NPL
National Physical Laboratory

NPL REPORT AC 25

A STUDY OF UNCERTAINTY PROPAGATION FOR AN END-TO-END DATA PROCESSING PIPELINE FOR AN APPLICATION IN UNDERWATER ACOUSTICS

PETER HARRIS, STEPHEN ROBINSON AND LIAN WANG

DECEMBER 2023

A STUDY OF UNCERTAINTY PROPAGATION FOR AN END-TO-END DATA PROCESSING PIPELINE FOR AN APPLICATION IN UNDERWATER ACOUSTICS

Peter Harris
Data Analytics and Modelling group

Stephen Robinson and Lian Wang
Ultrasound and Underwater Acoustics group

ISSN 1754-2936

https://doi.org/10.47120/npl.AC25

National Physical Laboratory
Hampton Road, Teddington, Middlesex, TW11 0LW

Approved on behalf of NPLML by
Louise Wright, Head of Science for the Data Science department.

CONTENTS

1 BACKGROUND

The data recorded by a sensor operating in the field, possibly as part of a deployed sensor network, are used by different end-user communities for different purposes. For example, in the case of underwater acoustic measurement, the data might be used for event detection and attribution to derive temporal, spatial and amplitude information about the event, which can be anthropogenic or natural. Another use might be for ocean noise monitoring in which the data are used to derive metrics for ambient noise maps for a given spatial region and time period, for example, the percentage of time that an exposure threshold for sound pressure level is exceeded. Yet another use might be for environmental monitoring in which the data are used to derive long-term (e.g., decadal) trends in sound pressure level and to correlate characteristics of the recorded sound with natural and anthropogenic sound sources.

The Joint Research Project 19ENV03 “Infra-AUV” of the European Metrology Programme for Innovation and Research (EMPIR) [1] has delivered traceable calibration of acoustic, underwater, and seismic sensors for measurements at low frequencies, as well as improved knowledge of their performance in-situ. Various case studies have been used to demonstrate the impact for different end-user communities of that traceability and improved knowledge. Specifically, the project has investigated methods for the propagation of uncertainty, including that associated with the calibration of a sensor, through models for high-level derived parameters related to various end-user applications. Quantifying reliably the uncertainty for estimates of these high-level parameters is essential when those estimates are used for decision-making or to inform policy, as well as to understand the comparability and consistency of estimates relating to different locations and times.

The objective of the study described in this report is to investigate the propagation of uncertainty through models related to a particular end user application, viz., a study of the data provided by acoustic sensors at the hydroacoustic stations that form part of the International Monitoring System operated by the Commission for the Comprehensive Nuclear-Test-Ban Treaty (CTBT) [2]. Each hydroacoustic station consists of three acoustic sensors placed in the deep ocean sound channel where the vertical sound speed profile exhibits a minimum. The locations of the stations allow good spatial coverage of the world’s oceans by taking advantage of the physical principles governing the propagation of sound in water. The sampling frequency for the sound pressure recordings is 250 Hz to provide information at acoustic frequencies up to 100 Hz, and a bit depth of 24 bits yields a maximum possible dynamic range of approximately 144 dB. Data is available over periods of years, and at some stations for periods in excess of ten years. Although the primary aim of the data is to support the monitoring of the Comprehensive Nuclear Test Ban Treaty, it is also made available by to support environmental research and, consequently, is used by users in many other and diverse applications.

The report is organised as follows. In Section 2 we describe the data processing chain from the raw measured data recorded by a hydroacoustic sensor to advanced derived parameters of interest to a user. In Sections 3 and 4 we describe, respectively, the sources of uncertainty considered to influence the data processing chain and the numerical method used to propagate those uncertainties through the chain. In Sections 5 to 8 we illustrate the results of that uncertainty propagation for a snapshot of data recorded by a single hydroacoustic sensor over a period of four days. A summary is given in Section 9. Finally, Section 12 contains the results in graphical and numerical forms.

2 DATA PROCESSING CHAIN

The data processing chain is composed of the following stages [3].

2.1 STAGE 1

The input to this stage is a time series of raw measured data in counts that are provided by the measuring system (an analogue to digital converter attached to a hydroacoustic sensor) with a sampling frequency of 250 Hz. The output is a time series of values of sound power spectral density level (SPSDL) in the unit of $\mathrm{dB\ re\ 1\ \mu Pa^2/Hz}$. The output depends on the choice of an averaging time interval for extracting frequency information and the choice of a frequency band. Examples of averaging time intervals are 1 minute and 10 minutes, and the choice of averaging time interval determines the sampling frequency of the output time series, i.e., the sampling frequency is 1 per minute for the choice of a 1-minute averaging time interval. Examples of frequency bands are 10—100 Hz (broadband), 1—10 Hz (very low), 10—40 Hz (low), 40—70 Hz (medium), and 70—100 Hz (high). The output is derived from the input using information about the measuring system in terms of a scaling factor (for converting counts to values of sound pressure in pascals) and its calibration in terms of a frequency response (for describing the frequency-dependent behaviour of the measuring system).

This stage of the data processing chain is described as follows. Let $(t_{0,i}, S_{0,i}), i = 1, \dots, n_0$, denote the input to stage 1 comprising a time series of raw measured data in counts. Let G denote the scaling factor, and $(f_{c,j}, A_{c,j}, \phi_{c,j}), j = 0, \dots n_c$, denote the amplitude and phase responses of the hydroacoustic sensor defined by calibration frequencies $f_{c,j}$ (with $f_{c,0} = 0\ \mathrm{Hz}$) and calibrated amplitudes $A_{c,j}$ and calibrated phases $\phi_{c,j}$. Furthermore, let ΔT_1 denote the averaging time interval, and (f_L, f_U) the frequency band.
The measurement function is specified as follows. For the rth averaging time interval defined by start-time $(r-1)\Delta T_1$ and end-time $r\Delta T_1$:

1.1 Identify those points with indices I_r for which $(r-1)\Delta T_1 \le t_{0,i} < r\Delta T_1$.

1.2 Apply the scaling factor: $S_{r,i} = GS_{0,i}. i \in I_r$.

1.3 Form the (singled-sided) FFT $(f_{r,j}, A_{r,j}, \phi_{r,j}), j = 0, \dots n_r$, of the signal values $S_{r,i}, i \in I_r$, defined by frequencies $f_{r,j} \in [0, F_s/2]$, amplitudes $A_{r,j}$ and phases $\phi_{r,j}$, where $F_s = 250\ \mathrm{Hz}$ is the sampling frequency for the signal values.

1.4 Interpolate the amplitude and phase responses of the hydroacoustic sensor to the frequencies of the FFT to give calibration data $(f_{r,j}, A'_{c,j}, \phi'_{c,j}), j = 0 = 1, \dots n_r$.

1.5 Remove the effect of the hydrophone response from the signal values: $A'_{r,j} = A_{r,j}/A'_{c,j}$ and $\phi'_{r,j} = \phi_{r,j} - \phi'_{c,j}, j = 0, \dots, n_r$.

1.6 Identify those points with indices J_r for which $f_L \le f_{r,j} \le f_U$.

1.7 Calculate $t_{1,r} = (r-1)\Delta T_1$.

1.8 Calculate $S_{1,r} = 20\log_{10}\sqrt{\sum_{j\in J_r} A'^2_{r,j}} - 10\log_{10}(f_U - f_L)$.

The output is the time series $(t_{1,r}, S_{1,r}), r = 1, \dots, n_1$.

2.2 STAGE 2

The input to this stage is the output from stage 1. The output is a time series of values of SPSDL representing summary statistics extracted from the distributions of values taken over some aggregation time interval. Examples of summary statistics are $P_1, P_{10}, P_{50}, P_{90}$ and P_{99}, where P_n denotes the nth percentile. Examples of aggregation time intervals include one day, one week, one month and one year, and the choice of aggregation time interval determines the sampling frequency of the output time series, i.e., the sampling frequency is 1 per day for the choice of a 1-day aggregation time interval.

This stage of the data processing chain is described as follows. Let $\left(t_{1,i}, S_{1,i}\right), i = 1, \ldots, n_1$, denote the output from stage 1 comprising a time series of values of SPSDL for a given averaging time interval and a given frequency band. Let ΔT_2 denote the aggregation time interval for this stage 2, and $P_\beta(\boldsymbol{y})$ the function that returns the sample β-percentile of the values in $\boldsymbol{y}$. For example, $P_{50}(\boldsymbol{y})$ returns the sample median of the values in $\boldsymbol{y}$, which is obtained by sorting the values of $\boldsymbol{y}$ into non-decreasing order and selecting the central value if the number of values in $\boldsymbol{y}$ is odd or the average of the two central values if the number is even.

The measurement function is specified as follows. For the rth aggregation time interval defined by start-time $(r-1)\Delta T_2$ and end-time $r\Delta T_2$:

2.1 Identify those points with indices I_r for which $(r-1)\Delta T_2 \leq t_{1,i} < r\Delta T_2$.

2.2 Calculate $t_{2,r} = (r-1)\Delta T_2$.

2.3 Calculate $S_{2,r} = P_\beta(S_{1,i}: i \in I_r)$.

The output is the time series $\left(t_{2,r}, S_{2,r}\right), r = 1, \ldots, n_2$.

2.3 STAGE 3

The input to this stage is the output from stage 2. The output from this stage is the value of an advanced parameter derived from the input having the purpose of meeting the requirement of a specific application or user. An example can be to use the output from stage 2 for the purpose of understanding the characteristics of the underwater noise at the location of the sensor. For example, whether there is a long-term trend in the noise or whether the characteristics of the noise are correlated with some other environmental variable, such as sea surface temperature. The challenge is to undertake an analysis of the time series data, when that analysis might be model-driven or data-driven, to extract information about those characteristics.

2.4 OTHER EXAMPLES

In another example of a data processing chain, stage 1 involves applying the scaling factor, removing the effect of the hydrophone response, and then applying the inverse FFT to return a time series of values of sound pressure in the unit of pascals at the same sampling frequency as the raw measured data. Stage 2 can then be to apply the stage 1 measurement function to each hydroacoustic sensor in a collection of spatially distributed sensors for the purpose of event detection and attribution, i.e., estimating the time and location of an event and correlating that information with known events. Here, an event of interest might be characterised by a particular acoustic signature or by a sound pressure that exceeds a particular threshold. The challenge is to identify when the event is registered by each sensor and to use that information (specifically the differences between those registration times and

the locations of the sensors) to determine the time and location of the event, which constitute the required advanced derived parameters. In this case, knowledge of phase information, and correcting for the phase response of the hydroacoustic sensor, is important for the accurate determination of the times at which events are detected by the different sensors.

3 SOURCES OF UNCERTAINTY

It is assumed that there are three sources of uncertainty that impact the data processing chain described above.

The first is random noise inherent in the sensor, which affects each raw measured count independently. The effect is considered to be modelled as the output of the analogue to digital converter applied to samples drawn randomly and independently from a continuous Gaussian distribution defined by an expectation of zero and a prescribed variance σ_N^2.

The second is the estimation of the scaling factor G. The effect is considered to be modelled by a continuous Gaussian distribution defined by an expectation equal to the estimate $\hat{G}$ of G and a prescribed variance σ_G^2.

The third is the calibration of the sensor. Suppose the calibration is defined by amplitude and phase responses given at discrete frequencies $f_{c,j}, j = 0, \dots, n_C$, and denoted, respectively, by $(f_{c,j}, A_{c,j})$ and $(f_{c,j}, \phi_{c,j}), j = 0, \dots, n_c$. The effect is considered to be modelled by a multivariate Gaussian distribution for each of amplitude and phase defined by an expectation vector comprising the estimates of the response ($\hat{A}_{c,1}, \dots, \hat{A}_{c,n_c}$ for amplitude and $\hat{\phi}_{c,1}, \dots, \hat{\phi}_{c,n_c}$ for phase) and prescribed variance matrices (V_A for amplitude and V_ϕ for phase). This model assumes the calibrations of amplitude and phase are independent, but a more general treatment would consider the effect for both amplitude and phase to be modelled by a single multivariate Gaussian distribution.

These sources of uncertainty will give rise to uncertainty in the values of SPSDL in the time series that constitute the output of stage 1 of the data processing chain and in the values of the summary statistics of SPSDL that constitute the output of stage 2. In this work we consider the evaluation of that uncertainty, and how it propagates through the data processing chain. We note that the second and third of these sources of uncertainty give rise to possible mutual correlation between the values of SPSDL of the summary statistics for those values.

4 METHOD OF UNCERTAINTY PROPAGATION

In this work we use a Monte Carlo (MC) method [4, 5] as a numerical approach for propagating measurement uncertainty through a defined measurement function or through the composition of a sequence of measurement functions. The advantages of the approach are that it does not require the explicit calculation of sensitivity coefficients, and it makes no assumptions about the nature of the measurement functions, i.e., that they are linear or mildly non-linear, or about the nature of the probability distributions for the outputs from those measurement functions, i.e., that they are Gaussian or approximately Gaussian. The disadvantage of the approach, however, is that it can be computationally expensive.

Given a measurement function $\boldsymbol{Y} = f(\boldsymbol{X})$ relating m output quantities $\boldsymbol{Y} = (Y_1, \dots, Y_m)^{\mathrm{T}}$ to N input quantities $\boldsymbol{X} = (X_1, \dots, X_N)^{\mathrm{T}}$ and a (joint) probability density function $g_{\boldsymbol{X}}(\boldsymbol{\xi})$ for $\boldsymbol{X}$, the Monte Carlo method involves repeating the following steps for $r = 1, \dots, M$:

I. Make a random draw $\boldsymbol{x}_r$ for $\boldsymbol{X}$ from the probability density function $g_{\boldsymbol{X}}(\boldsymbol{\xi})$;

II. Evaluate the measurement function to obtain a random draw $\boldsymbol{y}_r = f(\boldsymbol{x}_r)$ for Y from its (unknown) probability density function.

The (sample) average $\overline{\boldsymbol{y}}$ and (sample) covariance matrix $V_{\boldsymbol{Y}}$ of the samples $\boldsymbol{y}_r, r = 1, \ldots, M$, provide, respectively, an estimate of $\boldsymbol{Y}$ and its uncertainty.
For the problem treated in this work, one consideration is the computational effort involved that is dominated by the time to evaluate the measurement function representing stages 1 and 2 of the data processing chain. The required computational effort can limit the number M of Monte Carlo trials that can be performed in practice and, in turn, the reliability of the results that are obtained from the calculation. A further consideration is when the number m of output quantities is large, which is the case for the output of stage 1 of the data processing chain. In that case, it can be difficult to store the complete covariance matrix for those quantities, and it can be necessary to store the information captured by the covariance matrix in a compressed (and approximate) form (see, for example, [6]).

5 DATA

We consider data recorded by the hydroacoustic sensor H01W1, the first hydrophone at the IMS station west of Cape Leeuwin. Figure 1 illustrates the inputs to and outputs from stages 1 and 2 in the data processing chain described above. The top graph shows the input to stage 1 comprising a time series of raw measured data in counts over four days. The middle graph shows the output from stage 1 comprising a time series of values of SPSDL for an averaging time interval of 1 min and the frequency band 10 Hz ─ 40 Hz. The bottom graph shows the output from stage 2 comprising values of the summary statistics $P_1, P_{10}, P_{50}, P_{90}$ and P_{99} for an aggregation time interval of 1 day, and there are twenty such parameters.

6 RESULTS FORM STAGE 2 OUTPUTS

The results described in the section are each obtained using a Monte Carlo method and $M = 100$ trials. An “updating approach” is used to update the values of the estimates of the stage 2 outputs, as well as the covariance matrix for those estimates, which does not depend on storing the values of the stage 2 outputs for all the Monte Carlo trials [7]. Such an approach would become necessary as the number of stage 2 outputs and/or the number of Monte Carlo trials are increased.

Figure 2 shows results that summarise the influence of signal noise alone. The top graph shows the standard uncertainty for each of the stage 2 outputs, which comprise the summary statistics for SPDL for each of the four days of observation, as a function of signal-to-noise ratio σ_N (in percent), and the bottom graph shows the correlation matrix for the outputs corresponding to the largest value of σ_N. Similarly, Figure 3 shows results that summarise the influence of the scale factor alone as a function of the relative standard deviation σ_G (in percent), and Figure 4 considers the influence of the calibration alone. In this case, σ_A (in percent) denotes the relative standard deviation of the amplitudes in the frequency response with no uncertainty for the phases in that response.

We note that:

- Signal noise does not lead to correlation between the stage 2 outputs. The uncertainties of the P_{99} estimates are larger than the uncertainties for the other summary statistics. They vary proportionately with σ_N, whereas the uncertainties for the other summary statistics are essentially invariant to σ_N.

- The effect of the scaling factor is to produce perfect (positive) correlation between the stage 2 outputs. For a given choice of σ_G, there is no difference between the uncertainties for the outputs, and those uncertainties vary proportionately with σ_G.

- The effect of the calibration uncertainty (restricted to uncertainty in the amplitude response only) is to produce high correlation between the stage 2 outputs. For a given choice of σ_A, there are only small differences between the uncertainties for the outputs, and those uncertainties vary proportionately with σ_A.

Finally, Figure 5 illustrates results that summarise the influence of the combination of signal noise, scaling factor and calibration. The graph shows the correlation matrix for the stage 2 outputs corresponding to the largest values of σ_N, σ_G and σ_A. In this case, the correlations between the outputs are generally high except, most notably, for the P_{99} estimates and, to a lesser extent, for the P_1 estimates. We expect the influence of signal noise to dominate for the P_{99} estimates, and the observed behaviour then mimics that illustrated in Figure 2.

7 COMPACT REPRESENTATION OF THE UNCERTAINTY INFORMATION

In the examples considered above, for which there are only twenty stage 2 outputs, it is straightforward to store and work with the covariance matrix for those outputs. However, a challenge is to do so when the observation period lasts for many years, and the size of the covariance matrix becomes prohibitively large to store in full. For example, the hydrophone H01W1 has been recoding data for more than 15 years. Then, considering an aggregation time interval of 1 day for an observation period of 15 years and all five summary statistics would generate 27,375 stage 2 outputs, and a covariance matrix of 749,390,625 elements. Ignoring the symmetry of the matrix, i.e., storing it as a full matrix, and assuming each matrix element is stored as a 64-bit double, the matrix would require of the order of 5.5 GB of memory. However, the results from Section 6 suggest that the correlation matrix is structured and, consequently, there is the possibility to exploit that structure to provide a compact, albeit approximate, representation of the matrix.

We consider the case $\sigma_N = 0.16\ \%$, $\sigma_G = 0.1\ \%$ and $\sigma_A = 0.35\ \%$ and run three Monte Carlo calculations with $M = 1000$ as follows:

(MC/A) considering the influence of signal noise only, which produces uncorrelated effects in the stage 2 outputs,

(MC/B) considering scaling factor uncertainty and calibration uncertainty only, which produce correlated effects in the stage 2 outputs, and

(MC/C) considering the influences of signal noise, scaling factor uncertainty and calibration uncertainty.

To illustrate the output of the Monte Carlo calculation, Figure 6 shows the approximations to the probability density functions for the five stage 2 outputs (statistical percentiles) for the last aggregation period (day 4). The distributions are well-separated, and closer inspection shows that they appear “Gaussian” in their form.

Table 1 compares the estimates of the stage 2 outputs obtained from the three Monte Carlo calculations as well as directly in terms of the input data to stage 1 in the data processing pipeline (denoted by “DATA”). We note that there is agreement between the estimates for the methods DATA and MC/B, and between the estimates for MC/A and MC/C, with the estimates for the latter pair of methods being consistently greater than those for the former

pair. The methods MC/A and MC/C include the effect of signal noise, which in this case is interpreted through the summary statistics as an additional source of sound.

Let V_N, $V_{G,A}$ and V be the covariance matrices obtained from the three Monte Carlo calculations specified above. We define an approximation to V by

$$V' = D + \sigma_{G,A}^2 \mathbf{1},$$

where D is the diagonal matrix obtained by setting all off-diagonal elements in V_N to zero, $\sigma_{G,A}^2$ is the average of the elements of $V_{G,A}$, and $\mathbf{1}$ is a square matrix of dimension equal to the number of stage 2 outputs with all its elements set to one. The specification of V' is based on two assumptions: firstly, that the uncertainties associated with those influences giving rise to uncorrelated and correlated effects in the stage 2 outputs can be propagated separately and their variances combined additively; secondly, that the uncertainty information for those uncorrelated and correlated effects can be represented by covariance matrices having simple structures, viz, as a diagonal matrix and as a multiple of a unit matrix, respectively. For the example used as illustration above, this matrix is specified by 27,375 + 1 values if those elements known to be exactly zero or one are ignored. Figure 7 and Figure 8 show the correlation matrices obtained from V and V', respectively, which appear to have a very similar structure. The effect of replacing V by V' on the uncertainty associated with the estimate of a derived stage 3 output is investigated in the following section.

In practice and supported by the results from Section 6, we expect the elements of V_N and $V_{G,A}$ to depend only weakly on the period (or day over which the values of SPSDL are aggregated) but more strongly on the summary statistic. In this case, D can be defined approximately by five values, one for each summary statistic that can be read from pre-calculated graphs such as those shown in Figure 2. Furthermore, V' can be constructed from the covariance matrices V_N and $V_{G,A}$ evaluated for a (much) shorter observation period. In these ways, both the calculation and storage of V' can be made very efficient. Figure 12 Influence of noise factor alone: standard uncertainty for each of the stage 2 outputs, which comprise the summary statistics for SPDL for the four aggregation periods, as a function of signal to noise ratio (top), and the correlation matrix for the outputs corresponding to the largest value of signal noise (bottom).

8 IMPACT ON FURTHER ADVANCED DERIVED PARAMETERS

For the summary statistic P_β, let the stage 2 outputs comprise the data $(t_i, P_{\beta,i}), i = 1, \ldots, m$. In the examples considered before, β is 1, 10, 50, 90 or 99, $m = 4$, and t_i is 0, 1, 2 and 3. We define the stage 3 outputs to be the value a_β at $t_c = 1.5$ (the average of the values t_i), and the slope b_β, of a straight-line function fitted to the data by ordinary least-squares. The stage 3 output a_β provides a single representative value of P_β and the stage 3 output b_β the rate of change of P_β with respect to time (or its trend).

Define

$$\boldsymbol{c}_\beta = \begin{pmatrix} a_\beta \\ b_\beta \end{pmatrix}, \qquad \boldsymbol{c} = \begin{pmatrix} \boldsymbol{c}_1 \\ \vdots \\ \boldsymbol{c}_{99} \end{pmatrix}, \qquad \boldsymbol{p}_\beta = \begin{pmatrix} P_{\beta,1} \\ \vdots \\ P_{\beta,m} \end{pmatrix}, \qquad \boldsymbol{p} = \begin{pmatrix} \boldsymbol{p}_1 \\ \vdots \\ \boldsymbol{p}_{99} \end{pmatrix}, \qquad \boldsymbol{X}_\beta = \begin{pmatrix} 1 & t_1 - t_c \\ \vdots & \vdots \\ 1 & t_m - t_c \end{pmatrix},$$

and

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \boldsymbol{X}_{99} \end{pmatrix}.$$

Then, formally, the estimates $\boldsymbol{c}$ of the stage 3 outputs are given by

$$\boldsymbol{c} = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{p},$$

which defines a linear relationship between $\boldsymbol{p}$ and $\boldsymbol{c}$. Consequently, using a generalised form of the law of propagation of uncertainty (LPU) [5], the covariance matrices V_p and V_c for $\boldsymbol{p}$ and $\boldsymbol{c}$, respectively, are related by

$$V_c = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}V_p\boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}.$$

(In practice, and to avoid loss of numerical precision, the above calculations are undertaken in terms of an orthogonal factorisation of the design matrix $\boldsymbol{X}$.)

Let $\boldsymbol{p}$ be the estimates of the stage 2 outputs calculated directly in terms of the input data to stage 1 in the data processing pipeline (denoted by "DATA" in Table 1). We evaluate the stage 3 outputs in the following ways:

(MC) by running a Monte Carlo calculation considering the influences of signal noise, scaling factor uncertainty and calibration uncertainty,

(LPU/a) using the formal expressions for $\boldsymbol{c}$ and V_c given above with $V_p = V$ from Section 7,

(LPU/b) using the formal expressions for $\boldsymbol{c}$ and V_c given above with $V_p = V_N$ from Section 7, and

(LPU/c) using the formal expressions for $\boldsymbol{c}$ and V_c given above with $V_p = V'$ from Section 7.

Table 2 compares the estimates of the stage 3 outputs obtained in these different ways as well as directly in terms of the input data to stage 1 in the data processing pipeline (denoted by "DATA"). Figure 9 shows together this further stage of the data processing pipeline in terms of the outputs from stage 2, which are used as the inputs to stage 3, and the outputs of stage 3 displayed as trend-lines for each of the summary statistics. Four of those statistics shows an increasing trend whereas P_{99} shows a decreasing trend. We note that:

- As expected, the estimates of the stage 3 outputs from the methods DATA, LPU/a, LPU/b and LPU/c are identical because all the methods are implementing the same formulae for evaluating the stage 3 outputs in terms of the stage 2 outputs.

- However, the estimates obtained from the method MC are different from those obtained from the other methods. In Section 7 we saw that the stage 2 outputs obtained from a Monte Carlo method that considered the influence of signal were biased high, and this effects filters through to the estimation of the stage 3 outputs.

Table 3 compares the uncertainties associated with the estimates obtained in the different ways in terms of the standard uncertainties $u(a_\beta)$ and $u(b_\beta)$ associated with the pair of stage 3 outputs for stage 2 outputs P_β and their correlation coefficient $r(a_\beta, b_\beta)$. (The correlation

coefficients for stage 3 outputs corresponding to different summary statistics are not presented because they are considered to be of less importance and interest.) We note that:

- The standard uncertainty $u(a_\beta)$ obtained using the methods MC, LPU/a, and LPU/c are consistent, and the uncertainty obtained using the method LPU/b is underestimated. For this stage 3 output, the contributions to the uncertainty from the scale factor uncertainty and calibration uncertainty, which produce correlated effects in the stage 2 outputs, cannot be neglected.

- The standard uncertainty $u(b_\beta)$ obtained using all the methods are consistent. For this stage 3 output, the contributions to the uncertainty from the scale factor uncertainty and calibration uncertainty, which produce correlated effects in the stage 2 outputs, can be neglected.

- The compact, albeit approximate, representation V' of V would seem to be adequate for the evaluation of the uncertainties for the stage 3 outputs.

- The correlation coefficients are also close to zero, which indicates that the a_β and b_β are weakly correlated, which is a consequence of the chosen parametrisation of the straight-line trend function, viz., the parameter a_β is chosen to be the value of the function at the mean of the values $t_i, i = 1, \ldots, m$.

9 SUMMARY

We have investigated for a particular user application how uncertainty propagates through a data processing pipeline that starts with the raw data recorded by a sensor and finishes with advanced derived parameters which convey useful information. The application concerns raw data recorded by a hydroacoustic sensor taken from the International Monitoring System of the CTBT and parameters describing characteristics of summary statistics of the distributions of values of SPSDL derived from the raw data. In this application we find that the contribution from a given source of uncertainty depends on the nature of the source (e.g., signal noise compared to calibration uncertainty) and the nature of the parameter (e.g., a representative value of SPSDL compared to a temporal trend in SPSDL values). We also found that it was possible to construct compact, albeit approximate, representations of the uncertainty information at intermediary stages in the data processing pipeline that meant that the propagation of uncertainty could be done efficiently. It can be expected that these results are specific to the application, but it is hoped that the approach taken in the investigation can be used as a template for other applications.

10 ACKNOWLEDGEMENTS

This work has been undertaken within the Joint Research Project 19ENV03 “Infra-AUV” of the European Metrology Programme for Innovation and Research (EMPIR). The EMPIR is jointly funded by the EMPIR participating countries within EURAMET and the European Union. The views expressed in the paper are those of the authors and do not necessarily represent those of the CTBTO.

11 REFERENCES

[1] Joint Research Project 19ENV03 “Infra-AUV” Metrology for low-frequency sound and vibration https://www.ptb.de/empir2020/infra-auv/home/

[2] CTBTO, available online: www.ctbto.org.

[3] Harris, P., Sotirakopoulos, K., Robinson, S., Wang, L., and Livina V. A statistical method for the evaluation of long-term trends in underwater noise measurements, *J. Acoust. Soc. Am*. **145**, 228–241 (2019) https://doi.org/10.1121/1.5084040.

[4] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Supplement 1 to the 'Guide to the Expression of Uncertainty in Measurement' – Propagation of distributions using a Monte Carlo method, JCGM 101:2008. BIPM, 2008.

[5] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Supplement 2 to the 'Guide to the Expression of Uncertainty in Measurement' – Extension to any number of output quantities, JCGM 102:2011. BIPM, 2011.

[6] Humphreys, D.A., Harris, P.M., Rodríguez-Higuero, M., Mubarak, F.A., Zhao, D., and Ojasalo, K. Principal Component Compression Method for Covariance Matrices Used for Uncertainty Propagation. *IEEE Transactions on Instrumentation and Measurement* **64**, 2015.

[7] Cox, M.G., Harris, P.M., and Smith, I.M. Software specifications for uncertainty evaluation. NPL Report MS 7, 2010.

12 FIGURES AND TABLES

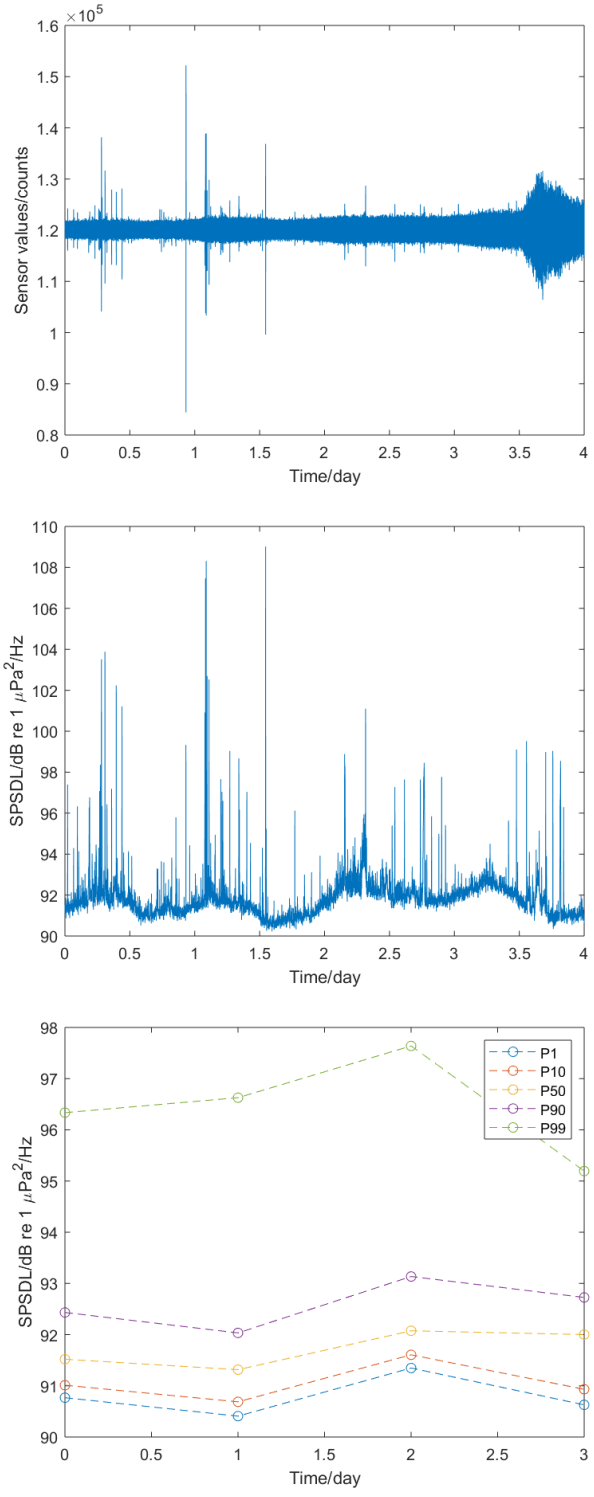

Figure 1 Stages in data processing chain: input to stage 1 (top), output of stage 1 and input to stage 2 (middle), and output of stage 2. (The dotted lines joining the daily values are included for purposes of visualisation only.)

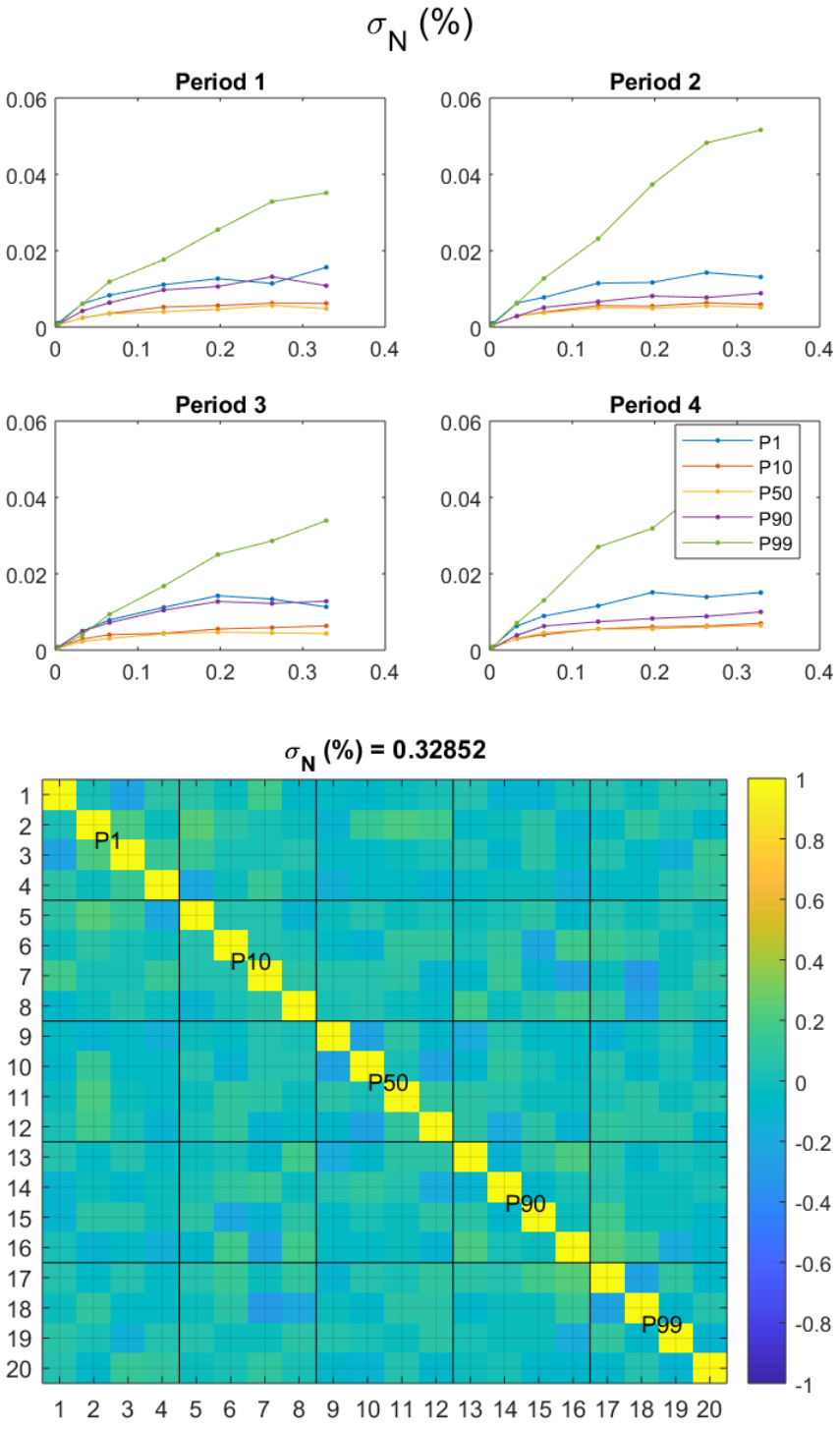

Figure 2 Influence of noise factor alone: standard uncertainty for each of the stage 2 outputs, which comprise the summary statistics for SPDL for the four aggregation periods, as a function of signal to noise ratio (top), and the correlation matrix for the outputs corresponding to the largest value of signal noise (bottom).

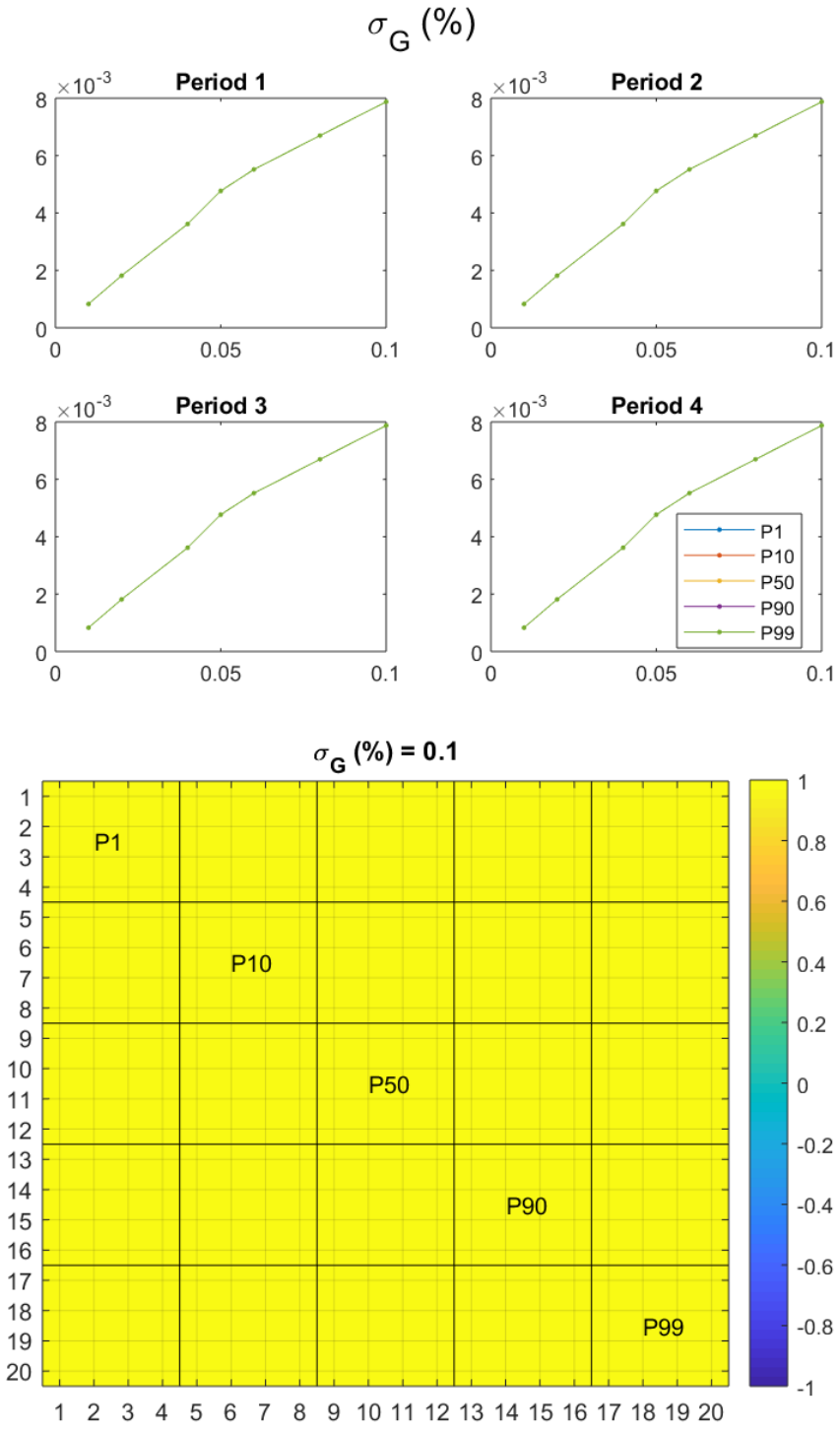

Figure 3 Influence of scaling factor alone: standard uncertainty for each of the stage 2 outputs, which comprise the summary statistics for SPDL for the four aggregation periods, as a function of the relative standard deviation of the scaling factor (top), and the correlation matrix for the outputs corresponding to the largest value of scaling factor uncertainty (bottom).

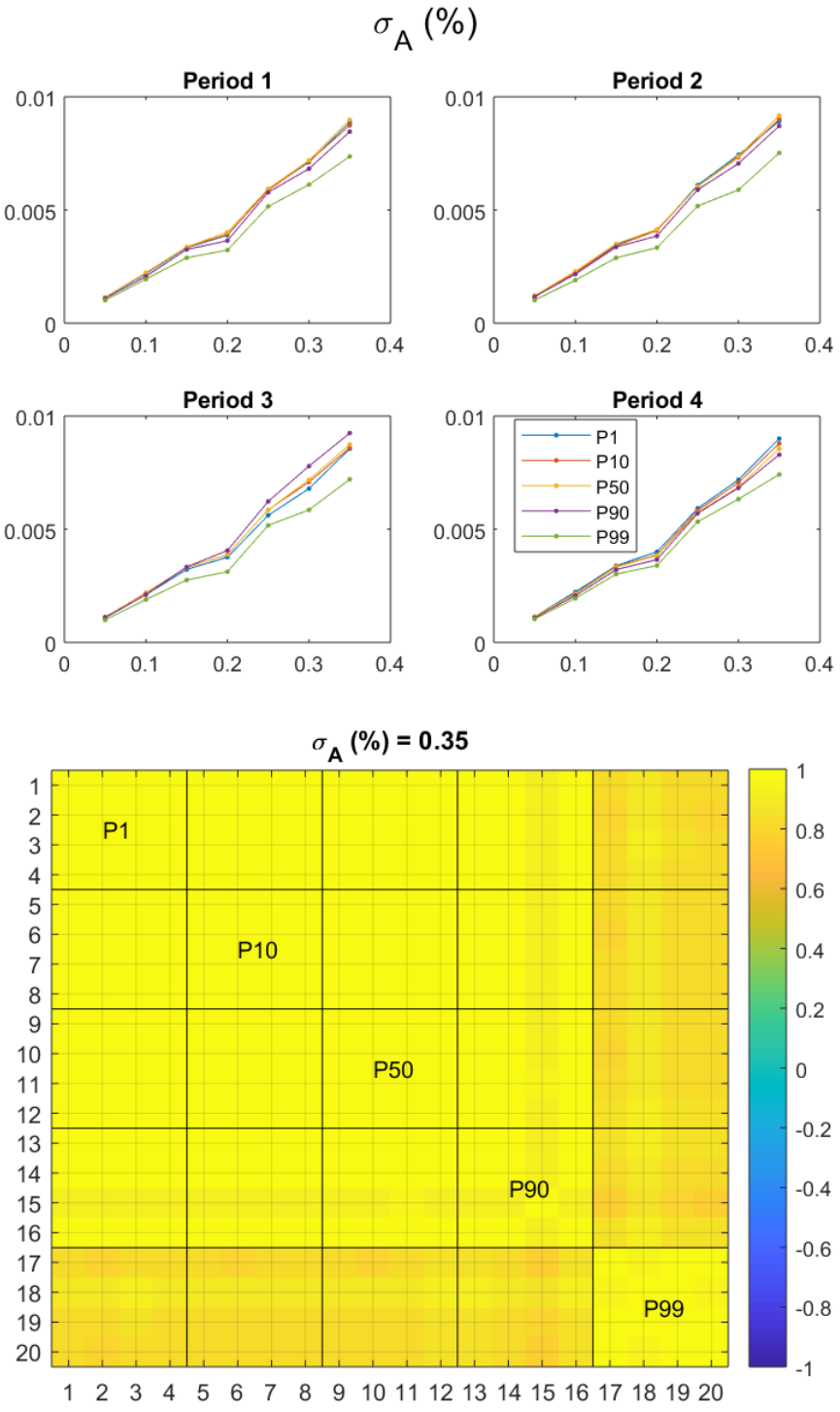

Figure 4 Influence of calibration alone: standard uncertainty for each of the stage 2 outputs, which comprise the summary statistics for SPDL for the four aggregation periods, as a function of relative standard deviation of the amplitudes in the frequency response (top), and the correlation matrix for the outputs corresponding to the largest value of calibration uncertainty (bottom).

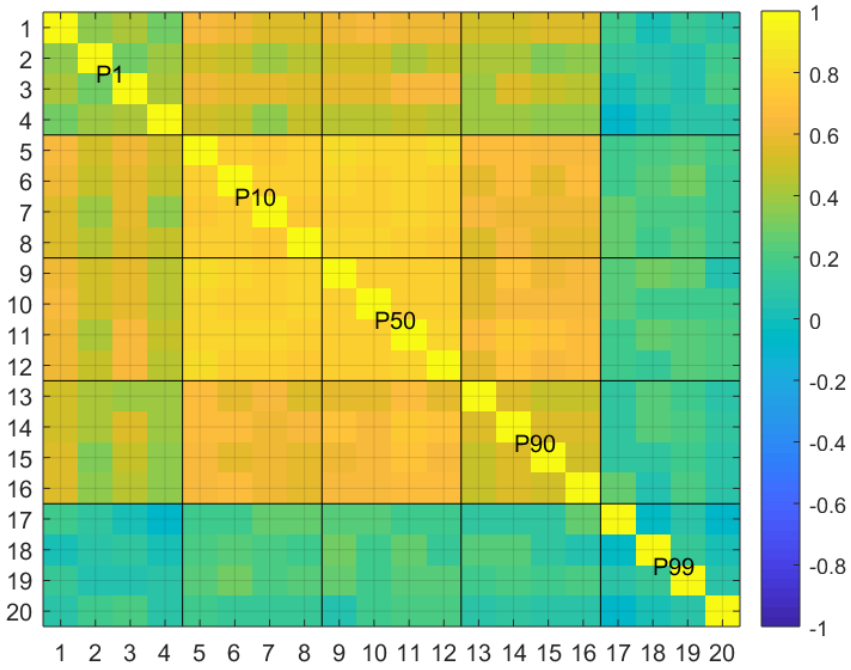

Figure 5 Combination of influences: the correlation matrix for the stage 2 outputs corresponding to the largest values of signal to noise ratio, scaling factor uncertainty and calibration uncertainty.

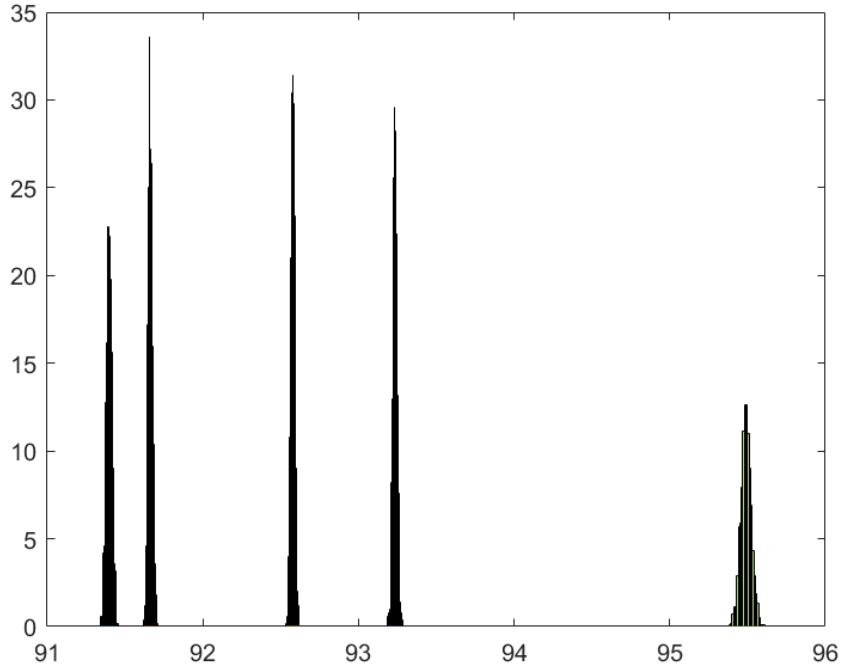

Figure 6 Combination of influences: approximations to the probability density functions for the five stage 2 outputs (statistical percentiles) for the last aggregation period obtained from a Monte Carlo calculation.

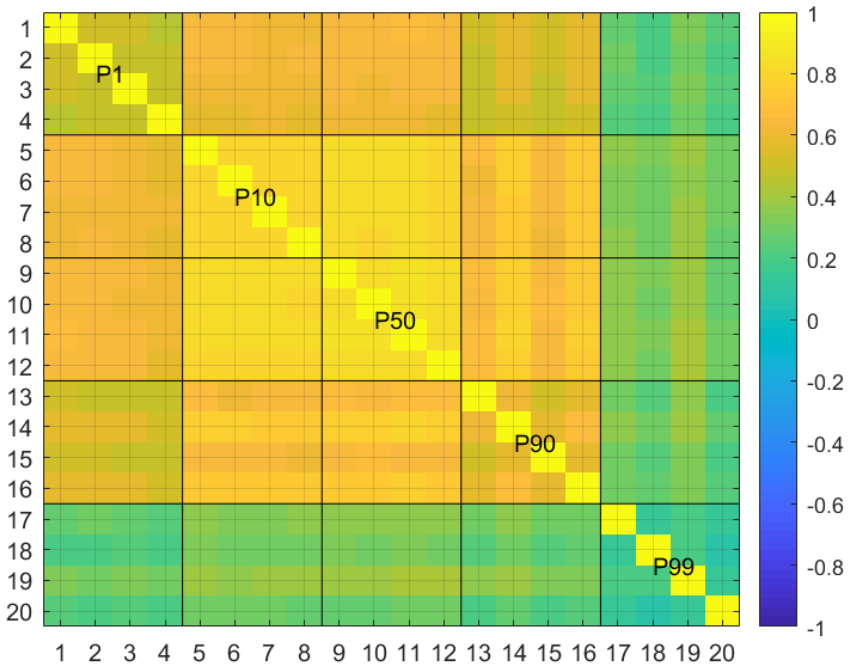

Figure 7 Combination of influences: correlation matrix for the stage 2 outputs accounting for the influence factors of signal noise, ratio, scaling factor and calibration.

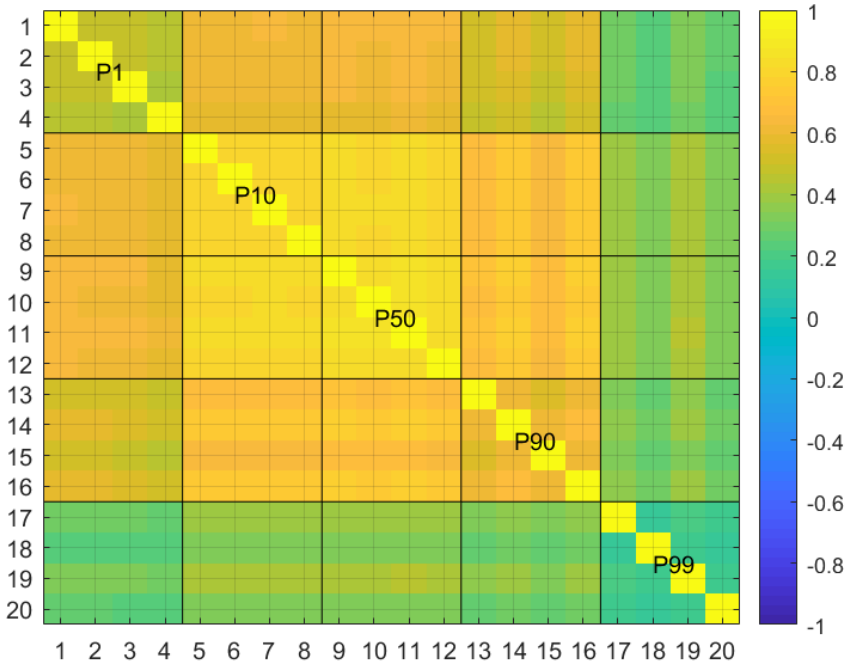

Figure 8 Combination of influences: an approximation to the correlation matrix for the stage 2 outputs accounting for the influence factors of signal noise, ratio, scaling factor and calibration.

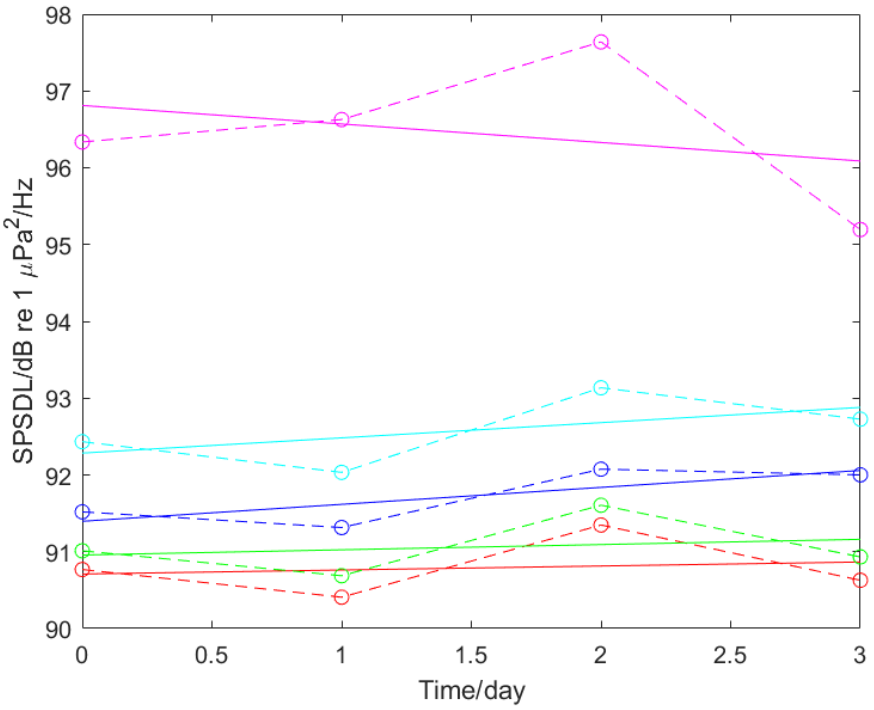

Figure 9 A further stage in the data processing chain: output of stage 2 and input to stage 3 (circles joined by dotted lines), and output of stage 3 (solid lines).

Table 1 Estimates of stage 2 outputs.

PERCENTILE	PERIOD	DATA	MC/A	MC/B	MC/C
P_1	1	90.7683	91.5096	90.7682	91.5091
	2	90.4079	91.2155	90.4076	91.2155
	3	91.3478	92.0013	91.3476	92.0019
	4	90.6284	91.3979	90.6284	91.3978
P_{10}	1	91.0117	91.7364	91.0116	91.7362
	2	90.6873	91.4558	90.6871	91.4555
	3	91.6058	92.2382	91.6058	92.2382
	4	90.9325	91.6625	90.9323	91.6620
P_{50}	1	91.5191	92.1681	91.5191	92.1679
	2	91.3151	91.9934	91.3145	91.9936
	3	92.0750	92.6559	92.0743	92.6559
	4	92.0009	92.5784	92.0007	92.5780
P_{90}	1	92.4327	92.9630	92.4326	92.9627
	2	92.0334	92.6133	92.0332	92.6133
	3	93.1352	93.5982	93.1349	93.5980
	4	92.7263	93.2353	92.7263	93.2348
P_{99}	1	96.3358	96.5553	96.3357	96.5557
	2	96.6281	96.8377	96.6281	96.8371
	3	97.6392	97.7947	97.6390	97.7935
	4	95.1949	95.4897	95.1949	95.4908

Table 2 Estimates of stage 3 outputs.

PERCENTILE P_β	METHOD	a_β	b_β
P_1	DATA	90.7881	0.0520
	MC	91.5311	0.0453
	LPU/a	90.7881	0.0520
	LPU/b	90.7881	0.0520
	LPU/c	90.7881	0.0520
P_{10}	DATA	91.0593	0.0681
	MC	91.7730	0.0560
	LPU/a	91.0593	0.0681
	LPU/b	91.0593	0.0681
	LPU/c	91.0593	0.0681
P_{50}	DATA	91.7275	0.2205
	MC	92.3489	0.1893
	LPU/a	91.7275	0.2205
	LPU/b	91.7275	0.2205
	LPU/c	91.7275	0.2205
P_{90}	DATA	92.5819	0.1983
	MC	93.1022	0.1801
	LPU/a	92.5819	0.1983
	LPU/b	92.5819	0.1983
	LPU/c	92.5819	0.1983
P_{99}	DATA	96.4495	-0.2412
	MC	96.6693	-0.2238
	LPU/a	96.4495	-0.2412
	LPU/b	96.4495	-0.2412
	LPU/c	96.4495	-0.2412

Table 3 Uncertainties associated with estimates of stage 3 outputs.

PERCENTILE P_β	METHOD	$u(a_\beta)$	$u(b_\beta)$	$r(a_\beta, b_\beta)$
P_1	MC	0.0132	0.0057	0.0226
	LPU/a	0.0132	0.0057	0.0226
	LPU/b	0.0062	0.0057	0.1278
	LPU/c	0.0129	0.0056	0.0539
P_{10}	MC	0.0119	0.0026	0.0670
	LPU/a	0.0119	0.0026	0.0670
	LPU/b	0.0028	0.0027	0.0206
	LPU/c	0.0117	0.0026	0.0033
P_{50}	MC	0.0118	0.0023	-0.0203
	LPU/a	0.0118	0.0023	-0.0203
	LPU/b	0.0024	0.0023	0.0765
	LPU/c	0.0116	0.0023	0.0147
P_{90}	MC	0.0125	0.0043	-0.0111
	LPU/a	0.0125	0.0043	-0.0111
	LPU/b	0.0047	0.0040	-0.1288
	LPU/c	0.0123	0.0041	-0.0357
P_{99}	MC	0.0171	0.0124	0.0846
	LPU/a	0.0171	0.0124	0.0846
	LPU/b	0.0127	0.0118	0.0660
	LPU/c	0.0173	0.0118	0.0479