Simultaneous Nadir Overpasses – Summary Notes and Unified Notation

Ross Bannister

May 23, 2018

1 Notation

The following table of notation summarises the symbols used in these summary notes.

Symbol	Description	Symbol	Description
i	Measurement index ($1 \leq i \leq M$)	k	Channel index ($1 \leq k \leq K$)
j	Ensemble member ($1 \leq j \leq N$)	l	General vector component index
$L^{\mathrm{t}}_{i(k)}$, $L^{\mathrm{r}}_{i(k)}$	Radiance measurement i of channel k taken with the target/reference instrument	$\mathbf{l}^{\mathrm{t}}_i$, $\mathbf{l}^{\mathrm{r}}_i$	K-element vectors comprising the elements of $L^{\mathrm{t}}_{i(k)}$, $L^{\mathrm{r}}_{i(k)}$, with particular error $\delta\mathbf{l}^{\mathrm{t}}_i$, $\delta\mathbf{l}^{\mathrm{r}}_i$
$\tilde{L}^{\mathrm{r}}_{i,\mathrm{mod}(l)}$, $\tilde{L}^{\mathrm{t}}_{i(l)}$	Projection of $\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}/\mathbf{l}^{\mathrm{t}}_i$ on to the lth uncorrelated component.	$\tilde{\mathbf{l}}^{\mathrm{t}}_i$, $\tilde{\mathbf{l}}^{\mathrm{r}}_i$	K-element vectors comprising the elements of $\tilde{L}^{\mathrm{r}}_{i,\mathrm{mod}(l)}$, $\tilde{L}^{\mathrm{t}}_{i(l)}$; each has error covariance $\mathbf{I}$
$\mathbf{B}^{\mathrm{A}}$, $\mathbf{B}^{\mathrm{B}}$	$K \times K$-element regression matrix for option A/B	$\mathbf{a}^{\mathrm{A}}$, $\mathbf{a}^{\mathrm{B}}$	K-element regression off-set for option A/B
$\mathbf{b}^{\mathrm{A}}$	K-element vector comprising the diagonal elements of $\mathbf{B}^{\mathrm{A}}$		
$\mathbf{R}^{\mathrm{t}}$, $\mathbf{R}^{\mathrm{r}}$	Error covariance of target/reference spectra	$\mathbf{l}^{\mathrm{t}}_{i,\mathrm{true}}$	Theoretical 'true' version of $\mathbf{l}^{\mathrm{t}}_i$
$\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}$	Estimated reference spectrum if it were made at the position of the target instrument (modified from $\mathbf{l}^{\mathrm{r}}_i$)	$\hat{\mathbf{l}}^{\mathrm{r}}_{i,\mathrm{mod}}$, $\hat{\tilde{\mathbf{l}}}^{\mathrm{r}}_{i,\mathrm{mod}}$	Regressed version of $\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}$, $\tilde{\mathbf{l}}^{\mathrm{r}}_{i,\mathrm{mod}}$
$\delta\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}$	Particular error in $\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}$	$\mathbf{R}^{\mathrm{r}}_{\mathrm{mod}}$	Error covariances of $\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}$
$\mathbf{x}^{\mathrm{t}}_i$, $\mathbf{x}^{\mathrm{r}}_i$	n-element vector of model's vertical profile for the scene of the target/reference instrument, with particular error $\delta\mathbf{x}^{\mathrm{t}}_i$, $\delta\mathbf{x}^{\mathrm{r}}_i$	$\delta\mathbf{x}^{\mathrm{t}}_{i,j}$, $\delta\mathbf{x}^{\mathrm{r}}_{i,j}$	n-element ensemble perturbations at the target/reference scene
θ^{t}_i	Viewing angle of the target instrument	θ^{r}_i	Viewing angle of the reference instrument
$\mathbf{h}(\mathbf{x}, \theta)$	Radiative transfer model. Inputs: n-element vector, and scalar, output: K-element vector	$\boldsymbol{\epsilon}^{\mathbf{h}}_{\mathrm{t}}$, $\boldsymbol{\epsilon}^{\mathbf{h}}_{\mathrm{r}}$	Error in the radiative transfer model when modelling the target/reference spectra
$\Delta\mathbf{x}^{\mathrm{rt}}_i$	$\mathbf{x}^{\mathrm{t}}_i - \mathbf{x}^{\mathrm{r}}_i$	$\Delta\theta^{\mathrm{rt}}_i$	$\theta^{\mathrm{t}}_i - \theta^{\mathrm{r}}_i$
$\mathbf{H}_{\mathbf{x}_0,\theta_0}$	$K \times n$-element Jacobian $\partial\mathbf{h}/\partial\mathbf{x}\vert_{\mathbf{x}_0,\theta_0}$	$\mathbf{H}_{\theta_0,\mathbf{x}_0}$	$K \times 1$-element Jacobian $\partial\mathbf{h}/\partial\theta\vert_{\theta_0,\mathbf{x}_0}$

2 The problem

We would like to relate radiance measurements made by a target instrument, $L^{\mathrm{t}}_{i(k)}$ to a reference instrument $L^{\mathrm{r}}_{i(k)}$ (i is the measurement number and k is the channel index). A calibration procedure is to be performed to allow this, and we assume that the two instruments share the same set of channels, but we do not assume that the instruments observe the same scene. The radiance measurements are placed in the K-element observation vectors (spectra) $\mathbf{l}^{\mathrm{t}}_i$ and $\mathbf{l}^{\mathrm{r}}_i$ respectively. The regression relationship takes the following form:

$$\hat{\mathbf{l}}^{\mathrm{r}}_{i,\mathrm{mod}} = \mathbf{B}^{\mathrm{A}}\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}} + \mathbf{a}^{\mathrm{A}}, \tag{1}$$

where $\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}$ is the spectrum that the reference instrument would have measured if it were made at the position of the target instrument (a modification to $\mathbf{l}^{\mathrm{r}}_i$ – see below) and $\hat{\mathbf{l}}^{\mathrm{r}}_{i,\mathrm{mod}}$ is the regressed version of this spectrum, which is designed to be as close as possible to the target spectrum $\mathbf{l}^{\mathrm{t}}_i$. The way that (1) is used is described in Sect. 4. We shall consider two forms of regression objects: one, $\mathbf{B}^{\mathrm{A}}$ and $\mathbf{a}^{\mathrm{A}}$, assumes that $\mathbf{B}^{\mathrm{A}}$ is diagonal and another, $\mathbf{B}^{\mathrm{B}}$ and $\mathbf{a}^{\mathrm{B}}$, that does not (see Sects. 5 and 6 respectively).

The following information is given.

- The spectra from the target instrument have error covariance $\mathbf{R}^{\mathrm{t}}$.
- The spectra from the reference instrument have error covariance $\mathbf{R}^{\mathrm{r}}$.
- The model's vertical profile at the target scene, $\mathbf{x}^{\mathrm{t}}_i$ and a set of correctly spread ensemble perturbations, $\delta\mathbf{x}^{\mathrm{t}}_{i,j}$.
- The model's vertical profile at the reference scene, $\mathbf{x}^{\mathrm{r}}_i$ and a set of correctly spread ensemble perturbations, $\delta\mathbf{x}^{\mathrm{r}}_{i,j}$.
- The target instrument's viewing angle, θ^{t}_i.
- The reference instrument's viewing angle, θ^{r}_i.

The previous document (Sect. 5.1) showed that the 'true' radiance that would be measured at the location of the target instrument is theorised to be

$$\mathbf{l}^{\mathrm{t}}_{i,\mathrm{true}} = \underbrace{\mathbf{l}^{\mathrm{r}}_i + \mathbf{H}_{\mathbf{x}^{\mathrm{r}}_i,\theta^{\mathrm{r}}_i}\Delta\mathbf{x}^{\mathrm{rt}}_i + \mathbf{H}_{\theta^{\mathrm{r}}_i,\mathbf{x}^{\mathrm{r}}_i}\Delta\theta^{\mathrm{rt}}_i}_{\substack{\text{known/calculable}\\ \mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}}} + \underbrace{-\delta\mathbf{l}^{\mathrm{r}}_i + \mathbf{H}_{\mathbf{x}^{\mathrm{r}}_i,\theta^{\mathrm{r}}_i}\delta\mathbf{x}^{\mathrm{r}}_i - \mathbf{H}_{\mathbf{x}^{\mathrm{t}}_i,\theta^{\mathrm{t}}_i}\delta\mathbf{x}^{\mathrm{t}}_i}_{\substack{\text{statistics known}\\ \delta\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}}} + \underbrace{\boldsymbol{\epsilon}^{\mathbf{h}}_{\mathrm{r}} - \boldsymbol{\epsilon}^{\mathbf{h}}_{\mathrm{t}}}_{\text{completely unknown}} \tag{2}$$

(see the table for the meanings of the symbols). This comprises a part that is known/calculable ($\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}$, the reference instrument's measured spectrum, $\mathbf{l}^{\mathrm{r}}_i$, modified with correction terms due to the different scenes), an error part ($\delta\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}$) that is not known, but whose statistics are known/calculable, and a completely unknown part. We shall ignore the completely unknown part for now and assume that its value is negligible compared to the other terms. The purpose of writing (2) was to derive its estimate, $\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}$, and its error, $\delta\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}$. The problem (posed here in terms of $\mathbf{a}^{\mathrm{A}}$ and $\mathbf{B}^{\mathrm{A}}$) is as follows.

What are the values of $\mathbf{a}^{\mathrm{A}}$ and $\mathbf{B}^{\mathrm{A}}$ such that $\mathbf{B}^{\mathrm{A}}\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}+\mathbf{a}^{\mathrm{A}}$, is as close as possible to $\mathbf{l}^{\mathrm{t}}_{i}$? This should be done in a way that is consistent with the error statistics of $\mathbf{l}^{\mathrm{t}}_{i}$ and $\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}$.

- Error statistics of $\mathbf{l}^{\mathrm{t}}_{i}$: $\mathbf{R}^{\mathrm{t}}$.
- Error statistics of $\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}$:

$$\begin{aligned}
\mathbf{R}^{\mathrm{r}}_{\mathrm{mod}} &= \left\langle \delta\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}\delta\mathbf{l}^{\mathrm{r}\,\mathrm{T}}_{i,\mathrm{mod}} \right\rangle \\
&\approx \mathbf{R}^{\mathrm{r}} + \left\langle \mathbf{H}_{\mathbf{x}^{\mathrm{r}}_{i},\theta^{\mathrm{r}}_{i}}\delta\mathbf{x}^{\mathrm{r}}_{i,j}\left(\mathbf{H}_{\mathbf{x}^{\mathrm{r}}_{i},\theta^{\mathrm{r}}_{i}}\delta\mathbf{x}^{\mathrm{r}}_{i,j}\right)^{\mathrm{T}} \right\rangle + \left\langle \mathbf{H}_{\mathbf{x}^{\mathrm{t}}_{i},\theta^{\mathrm{t}}_{i}}\delta\mathbf{x}^{\mathrm{t}}_{i,j}\left(\mathbf{H}_{\mathbf{x}^{\mathrm{t}}_{i},\theta^{\mathrm{t}}_{i}}\delta\mathbf{x}^{\mathrm{t}}_{i,j}\right)^{\mathrm{T}} \right\rangle \\
&\quad + \left\langle \left(\mathbf{H}_{\mathbf{x}^{\mathrm{r}},\theta^{\mathrm{r}}}\delta\mathbf{x}^{\mathrm{r}}_{i,j}\right)\left(\mathbf{H}_{\mathbf{x}^{\mathrm{t}},\theta^{\mathrm{t}}}\delta\mathbf{x}^{\mathrm{t}}_{i,j}\right)^{\mathrm{T}} \right\rangle + \left\langle \left(\mathbf{H}_{\mathbf{x}^{\mathrm{t}},\theta^{\mathrm{t}}}\delta\mathbf{x}^{\mathrm{t}}_{i,j}\right)\left(\mathbf{H}_{\mathbf{x}^{\mathrm{r}},\theta^{\mathrm{r}}}\delta\mathbf{x}^{\mathrm{r}}_{i,j}\right)^{\mathrm{T}} \right\rangle, \qquad (3)
\end{aligned}$$

  where the angled brackets are averages over ensemble members (index j).

The cost function that needs to be minimised to determine $\mathbf{a}^{\mathrm{A}}$ and $\mathbf{B}^{\mathrm{A}}$ is Eq. (33) of the previous document, with the modifications outlined in Sect. 5.1 of that document:

$$J[\mathbf{a}^{\mathrm{A}},\mathbf{B}^{\mathrm{A}}] = \frac{1}{2}\sum_{i=1}^{M}\left[\mathbf{l}^{\mathrm{t}}_{i}-\mathbf{a}^{\mathrm{A}}-\mathbf{B}^{\mathrm{A}}\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}\right]^{\mathrm{T}}\left(\mathbf{R}^{\mathrm{t}}+\mathbf{B}^{\mathrm{A}}\mathbf{R}^{\mathrm{r}}_{\mathrm{mod}}\mathbf{B}^{\mathrm{A}}\right)^{-1}\left[\mathbf{l}^{\mathrm{t}}_{i}-\mathbf{a}^{\mathrm{A}}-\mathbf{B}^{\mathrm{A}}\mathbf{l}^{\mathrm{r}}_{i,\mathrm{mod}}\right], \qquad (4)$$

which we will consider in two different ways – A (Sect. 5) and B (Sect. 6). ((4) has been set up for the specific case of option A.)

J in (4) should be minimised with respect to its arguments to determine the regression objects. This is not a so-called *ordinary least-squares problem* (which requires minimisation of a quadratic form) but is a *total least squares problem* (which is not quadratic). A total least squares problem includes uncertainties in the 'x' and 'y' data and is not a quadratic form owing to the presence of the regression object $\mathbf{B}^{\mathrm{A}}$ in the inverse matrix in (4). As this is a more difficult problem than in ordinary least squares, we consider two different approaches to this problem (see Sects. 5 and 6).

3 Note on the use of ensembles

Equation (3) uses ensembles to estimate contributions to $\mathbf{R}^{\mathrm{r}}_{\mathrm{mod}}$, where $\delta\mathbf{x}^{\mathrm{r}}_{i,j}$ and $\delta\mathbf{x}^{\mathrm{t}}_{i,j}$ represent vertical columns of model ensemble perturbations corresponding to the positions of the reference and target observations respectively (spectrum i, perturbation j). Even if we assume that $\mathbf{R}^{\mathrm{r}}$ is of fixed value for a given population of near-located measurements, $\mathbf{R}^{\mathrm{r}}_{\mathrm{mod}}$ will strictly not be of fixed value due to the i-dependence of the ensemble perturbations. A proposal, at least in the first instance, is to use an average over the population. For example, the first correction term to (3) would be:

$$\begin{aligned}
\left\langle \mathbf{H}_{\mathbf{x}^{\mathrm{r}}_{i},\theta^{\mathrm{r}}_{i}}\delta\mathbf{x}^{\mathrm{r}}_{i,j}\left(\mathbf{H}_{\mathbf{x}^{\mathrm{r}}_{i},\theta^{\mathrm{r}}_{i}}\delta\mathbf{x}^{\mathrm{r}}_{i,j}\right)^{\mathrm{T}} \right\rangle &\rightarrow \frac{1}{M}\sum_{i=1}^{M}\left\langle \mathbf{H}_{\mathbf{x}^{\mathrm{r}}_{i},\theta^{\mathrm{r}}_{i}}\delta\mathbf{x}^{\mathrm{r}}_{i,j}\left(\mathbf{H}_{\mathbf{x}^{\mathrm{r}}_{i},\theta^{\mathrm{r}}_{i}}\delta\mathbf{x}^{\mathrm{r}}_{i,j}\right)^{\mathrm{T}} \right\rangle \\
&= \frac{1}{M}\sum_{i=1}^{M}\frac{1}{N}\sum_{j=1}^{N}\mathbf{H}_{\mathbf{x}^{\mathrm{r}}_{i},\theta^{\mathrm{r}}_{i}}\delta\mathbf{x}^{\mathrm{r}}_{i,j}\left(\mathbf{H}_{\mathbf{x}^{\mathrm{r}}_{i},\theta^{\mathrm{r}}_{i}}\delta\mathbf{x}^{\mathrm{r}}_{i,j}\right)^{\mathrm{T}}. \qquad (5)
\end{aligned}$$

4 Using the regression objects once they are determined

Once $\mathbf{B}^{\mathrm{A}}$ and $\mathbf{a}^{\mathrm{A}}$ (or $\mathbf{B}^{\mathrm{B}}$ and $\mathbf{a}^{\mathrm{B}}$) have been determined, they are used to process target measurements. Since in (1) (or (14) below), $\hat{\mathbf{l}}^{\mathrm{r}}_{i,\mathrm{mod}}$ is designed to be as close as possible to the target spectrum, this symbol takes on the role of $\mathbf{l}^{\mathrm{t}}_{i}$. Given a target measurement, the 'corrected' version is

$$\mathbf{B}^{\mathrm{A}^{-1}}\mathbf{l}^{\mathrm{t}}_{i}-\mathbf{B}^{\mathrm{A}^{-1}}\mathbf{a}^{\mathrm{A}} \qquad (6)$$

(or with $\mathbf{B}^{\mathrm{A}}\rightarrow\mathbf{B}^{\mathrm{B}}$ and $\mathbf{a}^{\mathrm{A}}\rightarrow\mathbf{a}^{\mathrm{B}}$ when used with (14)).

5 Solution option A

This solution assumes that the matrix $\mathbf{B}^{\mathrm{A}}$ is diagonal ($\mathbf{B}^{\mathrm{A}} = \mathrm{diag}(\mathbf{b}^{\mathrm{A}})$), which means that the regression may be done channel-by-channel (although the determination of the regression elements must be done with all channels together given that errors in channels are correlated). The solution to this using a gradient descent approach has been coded and tested with synthetic data. The code is available here.

6 Solution option B

This solution forms a different problem to (4), and the details are given in Sect. 8 of the previous document. It relies on a solution provided by Reed (1989) using scalars.

The problem solved in this case is slightly different to (4). Recall that the input data are sets of vectors $\mathbf{l}_i^{\mathrm{t}}$ and $\mathbf{l}_{i,\mathrm{mod}}^{\mathrm{r}}$ with error covariances $\mathbf{R}^{\mathrm{t}}$ and $\mathbf{R}_{\mathrm{mod}}^{\mathrm{r}}$ respectively. We first find transforms that decorrelate errors in the input data. Let

$$\begin{aligned} \tilde{\mathbf{l}}_i^{\mathrm{t}} &= \mathbf{R}^{\mathrm{t}\,-1/2}\mathbf{l}_i^{\mathrm{t}}, \\ \tilde{\mathbf{l}}_{i,\mathrm{mod}}^{\mathrm{r}} &= \mathbf{l}_{i,\mathrm{mod}}^{\mathrm{r}}. \end{aligned} \tag{7}$$

The errors in the variables with a tilde are uncorrelated between elements. To see this consider an error $\delta\mathbf{l}_i^{\mathrm{t}}$, which has error covariance $\left\langle \delta\mathbf{l}_i^{\mathrm{t}}\delta\mathbf{l}_i^{\mathrm{t}\,\mathrm{T}} \right\rangle = \mathbf{R}^{\mathrm{t}}$. The corresponding error in the 'tilde' space is $\delta\tilde{\mathbf{l}}_i^{\mathrm{t}} = \mathbf{R}^{\mathrm{t}\,-1/2}\delta\mathbf{l}_i^{\mathrm{t}}$, which has error covariance $\left\langle \delta\tilde{\mathbf{l}}_i^{\mathrm{t}}\delta\tilde{\mathbf{l}}_i^{\mathrm{t}\,\mathrm{T}} \right\rangle = \left\langle \mathbf{R}^{\mathrm{t}\,-1/2}\delta\mathbf{l}_i^{\mathrm{t}} \left(\mathbf{R}^{\mathrm{t}\,-1/2}\delta\mathbf{l}_i^{\mathrm{t}}\right)^{\mathrm{T}} \right\rangle = \mathbf{R}^{\mathrm{t}\,-1/2} \left\langle \delta\mathbf{l}_i^{\mathrm{t}}\delta\mathbf{l}_i^{\mathrm{t}\,\mathrm{T}} \right\rangle \mathbf{R}^{\mathrm{t}\,-\mathrm{T}/2} = \mathbf{R}^{\mathrm{t}\,-1/2}\mathbf{R}^{\mathrm{t}}\mathbf{R}^{\mathrm{t}\,-\mathrm{T}/2} = \mathbf{I}$. The fact that errors of each component of $\tilde{\mathbf{l}}_i^{\mathrm{t}}$ and $\tilde{\mathbf{l}}_{i,\mathrm{mod}}^{\mathrm{r}}$ are uncorrelated means that we can apply Reed's approach separately to each component.

The version of the regression formula (1) for option B and for uncorrelated elements is

$$\hat{\tilde{\mathbf{l}}}_{i,\mathrm{mod}}^{\mathrm{r}} = \tilde{\mathbf{B}}^{\mathrm{B}}\tilde{\mathbf{l}}_{i,\mathrm{mod}}^{\mathrm{r}} + \tilde{\mathbf{a}}^{\mathrm{B}}, \tag{8}$$

$$\text{which has component } l\text{: } \hat{\tilde{L}}_{i,\mathrm{mod}(l)}^{\mathrm{r}} = \tilde{b}_l^{\mathrm{B}}\tilde{L}_{i,\mathrm{mod}(l)}^{\mathrm{r}} + \tilde{a}_l^{\mathrm{B}}. \tag{9}$$

The last line is now used as we will perform Reed's procedure on each component l separately. The cost function is

$$J[\tilde{a}_l^{\mathrm{B}}, \tilde{b}_l^{\mathrm{B}}] = \frac{1}{2}\sum_{i=1}^{M} \frac{\left(\tilde{L}_{i(l)}^{\mathrm{t}} - \tilde{a}_l^{\mathrm{B}} - \tilde{b}_l^{\mathrm{B}}\hat{\tilde{L}}_{i,\mathrm{mod}(l)}^{\mathrm{r}}\right)^2}{1 + \tilde{b}_l^{\mathrm{B}\,2}}. \tag{10}$$

Reed's solution is detailed in the previous document (Sect. 8.3), but the essence is given here in the unified notation.

For each l, solve the quadratic equation

$$g(\tilde{b}_l^{\mathrm{B}}) = -\sum_{i=1}^{M} U_i V_i \tilde{b}_l^{\mathrm{B}\,2} + \left(\sum_{i=1}^{M} V_i^2 - \sum_{i=1}^{M} U_i^2\right)\tilde{b}_l^{\mathrm{B}} + \sum_{i=1}^{M} U_i V_i = 0, \tag{11}$$

where $U_i = \hat{\tilde{L}}_{i,\mathrm{mod}(l)}^{\mathrm{r}} - \left\langle \hat{\tilde{L}}_{i,\mathrm{mod}(l)}^{\mathrm{r}} \right\rangle$ and $V_i = \tilde{L}_{i(l)}^{\mathrm{t}} - \left\langle \tilde{L}_{i(l)}^{\mathrm{t}} \right\rangle$ (the angled brackets here are averages over measurements i). There are two real roots of (11),

$$\tilde{b}_l^{\mathrm{B}} = \frac{\sum_{i=1}^{M} V_i^2 - \sum_{i=1}^{M} U_i^2 \pm \sqrt{\left(\sum_{i=1}^{M} V_i^2 - \sum_{i=1}^{M} U_i^2\right)^2 + 4\left(\sum_{i=1}^{M} U_i V_i\right)^2}}{2\sum_{i=1}^{M} U_i V_i}, \tag{12}$$

with corresponding values of $\tilde{a}_l^{\mathrm{B}}$ are

$$\tilde{a}_l^{\mathrm{B}} = \left\langle \tilde{L}_{i(l)}^{\mathrm{t}} \right\rangle - \tilde{b}_l^{\mathrm{B}} \left\langle \hat{\tilde{L}}_{i,\mathrm{mod}(l)}^{\mathrm{r}} \right\rangle, \tag{13}$$

but only one pair ($\tilde{a}_l^{\mathrm{B}}$, $\tilde{b}_l^{\mathrm{B}}$) represents the solution that we are after. The way to decide which one is to compute the cost associated with each pair in turn using (10) and to take the pair with the smaller cost.

The values $\tilde{a}_l^{\mathrm{B}}$ make up the components of $\tilde{\mathbf{a}}^{\mathrm{B}}$, and $\tilde{b}_l^{\mathrm{B}}$ make up the diagonal components of $\tilde{\mathbf{B}}^{\mathrm{B}}$. Once these have been determined for $1 \leq l \leq K$, the regression objects are required in the original channel space. Starting with (8) and using both lines of (7), we find that the corresponding regression (1), but for option B is

$$\hat{\mathbf{l}}_{i,\mathrm{mod}}^{\mathrm{r}} = \mathbf{B}^{\mathrm{B}} \mathbf{l}_{i,\mathrm{mod}}^{\mathrm{r}} + \mathbf{a}^{\mathrm{B}}, \tag{14}$$

where $\mathbf{a}^{\mathrm{B}} = {\mathbf{R}^{\mathrm{t}}}^{1/2} \tilde{\mathbf{a}}^{\mathrm{B}}$ and $\mathbf{B}^{\mathrm{B}} = {\mathbf{R}^{\mathrm{t}}}^{1/2} \tilde{\mathbf{B}}^{\mathrm{B}} {\mathbf{R}_{\mathrm{mod}}^{\mathrm{r}}}^{-1/2}$.

This approach has already been coded and tested with synthetic data. The code is available here.