Harmonisation Input File Format Definition

Sam Hunt, Ralf Quast, Jon Mittaz, Peter Harris, Emma Woolliams, Ralf Giering & Arta Dilo

National Physical Laboratory, Fast Opt & University of Reading

11/28/2017

FIDUCEO has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no. 638822

Contents

1 Introduction

Harmonisation is the process within the FIDUCEO project by which the project's Fundamental Climate Data Record products of sensor series Level 1 data will be recalibrated by processing the match-up data between the sensors in the series as well as with an external, well calibrated reference sensor.

Two algorithms have been developed within the project to perform this processing which aim to respect the full error correlation structure in the match-up series. The aim of this document is to describe what data is required from the sensor teams to achieve this and how it should be arranged into input netCDF files.

2 Data Structures

This section describes what sensor series match-up data is required for harmonisation and the structure of how that data should be stored. How that data maps into the required input file format is described in the following section.

2.1 Match-Up Data

For a set of N_s sensors each given sensor, labelled k, has a measurement equation,

$$L_k = f\left(X_{k,1}, \ldots, X_{k,m_k};\ a_{k,1}, \ldots, a_{k,N_a}\right),$$

where L_k is the sensor measurand, X_{k,n_X} are the quantities observed by sensor k (with $n_X \epsilon [1, m_k]$), and $a_{k,1}, \ldots, a_{k,N_a}$ are the N_a calibration parameters, to be harmonised.

Sensor-Sensor Match Up Series Data

For match-ups between two sensors labelled $k = 1$ and $k = 2$, we then have data for observed quantities $\underline{X}_{1,n_X}$ and $\underline{X}_{2,n_X}$, where the vector labelling indicates that we have observations for a series of match-ups. We arrange these data for storage in an arrays, *X1* and *X2*, as,

$$X1 = [\underline{X}_{1,1}, \ldots, \underline{X}_{1,m_1}];$$

$$X2 = [\underline{X}_{2,1}, \ldots, \underline{X}_{2,m_2}];$$

where the width of *X1* is m_1 and the width of *X2* is m_2. The number of rows for both is *M*, the number of matchups.

Reference Sensor Match-up Series Data

In harmonisation we also have match-ups between a sensor series sensor, labelled $k = 2$, and a reference sensor, labelled $k = 1$, which is not to be harmonised. Here in terms of data we have the determined reference sensor measurand, $\underline{L}_{ref}$, and $\underline{X}_{2,n_X}$ as before, with associated uncertainties. We arrange these data into *X1* and *X2* arrays with the same dimensions as for the sensor-sensor case, e.g. the $\underline{L}_{\mathrm{ref}}$ data fill the *X1* array with the length of the m_1 dimension 1.

Adjustment Factor, K

In the harmonisation process, we define the expected difference between the observed measurand of sensor 1 and 2 caused by differences between sensor spectral response functions and the match-up process (e.g. observation temporal, geo-location difference), as,

$$K = L_2 - L_1$$

An estimation of this difference should be stored per match-up as a vector $\underline{K}$ with length M.

Match-Up Time

The time of each match-up recorded by each sensor (which may not necessarily be the same) should be stored in two arrays of length M named *time1* and *time2,* for sensor 1 and sensor 2 respectively. This time should be stored as seconds since 1/1/1970.

2.2 Match-up Data Uncertainties

2.2.1 Observed Quantity Uncertainties

For every observed quantity, $\underline{X}_{k,n_X}$, in the *X1* and *X2* arrays uncertainty data must be provided, which also defines the error correlation structure between the measurements of a given observed quantity (it is not possible to define error correlations between measurements of different observed quantities). There are four classes of observed quantity error correlation structures it is possible to define within the files:

- Independent Error Correlation – Errors in the measurements of $\underline{X}_{k,n_X}$ are uncorrelated or random.
- Independent + Systematic Error Correlation – In addition to random errors, measurements of $\underline{X}_{k,n_X}$ have an error that is fully correlated or common between all measurements.
- Structured Correlation – Other intermediate error correlation structure for measurements of $\underline{X}_{k,n_X}$.
- Structured + Systematic Error Correlation – In addition to structured errors, measurements of $\underline{X}_{k,n_X}$ have an error that is fully correlated or common between all measurements.

This sections details how to assign an observed quantity a given error correlation structure, and what uncertainty data is required in each case.

2.2.1.1 Independent Measurement Error Correlation

Each datum in both *X1* and *X2* may have an associated random component of uncertainty (i.e. there is no correlation between the errors of each match-up for a given observed quantity). The covariance matrix of a series of measurements of a given observed parameter in this case is,

$$V(\underline{X}_{k,n_X}) = U_r(\underline{X}_{k,n_X})U_r^{\mathrm{T}}(\underline{X}_{k,n_X}),$$

where $U_r(\underline{X}_{k,n_X}) = \mathrm{diag}[\underline{u}_r(\underline{X}_{k,n_X})]$, with $\underline{u}_r(\underline{X}_{k,n_X})$ the vector of uncertainties in $\underline{X}_{k,n_X}$ due to the independent error effects.

These data should be stored in the same format as *X1* and *X2* in arrays *Ur1 and Ur2* as,

$$Ur1 = [\underline{u}_r(\underline{X}_{1,1}), \dots, \underline{u}_r(\underline{X}_{1,m_1})]\ ;$$

$$Ur2 = [\underline{u}_r(\underline{X}_{2,1}), \ldots, \underline{u}_r(\underline{X}_{2,m_2})] \,.$$

2.2.1.2 Systematic Measurement Error Correlation

Each datum in both *X1* and *X2* may also have an additional fully systematic component of uncertainty (i.e. the error of each match-up is fully correlated for a given observed quantity). The covariance matrix of a series of measurements of a given observed parameter in this case is,

$$V(\underline{X}_{k,n_X}) = U_r(\underline{X}_{k,n_X})U_r^{\mathrm{T}}(\underline{X}_{k,n_X}) + \underline{u}_s(\underline{X}_{k,n_X})\underline{u}_s^{\mathrm{T}}(\underline{X}_{k,n_X}),$$

where $\underline{u}_s(\underline{X}_{k,n_X})$ is the vector systematic uncertainties of the observed quantities $\underline{X}_{k,n_X}$.

Typically, $\underline{u}_s(\underline{X}_{k,n_X})$ will be single valued. This is the common case where a given measurement in a time series, $X_{k,n_X}(t)$, is subject to a systematic error and can be written as,

$$X_{k,n_X}(t) = X_{k,n_X}^{\mathrm{r}}(t) + X_{k,n_X}^{\mathrm{sys}} \,,$$

where $X_{k,n_X}^{\mathrm{r}}(t)$ is the unknown true value plus a random error and X_{k,n_X}^{sys} is the systematic error, which is shared by all measurements in the time series. The systematic uncertainties for this series of measurements, $\underline{X}_{k,n_X}$, is then,

$$\underline{u}_s(\underline{X}_{k,n_X}) = u\left(X_{k,n_X}^{\mathrm{sys}}\right)\underline{j},$$

where $\underline{j}$ is a vector of length *M* with all elements equal to 1 and $u(X_{k,n_X}^{\mathrm{sys}})$ is the uncertainty in the value of the common systematic error.

In some more complex cases $\underline{u}_s(\underline{X}_{k,n_X})$ might not be single valued. For example, take a case where each measurement of the observed quantity in a time series $X_{k,n_X}(t)$ is itself derived from an underlying observation of some other quantity, $q(t)$, as, $X_{k,n_X}(t) = f(q(t))$. If the all the measurements in time series, $\underline{q}$, are subject to a common systematic error as,

$$q(t) = q^{\mathrm{r}}(t) + q^{\mathrm{sys}},$$

they have associated systematic uncertainties as before,

$$\underline{u}_s\left(\underline{q}\right) = u(q^{\mathrm{sys}})\underline{j}.$$

This then propagates to the measurements in $\underline{X}_{k,n_X}$ as,

$$\underline{u}_s(\underline{X}_{k,n_X}) = u(q^{\mathrm{sys}})\underline{c}_{X_{k,n_X}},$$

where $\underline{c}_{X_{k,n_X}}$ is the vector of sensitivities of $\underline{X}_{k,n_X}$ to $\underline{q}$,

$$\underline{c}_{X_{k,n_X}} = \frac{\partial \underline{X}_{k,n_X}}{\partial \underline{q}}.$$

In the FIDUCEO project an example of this would be the internal calibration target radiance measurements, $\underline{L}_{\mathrm{ICT}}$, which appears in the measurement equation of the AVHRR. $\underline{L}_{\mathrm{ICT}}$ is itself derived from a set internal calibration target temperature PRT measurements, $\underline{T}_{\mathrm{PRT}}$, which are subject to a systematic error. So the systematic uncertainty in the of $\underline{L}_{\mathrm{ICT}}$ is given as,

$$\underline{u}_s(\underline{L}_{\mathrm{ICT}}) = u(T_{\mathrm{PRT}}^{\mathrm{sys}})\underline{c}_{\underline{L}_{\mathrm{ICT}}}, \quad \text{with } \underline{c}_{\underline{L}_{\mathrm{ICT}}} = \frac{\partial \underline{L}_{\mathrm{ICT}}}{\partial \underline{T}_{\mathrm{PRT}}}$$

These data should be stored in the same format as *X1, X2, Ur1* and *Ur2* in arrays *Us1* and *Us2* as,

$$Us1 = [\underline{u}_s(\underline{X}_{1,1}), \ldots, \underline{u}_s(\underline{X}_{1,m_1})] \; ;$$

$$Us2 = [\underline{u}_s(\underline{X}_{2,1}), \ldots, \underline{u}_s(\underline{X}_{2,m_2})].$$

2.2.1.3 Structured Error Correlation Contributions

This section explains how to extend this to more complex error correlation structures. This involves introducing the concept of *W* matrices, which is first described below with examples. This is followed an explanation of how to store this data in the harmonisation input files in a sparse CSR representation.

Expressing Correlation with W Matrices

In a series of measurements over time, t, sensor k determines measurands $L_k(t = t_1)$ to $L_k(t = t_N)$. In general, the error of a given sensor state variable in the $L_k(t_a)$ measurement equation, $X_{k,n_X}(t_a)$, may not be metrologically independent to that of the corresponding sensor state variable in the $L_k(t_b)$ measurement equation, $X_{k,n_X}(t_b)$. For example, they may share a common systematic error component.

In general, the series of error-correlated observed quantities, $\underline{X}_{k,n_X}$, can be related to a set of error-independent observed quantity $\underline{\tilde{X}}_{k,n_X}$ by the transformation,

$$\underline{X}_{k,n_X} = W_{k,n_X}\underline{\tilde{X}}_{k,n_X}$$

For a measurement series there will be m_k such equations, featuring a set of *W* matrices $W_{k,1}$ to W_{k,m_k}, one for each sensor state variable in the measurement equation of sensor k. The *W* matrices here encode the full information of the correlation for the set of parameters in the full series of measurements.

This allows one to build a covariance matrix for $\underline{X}_{k,n_X}$ of,

$$V(\underline{X}_{k,n_X}) = W_{k,n_X} U(\underline{\tilde{X}}_{k,n_X})U^{\mathrm{T}}(\underline{\tilde{X}}_{k,n_X})W_{k,n_X}^{\mathrm{T}},$$

where $U(\underline{\tilde{X}}_{1/2,n_X}) = \mathrm{diag}[u(\underline{\tilde{X}}_{1/2,n_X})]$ with $u(\underline{\tilde{X}}_{1/2,n_X})$ the uncertainties of the error independent set of values $\underline{\tilde{X}}_{1/2,n_X}$.

<u>Example 1:</u> *W matrix for a measurement series with an independent error correlation structure*

This is the case described in Section 2.2.1.1 and is handled in our data structure with the *Ur1* and *Ur2* arrays. However, if we were to take the W matrix approach to independent measurements we would simply have,

$$\underline{X} = I\,\underline{\tilde{X}},$$

where I is the identity matrix. The *U* matrix in this case would be given by,

$$U = diag(u(\underline{X}))$$

NB: $u(\underline{X})$ are the values contained in the *Ur1*/*Ur2* arrays.

Example 2: *W matrix for a measurement series with random and systematic error correlation structure*

This is the case described in Section 2.2.1.2 and is handled in our data structure with the *Ur1*, *Ur1*, *Us1* and *Us2* arrays. Again, though, in this example we look how this could be represented in *W* matrix form.

It is very similar to the previous independent case, except an additional element is introduced into the $\tilde{X}$ vector to represent the systematic error effect, which is shared by the elements of the $\underline{X}$ vector. For example, for a series of 5 measurements the W matrix equation would look like,

$$\begin{pmatrix} X(t_1) \\ X(t_2) \\ X(t_3) \\ X(t_4) \\ X(t_5) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \tilde{X}^{\mathrm{r}}(t_1) \\ \tilde{X}^{\mathrm{r}}(t_2) \\ \tilde{X}^{\mathrm{r}}(t_3) \\ \tilde{X}^{\mathrm{r}}(t_4) \\ \tilde{X}^{\mathrm{r}}(t_5) \\ \tilde{X}^{\mathrm{sys}} \end{pmatrix}$$

Here we assume that a measured $X(t)$ is equal to the unknown true value with a random error, $\tilde{X}^{\mathrm{r}}(t)$, combined with the systematic error, $\tilde{X}^{\mathrm{sys}}$, thus, $X(t_n) = \tilde{X}^{\mathrm{r}}(t_n) + \tilde{X}^{\mathrm{sys}}$. The *U* matrix in this case is therefore given by,

$$U = \begin{pmatrix} u_{\mathrm{r}}(X(t_1)) & 0 & 0 & 0 & 0 & 0 \\ 0 & u_{\mathrm{r}}(X(t_2)) & 0 & 0 & 0 & 0 \\ 0 & 0 & u_{\mathrm{r}}(X(t_3)) & 0 & 0 & 0 \\ 0 & 0 & 0 & u_{\mathrm{r}}(X(t_4)) & 0 & 0 \\ 0 & 0 & 0 & 0 & u_{\mathrm{r}}(X(t_5)) & 0 \\ 0 & 0 & 0 & 0 & 0 & u_{\mathrm{s}}(\tilde{X}^{\mathrm{sys}}) \end{pmatrix}$$

NB: $u_r(X(t_n))$ are the values contained in *Ur1*/*Ur2* array and $u_{\mathrm{s}}(X^{\mathrm{sys}})$ is the repeated value contained in the corresponding *Us1*/*Us2* array.

Example 3: *W matrix for a measurement series with triangular error correlation form*

For parameters in a measurement series determined by performing a rolling average on a series uncorrelated measurements the resulting correlation form is referred to as “triangular”. This time a *W* matrix would be required to represent this in our data structure.

W takes the form of a banded diagonal matrix. For example, for a series of 5 measurements each determined by performing a rolling averaging over a set uncorrelated “raw” measurements with a kernel size of 3 measurements, we would have,

$$\begin{pmatrix} X(t_1) \\ X(t_2) \\ X(t_3) \\ X(t_4) \\ X(t_5) \end{pmatrix} = \begin{pmatrix} ⅓ & ⅓ & ⅓ & 0 & 0 & 0 & 0 \\ 0 & ⅓ & ⅓ & ⅓ & 0 & 0 & 0 \\ 0 & 0 & ⅓ & ⅓ & ⅓ & 0 & 0 \\ 0 & 0 & 0 & ⅓ & ⅓ & ⅓ & 0 \\ 0 & 0 & 0 & 0 & ⅓ & ⅓ & ⅓ \end{pmatrix} \begin{pmatrix} \tilde{X}(t_0) \\ \tilde{X}(t_1) \\ \tilde{X}(t_2) \\ \tilde{X}(t_3) \\ \tilde{X}(t_4) \\ \tilde{X}(t_5) \\ \tilde{X}(t_6) \end{pmatrix}$$

Here the *U* matrix is given by,

$$U = diag\left(u(\underline{\tilde{X}})\right)$$

NB: This is data not contained in the *Ur1*, *Ur2*, *Us1* or *Us2* arrays.

Storing W Matrices in Sparse Matrix Representation

For the harmonisation process if any of the sensor observed quantity, $\underline{X}_{1/2,n_X}$ contained the *X1* and *X2* arrays has an error correlation structure between the match-ups that is more complex than fully independent or fully systematic (which can be represented by populating the corresponding columns of the *Ur* and *Us* arrays), the *W* matrix describing this must be included in the harmonisation input file. In addition, the uncertainties of the values in the corresponding $\underline{\tilde{X}}_{1/2,n_X}$, $u(\underline{\tilde{X}}_{1/2,n_X})$, are also required. This section explains how that is a achieved in the match up dataset file format.

W Matrices in Compressed Sparse Row Representation

In general, *W* matrices are large (with dimensions of *M* by the length of $\underline{\tilde{X}}_{1/2,n_X}$) and sparse. In order to exploit this sparsity a format of matrix storage called compressed sparse row (CSR) is used in the harmonisation input files. CRS is a standard format which represents a matrix with three 1D arrays that respectively contain nonzero values, the extents of rows, and column indices. We will refer to these three arrays as the following:

- *A* – non-zero values of the matrix;
- *I*A – indices of the points in *A* and *JA* that a new row starts and ends;
- *JA* – column indices of the non-zero elements in the *A* vector.

The following shows how this is utilised for the storage of our *W* matrices.

Since there may in general be a maximum of $m = m_1 + m_2$ *W* matrices for each harmonisation input file (one for each observed quantity of both sensors) this format must be extended to allow for the inclusion of a number, N_W, of *W* matrices. Each *W* matrix, $\underline{W}_{1/2,n_X}$, therefore has its own set of arrays $A_{1/2,n_X}$, $IA_{1/2,n_X}$, and $JA_{1/2,n_X}$,. We will now label these as W_{n_W}, with $n_W \epsilon [1, N_W]$. The variables in the harmonisation input file are concatenations of these individual arrays in the following way:

w_matrix_val = $[A_1; \dots; A_{N_W}]$

w_matrix_col = $[JA_1; \dots; JA_{N_W}]$

w_matrix_row= $[[IA_1], \dots , [IA_{N_W}]]$

Then an additional variable storing the length of each A_{n_W} is introduced to allow the book keeping of this data:

$$w\_matrix\_nnz = [\mathrm{len}(A_1), \ldots, \mathrm{len}(A_{N_W})]$$

A final pair of arrays, *w_matrix_use1* and *w_matrix_use2* are then introduced to map between the *W* matrices and their corresponding observed quantities, $\underline{X}_{1/2,n_X}$ - the columns of the *X1* and *X2* arrays. They respectively have one entry per *X1* or *X2* column and so are lengths *m1* and *m2*. They are filled in the following way,

- If $\underline{X}_{1/2,n_X}$ has no associated *W* matrix (it has an error correlation form of fully independent or systematic) the corresponding entry in *w_matrix_use* is 0.
- If $\underline{X}_{1/2,n_X}$ is associated with a *W* matrix numbered n_W the corresponding entry in *w_matrix_use* is n_W.

U Matrix Representation

As previously mentioned harmonisation requires the uncertainties for the independent measurements to build the matrices $U(\underline{\tilde{X}}_{1/2,n_X})$ (where $U(\underline{\tilde{X}}_{1/2,n_X}) = \mathrm{diag}[u(\underline{\tilde{X}}_{1/2,n_X})]$). In general, N_u such *U* matrices are required, up to *m*, each now labelled as U_{n_u} (where $U_{n_u} = \mathrm{diag}[u_{n_u}]$). These data are concatenated into harmonisation input file variables in a similar way to the *W* matrix values described above, in the following way,

$$u\_matrix\_val = [u_1, \ldots, u_{N_U}]$$

$$u\_matrix\_row\_count = [\mathrm{len}(u_1), \ldots, \mathrm{len}(u_{N_u})]$$

Again a final pair of arrays, *u_matrix_use1* and *u_matrix_use2* are then introduced to map between the *U* matrices and their corresponding observed quantities, $\underline{X}_{1/2,n_X}$ - the columns of the *X1* and *X2* arrays. They respectively have one entry per *X1* or *X2* column and so are lengths *m1* and *m2*. They are filled in the following way,

- If $\underline{X}_{1/2,n_X}$ has no associated *W* matrix (it has an error correlation form of fully independent or systematic) the corresponding entry in *u_matrix_use* is 0.
- If $\underline{X}_{1/2,n_X}$ is associated with a *W* matrix and *U* matrix the corresponding entry in *u_matrix_use* is n_u.

NB: In general it is possible that $N_u \neq N_W$. For example, the same *W* matrix may be used for multiple $\underline{X}_{1/2,n_X}$ as they may be correlated in the same way. However, wherever a *W* matrix is located in *w_matrix_use* there must also be an *U* matrix located in *u_matrix_use*.

2.2.1.4 Structured + Systematic Error Correlation

This is the case where $\underline{X}_{1/n_X}$ is subject to a structured error correlation effect which may be described as,

$$X_{k,n_X} = W_{k,n_X}\underline{\tilde{X}}_{k,n_X}$$

where $\underline{\tilde{X}}_{k,n_X}$ is also subject to a systematic error correlation. In principle, this could be described by one *W* matrix still but in practice it is worthwhile to separate as a special case. For this uncertainty type data is required for both structured and systematic uncertainties, see sections *2.2.1.2* and *2.2.1.3* for further details.

2.2.1.5 Assigning Uncertainty

We have now introduced the four varieties of match-up error correlation and the data required in each case. In summary:

1. Independent Error Correlation:
   - Populate appropriate column of *Ur1/Ur2* array with uncertainties.
2. Systematic and Independent Error Correlation
   - Populate appropriate column of *Ur1/Ur2* and *Us1/Us2* arrays with uncertainties.
3. Structured Error Correlation
   - Populate W matrix variables, add appropriate entry to *w_matrix_use* and *u_matrix_use.*
4. Structured + Systematic Error Correlation
   - Populate W matrix variables, add appropriate entry to *w_matrix_use* and *u_matrix_use.*
   - Populate appropriate column of *Us1/Us2* arrays with uncertainties.

Additionally another pair of variables, *uncertainty_type1* and *uncertainty_type2*, are introduced to indicate the error correlation structure each observed quantity, $\underline{X}_{1/2,n_X}$ - columns of the *X1* and *X2* arrays. They respectively have one entry per *X1* or *X2* column, and so are lengths *m1* and *m2*. The entries are labelled as above:

- 1 – Independent Error Correlation
- 2 – Independent + Systematic Error Correlation
- 3 – Structured Error Correlation
- 4 – Structured + Systematic Error Correlation

2.2.2 Match-up adjustment factor Uncertainty

Uncertainties are also required for the estimates of the match-up adjustment factor, K. It currently is assumed that all errors in K are independent, though are separated into two categories. Uncertainties due to the SRF differences should be also be stored in a vector $\underline{K}_r$ and the uncertainties due to the match-up process should be stored in a vector $\underline{K}_s$.

3 File Format Description

The required data, described in the previous section, should be stored in a series netCDF files, with the match-up data for each pair sensors stored in its own file. The structure of each match-up series file is described in this section.

3.1 Naming Convention

There is no requirement for the filename of each match-up series data file, but it should be clearly named such it can be easily identified.

3.2 Match-Up Series Harmonisation Input File

This section defines the requirements for the contents of each sensor pair match-up series netCDF file.

Attributes

- *sensor_1_name* – Identification number of the first sensor in the match-up series
- *sensor_2_name* – Identification number of the second sensor in the match-up series

Dimensions

- *M* – Total number of match-ups in file
- *m1* – Number of data variables in the first sensor measurement equation
- *m2* – Number of data variables in the second sensor measurement equation

If *W* matrix data stored to express the error correlation structure of observed quantity match-up data also include:

- *w_matrix_count* – Number of *W* matrices included in dataset
- *w_matrix_row_count* – Number of elements in the sparse representation of each *W* matrix row starts variable (M+1).
- *w_matrix_nnz_sum* – Combined number of non-zero elements in all w matrices in file
- *u_matrix_count* – Number of *u* matrices included in the file.
- *u_matrix_row_count_sum* – Combined number of *u* matrix non-zero elements.

Variables - Required

- float *X1*[*M*, *m1*]
  Description = "Measurement equation data variables per matchup for sensor 1"
- float *X2*[*M*, *m2*]
  Description = "Measurement equation data variables per matchup for sensor 2"
- float *Ur1*[*M*, *m1*]
  Description = "Uncertainties for *X1* array with random error correlation structure"
- float *Ur2*[*M*, *m2*]
  Description = "Uncertainties for *X2* array with random error correlation structure"
- float *Us1*[*M*, *m1*]
  Description = "Uncertainties for *X1* array with systematic correlation error structure"
- float *Us2*[*M*, *m2*]
  Description = "Uncertainties for *X2* array with systematic correlation error structure"
- int *uncertainty_type1*[*m1*]
  Description="Uncertainty correlation type per *X1* column, labelled as,
  1 – Independent Error Correlation
  2 – Independent + Systematic Error Correlation
  3 – Structured Error Correlation
  4 – Structured + Systematic Error Correlation
- int *uncertainty_type2*[*m2*]
  Description="Uncertainty correlation type per *X2* column, labelled as,
  1 – Independent Error Correlation
  2 – Independent + Systematic Error Correlation

3 – Structured Error Correlation
4 – Structured + Systematic Error Correlation"

- float *K*[*M*]
  Description="Sensor to sensor differences"
- float *Kr*[*M*]
  Description="Random uncertainties of sensor-to-sensor differences"
- float *Ks*[*M*]
  Description="Systematic uncertainties of sensor-to-sensor differences"
- double *time1*[*M*]
  Description="Sensor 1 match-up time (seconds since 1970-01-01)"
- double *time2*[*M*]
  Description="Sensor 2 match-up time (seconds since 1970-01-01)"

Variables – Optional Indices

These variables may be required to evaluate some but not all measurement equations. Each is optional to include.

- int across_track_index1[*M*]
  Description="Across track index for sensor 1"
- int across_track_index2[*M*]
  Description="Across track index for sensor 2"
- int along_track_index1[*M*]
  Description="Along track index for sensor 1"
- int along_track_index2[*M*]
  Description="Along track index for sensor 2"

Variables – Additional W Matrix Variables

These variables are only required if *W* matrices are required to describe the error correlation structure of observed parameters (columns of *X1* and *X2*) between match-ups. If they are required, the whole set must be included.

- float *w_matrix_val*[*w_matrix_nnz_sum*]
  Description="values of w matrix non-zero elements"
- int *w_matrix_row*[*w_matrix_count*, *w_matrix_row_count*]
  Description="locations within w_matrix_val of new rows per w matrix"
- int *w_matrix_col*[*w_matrix_sum_nnz*]
  Description="column index of w matrix non-zero elements"
- int *w_matrix_nnz* [*w_matrix_count*]
  Description="number of non-zero elements per w matrix"
- int *w_matrix_use1*[*m1*]
  Description="mapping of w matrices to *X1* columns"
- int *w_matrix_use2*[*m2*]
  Description="mapping of w matrices to *X2* columns"
- int *u_matrix_row_count* [*u_matrix_count*]
  Description="locations within *u_matrix_val* of new *U* matrix"
- double *u_matrix_val*[*u_matrix_row_count_sum*]

Description="values *U* matrix non-zero elements"

- int *u_matrix_use1*[*m1*]
  Description="mapping of *U* matrices to *X1* columns"
- int *u_matrix_use2*[*m2*]
  Description="mapping of *U* matrices to *X2* columns"

Variables – Diagnostic

Optional for the harmonisation process itself, though recommended for analysis of its results, are further variables with data per matchup (of dimension *M*) – e.g. nominal measurands, longitudes, latitudes and observation angles. If the variable is specific to sensor 1 or 2 is should be suffixed with "1" or "2" for the reference of the harmonisation plotting software. The current list of useful general diagnostic variables is as follows:

- double *nominal_measurand1*[*M*]
  Description="Nominal measurand value for sensor 1"
- double *nominal_measurand2*[*M*]
  Description="Nominal measurand value for sensor 2"
- double *lon1*[*M*]
  Description="Longitude of match-up Earth location for Sensor 1"
- double *lon2*[*M*]
  Description=" Longitude of match-up Earth location for Sensor 2"
- double *lat1*[*M*]
  Description="Latitude of match-up Earth location for Sensor 1"
- double *lat2*[*M*]
  Description=" Latitude of match-up Earth location for Sensor 2"
- double *sza1*[*M*]
  Description="Sun-zenith angle for Sensor 1"
- double *sza2*[*M*]
  Description="Sun-zenith angle for Sensor 2"
- double *vza1*[*M*]
  Description="Satellite-zenith angle for Sensor 1"
- double *vza2*[*M*]
  Description="Satellite-zenith angle for Sensor 2"
- double *vaa1*[*M*]
  Description="Azimuth angle for Sensor 1"
- double *vaa2*[*M*]
  Description="Azimuth angle for Sensor 2"