymbol{J}_{\boldsymbol{\theta}_c,hg_si}(\boldsymbol{\beta}_h^{(t_{hg_s})}, w_{hg_si}, \boldsymbol{\theta}_c)$, for $i = 1, \ldots, M_{hg_s}$. The vector $\boldsymbol{w}_{hg_s} = (w_{hg_s1}, w_{hg_s2}, \ldots, w_{hg_sM_{hg_s}})$ contains the unknown benchmark yields corresponding to the source group "hg_s". The partial derivatives of the log-likelihood function for the benchmark data with respect to log-yield w_s for $s \in \mathcal{S}$ are given by

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

with relevant second partial derivatives given by

$$\frac{\partial^2 \mathcal{L}_c}{\partial w_s^2} = -\sum_{h \in \mathcal{H}_s} \left(\boldsymbol{g}_{hg_s}(\boldsymbol{\beta}_h^{(t_{hg_s})}, w_s, \boldsymbol{\theta}_c) \right)^\top \boldsymbol{\Omega}_{hg_s}^{-1} \boldsymbol{g}_{hg_s}(\boldsymbol{\beta}_h^{(t_{hg_s})}, w_s, \boldsymbol{\theta}_c)$$

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

$$\frac{\partial^2 \mathcal{L}_c}{\partial w_s \partial w_t} = -\sum_{h \in \mathcal{H}_s \cap \mathcal{H}_t} \left(\boldsymbol{g}_{hg_s}(\boldsymbol{\beta}_h^{(t_{hg_s})}, w_s, \boldsymbol{\theta}_c) \right)^\top \boldsymbol{\Omega}_{hg_s}^{-1} \boldsymbol{g}_{hg_t}(\boldsymbol{\beta}_h^{(t_{hg_t})}, w_t, \boldsymbol{\theta}_c)\, \chi\left[hg_t = hg_s\right], \quad \text{for } s \neq t\,,$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial w_s \partial \boldsymbol{\beta}_{hu}^\top} = -\left(\boldsymbol{g}_{hg_s}(\boldsymbol{\beta}_h^{(t_{hg_s})}, w_s, \boldsymbol{\theta}_c) \right)^\top \boldsymbol{\Omega}_{hg_s}^{-1} \boldsymbol{J}_{\boldsymbol{\beta}_{hu}, hg_s}(\boldsymbol{\beta}_h^{(t_{hg_s})}, \boldsymbol{w}_{hg_s}, \boldsymbol{\theta}_c)$$

$$+ \left(\boldsymbol{r}'_{hg_s t_{hg_s}} \right)^\top \boldsymbol{\Omega}_{hg_s}^{-1} \left(\frac{\partial \boldsymbol{g}_{hg_s}(\boldsymbol{\beta}_h^{(t_{hg_s})}, w_s, \boldsymbol{\theta}_c)}{\partial \boldsymbol{\beta}_{hu}^\top} \right) \quad \text{for } h \in \mathcal{H}_s \text{ and } u \in \{0, t_{hg_s}\}$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial w_s \partial \boldsymbol{\beta}_{hu}^\top} = \mathbf{0}_{1,p_{hu}} \quad \text{for } h \notin \mathcal{H}_s \text{ or } h \in \mathcal{H}_s \text{ and } u \notin \{0, t_{hg_s}\}$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial w_s \partial \boldsymbol{\theta}_c^\top} = -\sum_{h \in \mathcal{H}_s} \left(\boldsymbol{g}_{hg_s}(\boldsymbol{\beta}_h^{(t_{hg_s})}, w_s, \boldsymbol{\theta}_c) \right)^\top \boldsymbol{\Omega}_{hg_s}^{-1} \boldsymbol{J}_{\boldsymbol{\theta}_c, hg_s}(\boldsymbol{\beta}_h^{(t_{hg_s})}, \boldsymbol{w}_{hg_s}, \boldsymbol{\theta}_c)$$

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

$$\frac{\partial^2 \mathcal{L}_c}{\partial w_s \partial \boldsymbol{\theta}_0^\top} = \mathbf{0}_{1,q}$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial w_s \partial \varsigma_{hk}} = -\left(\boldsymbol{g}_{hg_s}(\boldsymbol{\beta}_h^{(t_{hg_s})}, w_s, \boldsymbol{\theta}_c) \right)^\top \boldsymbol{\Omega}_{hg_s}^{-1} \frac{\partial \boldsymbol{\Omega}_{hg_s}}{\partial \varsigma_{hk}} \boldsymbol{\Omega}_{hg_s}^{-1} \boldsymbol{r}'_{hg_s t_{hg_s}} \quad \text{for } h \in \mathcal{H}_s$$

$$\frac{\partial^2 \mathcal{L}_c}{\partial w_s \partial \varsigma_{hk}} = 0 \quad \text{for } h \notin \mathcal{H}_s$$

with $\chi[A]$ the indicator function of condition A. The log-likelihood function for pre-shot design or post-shot measured log-yields $\widetilde{w}_s$ with $s \in \mathcal{S}$ corresponding to the benchmark 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

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

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

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 $(\boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{\theta}_c, \boldsymbol{\theta}_0, \boldsymbol{\varsigma})$:

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

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

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

$$\mathcal{I}^c_{w_s,w_t} = \sum_{h \in \mathcal{H}_s \cap \mathcal{H}_t} \left(\boldsymbol{g}_{hg_s}(\boldsymbol{\beta}_h^{(t_{hg_s})}, w_s, \boldsymbol{\theta}_c)\right)^\top \boldsymbol{\Omega}^{-1}_{hg_s}\, \boldsymbol{g}_{hg_t}(\boldsymbol{\beta}_h^{(t_{hg_t})}, w_t, \boldsymbol{\theta}_c)\, \chi\left[hg_t = hg_s\right], \quad \text{for } s \neq t\,,$$

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

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

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

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

$$\left[\mathcal{I}^c_{\boldsymbol{w},\boldsymbol{\theta}_c}\right]_s = \mathcal{I}^c_{w_s,\boldsymbol{\theta}_c}\,, \quad \text{for } s = 1, \ldots, |\mathcal{S}|\,,$$

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

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

$$\mathcal{I}^{c,w,0}_{\boldsymbol{\theta}_0|\boldsymbol{\beta},\boldsymbol{w},\boldsymbol{\theta}_c} = \mathcal{I}^0_{\boldsymbol{\theta}_0,\boldsymbol{\theta}_0} - \left(\mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^\top \left(\mathcal{I}^{c,0}_{\boldsymbol{\beta}|\boldsymbol{\theta}_c} - \mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{w}|\boldsymbol{\theta}_c}\left(\mathcal{I}^c_{\boldsymbol{w}|\boldsymbol{\theta}_c}\right)^{-1}\left(\mathcal{I}^c_{\boldsymbol{\beta},\boldsymbol{w}|\boldsymbol{\theta}_c}\right)^\top\right)^{-1} \mathcal{I}^0_{\boldsymbol{\beta},\boldsymbol{\theta}_0} \tag{25}$$

for

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

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

From (25) it is seen that the second term on the right-hand side of the equality represents the loss of information in estimating the new event quantities of interest $\boldsymbol{\theta}_0$ due to estimating the forward model parameters $\boldsymbol{\beta}$. In turn, there is an additional loss of information in estimating $\boldsymbol{\beta}$ due to estimating the log-yields $\boldsymbol{w}$ rather than assuming they are known and fixed at their design/measured values. Finally, estimation of the calibration parameters $\boldsymbol{\theta}_c$ causes additional information loss in estimating both $\boldsymbol{\beta}$ and $\boldsymbol{w}$.

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

$$\mathcal{I}^{c,w,0}_{\boldsymbol{\theta}_c|\boldsymbol{\beta},\boldsymbol{w},\boldsymbol{\theta}_0} = \mathcal{I}^{c,0}_{\boldsymbol{\theta}_c|\boldsymbol{\beta},\boldsymbol{\theta}_0} - \left(\mathcal{I}^{c,0}_{\boldsymbol{w},\boldsymbol{\theta}_c|\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^{\top} \left(\mathcal{I}^{c,0}_{\boldsymbol{w}|\boldsymbol{\beta},\boldsymbol{\theta}_0}\right)^{-1} \mathcal{I}^{c,0}_{\boldsymbol{w},\boldsymbol{\theta}_c|\boldsymbol{\beta},\boldsymbol{\theta}_0} \tag{26}$$

for

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

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

From (26) it is seen that the first term on the right-hand side of the equality represents the loss of information in estimating the calibration parameters $\boldsymbol{\theta}_c$ due to estimating the forward model parameters $\boldsymbol{\beta}$ and the new event quantities of interest $\boldsymbol{\theta}_0$. In turn, there is an additional loss of information in estimating $\boldsymbol{\theta}_c$ due to estimating the log-yields $\boldsymbol{w}$ rather than assuming they are known and fixed at their design/measured values.

In practice, increased standard errors in estimates of $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_c$ suggested by expressions (25) and (26) 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. The consequence is an inability to arrive at any firm conclusion regarding the impact of errors-in-variables treatment on uncertainty quantification for $\boldsymbol{\theta}_0$ or $\boldsymbol{\theta}_c$.

As discussed in Appendix A.2, the conditional independence assumption of the new event and benchmark data may be relaxed for *crossed* paths by incorporating the new event data into the development of Appendix A.1 as an additional source.

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

Returning to the linear model formulation, define

$$\overline{\boldsymbol{x}}_{hu1} = \boldsymbol{X}^{\top}_{hu1} \boldsymbol{\Omega}^{-1}_{hu} \mathbf{1}_{n_{hu1}}$$

$$\widetilde{X}_{hu1} = \overline{\boldsymbol{x}}^{\top}_{hu1} \left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\boldsymbol{\mu}_h}\right)^{-1} \mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\omega_h}$$

and

$$q_h(u,v) = \overline{\boldsymbol{x}}_{hu1}^\top \left(\mathcal{I}^{c,0}_{\boldsymbol{\mu}_h,\boldsymbol{\mu}_h}\right)^{-1} \overline{\boldsymbol{x}}_{hv1} + \left(\widetilde{X}_{hu1} - W_{u1}\mathbf{1}_{n_{hu1}}^\top \boldsymbol{\Omega}_{hu}^{-1}\mathbf{1}_{n_{hu1}}\right)\left(\mathcal{I}^{c,0}_{\omega_h|\boldsymbol{\mu}_h}\right)^{-1}\left(\widetilde{X}_{hv1} - W_{v1}\mathbf{1}_{n_{hv1}}^\top \boldsymbol{\Omega}_{hv}^{-1}\mathbf{1}_{n_{hv1}}\right).$$

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

$$V_{\text{eiv}}^{-1}\left[\widehat{W}_{01}\right] = V^{-1}\left[\widehat{W}_{01}\right] - \boldsymbol{\omega}^\top \boldsymbol{Q}_0 \left(\mathcal{I}^{c,w,0}_{\boldsymbol{w}|\boldsymbol{\beta}}\right)^{-1} \boldsymbol{Q}_0^\top \boldsymbol{\omega}$$

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

$$\left[\sum_{h\in\mathcal{H}_u} \omega_h^2\, \mathbf{1}_{n_{hu1}}^\top \boldsymbol{\Omega}_{hu}^{-1}\mathbf{1}_{n_{hu1}} + \frac{1}{\sigma_u^2}\right]\chi\left[u=v\right] - \sum_{h\in\mathcal{H}_u\cap\mathcal{H}_v} \omega_h^2\, q_h(u,v)$$

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 physical/empirical models will be simplified to remove covariate information from the model for source and path effects. That is, both $\boldsymbol{Z}_{hgir,j}$ and $\boldsymbol{Z}_{hjr,gi}$ are set to $\mathbf{1}_{n_{hgijr}}$. 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 ("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 $\boldsymbol{\Omega}_{hg}$ corresponding to the source and path errors are given as follows,

$$\frac{\partial \boldsymbol{\Omega}_{hg}}{\partial \tau_{S_{hr}}} = \frac{\partial \boldsymbol{\Xi}^{(S)}_{hg}}{\partial \tau_{S_{hr}}} = \text{Diag}\left(\boldsymbol{\Delta}^S_{hg1}, \boldsymbol{\Delta}^S_{hg2}, \ldots, \boldsymbol{\Delta}^S_{hgM_{hg}}\right)$$

$$\frac{\partial \boldsymbol{\Omega}_{hg}}{\partial \tau_{P_{hr}}} = \frac{\partial \boldsymbol{\Xi}_{hg}^{(P)}}{\partial \tau_{P_{hr}}} = \begin{bmatrix} \boldsymbol{\Delta}_{hg}^{P}[1,1] & \boldsymbol{\Delta}_{hg}^{P}[1,2] & \cdots & \boldsymbol{\Delta}_{hg}^{P}[1,M_{hg}] \\ \left(\boldsymbol{\Delta}_{hg}^{P}[1,2]\right)^{\top} & \boldsymbol{\Delta}_{hg}^{P}[2,2] & \cdots & \boldsymbol{\Delta}_{hg}^{P}[2,M_{hg}] \\ \vdots & \vdots & \ddots & \vdots \\ \left(\boldsymbol{\Delta}_{hg}^{P}[1,M_{hg}]\right)^{\top} & \left(\boldsymbol{\Delta}_{hg}^{P}[2,M_{hg}]\right)^{\top} & \cdots & \boldsymbol{\Delta}_{hg}^{P}[M_{hg},M_{hg}] \end{bmatrix},$$

where

$$\boldsymbol{\Delta}_{hgi}^{S} = \mathrm{Diag}\left(\mathbf{0}_{n_{hgi1},n_{hgi1}}, \ldots, \sigma_{S_{hr}}^{2} \boldsymbol{Z}_{hgir}^{S} \left(\boldsymbol{Z}_{hgir}^{S}\right)^{\top}, \ldots, \mathbf{0}_{n_{hgiR_h},n_{hgiR_h}}\right)$$

$$\boldsymbol{\Delta}_{hg}^{P}[i_1,i_2] = \begin{bmatrix} & & & \\ \mathbf{0}_{n_{hgi_1},n_{hgi_21}} & \cdots & \begin{pmatrix} \mathbf{0}_{n_{hgi_11},n_{hgi_2r}} \\ \vdots \\ \sigma_{P_{hr}}^{2} \boldsymbol{Z}_{hgi_1r}^{P} \left(\boldsymbol{Z}_{hgi_2r}^{P}\right)^{\top} \\ \vdots \\ \mathbf{0}_{n_{hgi_1R_h},n_{hgi_2r}} \end{pmatrix} & \cdots & \mathbf{0}_{n_{hgi_1},n_{hgi_2R_h}} \\ & & & \end{bmatrix},$$

for $i_1, i_2 \in \{1, \ldots, M_{hg}\}$. The terms in these matrices are given by

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

$$\boldsymbol{Z}_{hgi_1r}^{P} \left(\boldsymbol{Z}_{hgi_2r}^{P}\right)^{\top} = \begin{cases} \boldsymbol{D}_{hgi_1r} \boldsymbol{P}_{hgi_1r} \boldsymbol{P}_{hgi_2r}^{\top} \boldsymbol{D}_{hgi_2r}^{\top}, & r \in \mathcal{R}_h^P \\ \mathbf{0}_{n_{hgi_1r},n_{hgi_2r}}, & r \notin \mathcal{R}_h^P \end{cases},$$

with $\boldsymbol{J}_m$ the $m \times m$ matrix of ones and

$$\boldsymbol{D}_{hgir} = \mathrm{Diag}\left(\mathbf{1}_{n_{hgi1r}}, \mathbf{1}_{n_{hgi2r}}, \ldots, \mathbf{1}_{n_{hgiM_{hgi}r}}\right).$$

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

$$\frac{\partial \boldsymbol{\Omega}_{hg}}{\partial \upsilon} = \frac{\partial \boldsymbol{\Sigma}_{hg}}{\partial \upsilon} = \mathrm{Diag}\left(\frac{\partial \boldsymbol{\Sigma}_{hg1}}{\partial \upsilon}, \frac{\partial \boldsymbol{\Sigma}_{hg2}}{\partial \upsilon}, \ldots, \frac{\partial \boldsymbol{\Sigma}_{hgM_{hg}}}{\partial \upsilon}\right)$$

where

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

for

$$\left[\frac{\partial \boldsymbol{\Sigma}_{hgi}}{\partial \sigma_{h,r_1 r_2}}\right]_{r_1,r_2} = \sum_{(k_1,k_2) \in \mathcal{J}_{r_1,r_2}^{hgi}} \boldsymbol{e}_{hgir_1,k_1} \boldsymbol{e}_{hgir_2,k_2}^{\top}$$

$$\left[\frac{\partial \boldsymbol{\Sigma}_{hgi}}{\partial \sigma_{h,r_1 r_2}}\right]_{r_2,r_1} = \sum_{(k_1,k_2) \in \mathcal{J}_{r_1,r_2}^{hgi}} \boldsymbol{e}_{hgir_2,k_2} \boldsymbol{e}_{hgir_1,k_1}^{\top}$$

$$\left[\frac{\partial \boldsymbol{\Sigma}_{hgi}}{\partial \sigma_{h,r_1 r_2}}\right]_{rs} = \mathbf{0}_{n_{hgir},n_{hgis}} \,,\ (r,s) \notin \{(r_1,r_2),(r_2,r_1)\}$$

and

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

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

for $s > r$. In this treatment,

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