Taylor & Francis
Taylor & Francis Group

Efficient forward modelling by matrix representation of sensor responses

PATRICK ERIKSSON*†, MATTIAS EKSTRÖM†, CHRISTIAN
MELSHEIMER‡ and STEFAN A. BUEHLER‡

†Department of Radio and Space Science, Chalmers University of Technology,
SE-41296 Gothenburg, Sweden
‡Institute of Environmental Physics, University of Bremen, D-28259 Bremen, Germany

The polarization, frequency and spatial responses of the sensor can be considered by calculating the Stokes vector of monochromatic pencil beam radiances for a set of frequencies and viewing directions, and weight these radiances with the instrument responses. This paper presents a highly efficient solution for this calculation procedure. The basic idea is to pre-calculate a matrix that represents the mapping from polarisation, frequency and spatial values to measured data. Sensor impacts can then be included by a simple matrix multiplication. The full sensor matrix can be obtained by determining the response matrix for the sensor parts individually. Data reduction methods can also be incorporated. A simple method for optimizing the calculation grids is further presented. The described approach for sensor modeling has been implemented in two public available softwares for atmospheric radiative transfer simulations.

1. Introduction

The extraction of geophysical data from remote sensing observations, the inversion or retrieval, requires that all significant radiative transfer and sensor impacts are understood and can be accurately modelled. Examples on such impacts are surface volume scattering, atmospheric emission and the frequency response of the sensor. The task of predicting the measurement, with the assumption that all relevant geophysical and sensor parameters are known, is often denoted as forward modelling. The complexity and the need of high calculation efficiency of the forward model are increasing as remote sensing instruments are becoming capable of making more numerous and advanced measurements, such as obtaining data at more frequencies and polarisation states. The trend is also that the measurements are based on radiative interactions that require more demanding simulations, such as measuring cloud microphysical parameters based on scattering of radiation, in either active or passive mode. This means that even with continuously faster computers, the forward modelling can limit the usage of measured data for many applications, where numerical weather prediction is one example.

   Both the implementation and presentation of forward models put in general a high emphasis on the interaction between radiation and matter, while the modelling of sensor impacts is given less emphasis. The frequency and angular resolution can be assumed to be infinite in many cases, and a single monochromatic calculation

*Corresponding author. Email: patrick@eriksson.chalmers.se

along a pencil beam direction is then sufficient. On the other hand, there exist many cases where both the limited frequency and angular resolution must be considered. This latter situation is common, at least, for passive atmospheric measurements. The problem at hand is to estimate the weighting of monochromatic pencil beam radiances (MPBRs) with the sensor responses. The standard method to solve this problem is to calculate the MPBR for a discrete set of frequencies and viewing directions and estimate the weighting integral numerically by some quadrature rule, and this procedure is then repeated for each calculation loop (e.g. Zorn *et al.* 2000, SMILES science team 2001, Urban *et al.* 2004). This paper follows the same basic approach, but instead of repeating a direct numerical integration, the quadrature weights are pre-calculated in the form of a matrix, and the weighting between MPBRs and sensor responses is performed by a matrix–vector multiplication. An important aspect is that a speed improvement is achieved without any deterioration of the calculation accuracy.

The efficiency of this method depends on how many times the calculations are repeated, but this number can be high. The number of such repetitions is normally high for operational retrievals, especially if an iterative inversion method based on Jacobians of the forward model is used. For example, the method described in this paper has been successfully applied for the operational inversions of data from the sub-mm limb sounder on-board the Odin satellite (Eriksson *et al.* 2002b), where the sensor part of the forward model is at a rough estimate two orders of magnitude faster than without this new approach.

2. Theory

2.1 *Theoretical formalism*

A remote sensing measurement, **y**, can be expressed as (Rodgers 2000)

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \mathbf{b}) + \varepsilon, \tag{1}$$

where $\mathcal{F}$ is the forward model, **x** and **b** together give a complete description of the geophysical and sensor states, and ε is the measurement error. The vector **x** holds variables to be retrieved from the measurement, while all other variables are placed in **b**.

Retrieval methods where the forward model is inverted (with some constraint when the problem is ill-posed) and error characterization, make use of multivariate derivatives of $\mathcal{F}$, the Jacobians. The division of forward model input between **x** and **b** gives the following Jacobians:

$$\mathbf{K_x} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}}, \tag{2}$$

for the state vector and

$$\mathbf{K_b} = \frac{\partial \mathbf{y}}{\partial \mathbf{b}}, \tag{3}$$

for the model parameters.

For the purpose of this paper, the forward model $\mathcal{F}$ is divided into two sections, a part treating radiative transfer for pencil beam monochromatic signals

$$\mathbf{i} = \mathcal{F}_r(\mathbf{x}_r, \mathbf{b}_r), \tag{4}$$

and a sensor part including the instrumental responses

$$\mathbf{y} = \mathcal{F}_s(\mathbf{i}, \mathbf{x}_s, \mathbf{b}_s) + \varepsilon. \tag{5}$$

The vector $\mathbf{i}$ holds the Stokes components for the MPBRs for the considered set of monochromatic frequencies and pencil beam directions. The state and model parameter vectors are separated accordingly and can be expressed as

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_s \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}_r \\ \mathbf{b}_s \end{bmatrix}. \tag{6}$$

2.2 Example on instrumental expressions

A passive heterodyne instrument will here be used to exemplify the sensor modelling approach, and this section gives the analytical expressions to implement in the forward model. Three instrument parts will be considered: the antenna, the sideband folding, and the spectrometer. In addition, inclusion of the polarization response and data reduction will also be discussed.

2.2.1 Polarization response. The intensity and polarisation state of radiation are normally described by means of the Stokes vector, $\mathbf{s}$, defined as

$$\mathbf{s} = \begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} = \frac{1}{2}\sqrt{\frac{\varepsilon}{\mu}} \begin{bmatrix} \langle E_v E_v^* + E_h E_h^* \rangle \\ \langle E_v E_v^* - E_h E_h^* \rangle \\ \langle E_v E_h^* - E_h E_v^* \rangle \\ \langle E_h E_v^* - E_v E_h^* \rangle \end{bmatrix}, \tag{7}$$

where $\langle \cdot \rangle$ signifies time average, E_v and E_h are the complex amplitudes for vertical and horizontal polarization and $.^*$ denotes complex conjugate. The first Stokes component (I) is the total intensity, the second component (Q) is the difference between vertical and horizontal polarization, and the last two components, U and V, correspond to linear $\pm 45°$ and circular polarization, respectively.

The measured intensity, I_p, if assuming no losses, is

$$I_p = \frac{1}{2}\mathbf{ps}, \tag{8}$$

where $\mathbf{p}$ is a row vector of length 4 describing the sensor polarization response. For an instrument measuring a single polarization, the first element of $\mathbf{p}$ shall be one and the three last elements shall fulfil $p_2^2 + p_3^2 + p_4^2 = 1$, where p_2 is the second element of $\mathbf{p}$, etc. For example, if vertical polarization is measured, then $\mathbf{p} = [1\ 1\ 0\ 0]$ and for linear $\pm 45°$ polarization $\mathbf{p} = [1\ 0\ \mp 1\ 0]$.

If a rotation of an angle χ is needed to obtain consistent definition between the polarization directions for atmospheric radiation and sensor response, the transformation matrix $\mathbf{L}(\chi)$ can be applied, where (Liou 2002)

$$\mathbf{L}(\chi) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\chi & \sin 2\chi & 0 \\ 0 & -\sin 2\chi & \cos 2\chi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{9}$$

The measured intensity is then

$$\mathbf{I} = \frac{1}{2}\mathbf{p}\mathbf{L}(\chi)\mathbf{s}. \tag{10}$$

2.2.2 Antenna response.

The influence of the antenna can be expressed as a weighting between the normalised antenna response, w_a, as a function of viewing direction and the MPBR field, I:

$$I_a = \int_{4\pi} I(\Omega + \Omega_0)w_a(\Omega)\mathrm{d}\Omega \quad \text{with} \quad \int_{4\pi} w_a(\Omega)\mathrm{d}\Omega = 1, \tag{11}$$

where I_a is the apparent radiation intensity after the antenna, Ω is solid angle and Ω_0 is some reference point for the antenna pattern, normally the point of highest response.

Here it will be assumed that the variation of I in one angular dimension can be neglected, and that the weighting can be described by one-dimensional integral:

$$I_a = \int_{-\pi}^{\pi} I(\theta + \theta_0)w_a^\theta(\theta)\mathrm{d}\theta, \tag{12}$$

$$w_a^\theta(\theta) = \int_0^{2\pi} w_a(\theta,\phi)\cos(\theta)\mathrm{d}\phi, \tag{13}$$

where θ_0 is the reference angle of the antenna.

2.2.3 Sideband folding.

For situations when the measurement is performed at frequencies for which no acceptable direct detection is possible, the measurement signal is converted to a lower frequency by mixing with a local oscillator signal (LO). This translates the measurement frequency, v, to the intermediate frequency, v_{IF}:

$$v_{IF} = |v - v_{LO}|, \tag{14}$$

where v_{LO} is the frequency of the LO. The frequency v is for historical reasons often denoted as the radio frequency.

This procedure has the consequence that frequencies on both sides of v_{LO}, the lower and upper sideband, are folded together. More exactly, frequencies with the same distance to v_{LO} are translated to the same v_{IF}. Thus, the 'image' frequency, v', is

$$v' = 2v_{LO} - v. \tag{15}$$

The signal from v and v' is normally not weighted equally. This can be an effect of a frequency-dependent response of elements from the antenna to the mixer, or caused by a sideband filter, with the purpose of minimizing the contribution from one of the bands. These terms are here lumped together into one term, the sideband response w_{sb}. The apparent intensity after the mixer, I_{IF}, can then be expressed as:

$$I_{IF}(v_{IF}) = \frac{w_{sb}(v)I(v) + w_{sb}(v')I(v')}{w_{sb}(v) + w_{sb}(v')}. \tag{16}$$

The division with $(w_{sb}(v) + w_{sb}(v'))$ ensures a correct normalization, which for the practical measurements can be part of the calibration procedure.

2.2.4 Spectrometer response. The signal, after amplification, is finally recorded by some kind of spectrometer. The output is a weighted mean of the signal around some discrete frequency positions, the channels, that constitutes the spectrum. For a normalized channel response, w_{ch}, the output for channel i is

$$\mathbf{y}_i = \int_0^\infty I(v) w_{ch}^i(v) \, dv \quad \text{with} \quad \int_0^\infty w_{ch}^i(v) \, dv = 1. \tag{17}$$

2.2.5 Data reduction. Data reduction is not strictly a sensor operation, but in this context it can be seen as an imaginary last part of the sensor. The aim of the data reduction is to increase the calculation speed of the inversion process, by decreasing the length of $\mathbf{y}$ without any significant loss of measurement information. Two common approaches are binning and eigenvector transformations (both linear operations), that are discussed further in section 3.6.

3. Method

3.1 *Calculation approach*

All of the instrumental equations in section 2.2 express a weighting of the radiation intensity. Inside the forward model, the intensity is known only at some discrete positions, where the primary input is the vector $\mathbf{i}$ of equation (5). The problem at hand is then to find a linear weighting of the values in $\mathbf{i}$ to mimic the instrumental expressions correctly. This observation lead to the idea (Eriksson *et al.* 2002b) to implement the sensor model $\mathcal{F}_s$ as a matrix operation (cf. equation (5)):

$$\mathbf{y} = \mathbf{H}i + \varepsilon. \tag{18}$$

An important demand for applying this equation is that the instrumental expressions are all linear operations, which should be an aim when designing an instrument. On the other hand, there exist non-linear data reduction methods that cannot be incorporated into $\mathbf{H}$.

The task of creating the response matrix $\mathbf{H}$ is facilitated by the fact that the problem can be broken down to determining the matrix for each individual sensor part:

$$\mathbf{H} = \mathbf{H}_n \dots \mathbf{H}_2 \mathbf{H}_1, \tag{19}$$

where n is the number of considered sensor parts.

If the matrix $\mathbf{H}$ could only be used to calculate $\mathbf{y}$, little would have been gained, but it can also be used in connection with other quantities. Maybe most important is the calculation of Jacobians. On the condition that $\mathbf{x}$ does not contain any sensor variables, the chain rules gives

$$\mathbf{K} = \frac{\partial \mathbf{y}}{\partial \mathbf{i}} \frac{\partial \mathbf{i}}{\partial \mathbf{x}} = H \frac{\partial \mathbf{i}}{\partial \mathbf{x}}. \tag{20}$$

However, it is often possible to apply equation (20) even if an element of $\mathbf{x}$ is a sensor parameter. For example, the Jacobian for a frequency off-set of the spectrometer (a constant shift over all channels) can be calculated by perturbing the frequencies of $\mathbf{i}$ instead of shifting the position of the spectrometer channels.

3.2 Discretization

The MPBR field and sensor responses are continuous functions, but in the computer they must be represented in some discrete manner. The functions are here assumed to be known at a set of points, the calculation grid, and are treated to vary linearly between these points. That is, a piece-wise linear description is used. The calculation strategy presented here is not limited to a piece-wise linear treatment of the functions. For example, expressions for piece-wise cubic functions have been implemented and tested, but this representation was found to be more sensitive with respect to the selection of calculation grids, and the more robust and user friendly treatment of just allowing the piece-wise linear option was selected.

It is allowed that the various grids can be defined independently of each other. This to make the usage of the forward model as straightforward as possible. An important aspect of the used instrumental expressions is that all information is preserved, and the calculations will be exact as far as the selected grids are capable of representing the corresponding continuous functions.

The sorting of elements in $\mathbf{i}$ and $\mathbf{H}$ with respect to polarization, frequency and viewing angle must of course be consistent. The (column) vector $\mathbf{i}$ is here assumed to be arranged as

$$\mathbf{i} = [I(v_1, \theta_1), Q(v_1, \theta_1), \ldots, I(v_2, \theta_1), \ldots, I(v_2, \theta_2), \ldots, V(v_k, \theta_l)]^T, \quad (21)$$

where I, Q and V are Stokes components (section 2.2.1), θ represents viewing direction, k is the length of the frequency grid, and l is the length of the viewing angle grid. The same sorting order applies for the measurement vector ($\mathbf{y}$).

3.3 Antenna and spectrometer

3.3.1 Basic expressions. The instrumental responses for the antenna and the spectrometer (equations (12) and (17)) both consist of a weighting of the signal by some instrumental response function. This corresponds to an integration of the product of the signal and instrument functions. To incorporate this weighting into the response matrix $\mathbf{H}$, we want to express the weighting as a multiplication between two vectors:

$$\mathbf{hg} = \int f(x)g(x)\mathrm{d}x, \quad (22)$$

where x is either angle or frequency, $f(x)$ is the instrument response, $g(x)$ is the spectral signal, $\mathbf{g}$ is the vector representation of $g(x)$ and $\mathbf{h}$ is a row vector. The vector $\mathbf{h}$ corresponds to elements of a row in $\mathbf{H}$, where row and element positions depend on considered response and sorting order between Stokes components, frequencies and viewing directions. It should be noted that the actual values of g are not known when creating $\mathbf{h}$, only the grid used for $\mathbf{g}$. In addition, the strength of the described method is that the same $\mathbf{H}$ can be used repeatedly, and then in conjunction with a varying $\mathbf{g}$.

The functions f and g can be represented using different grids, as described in section 3.2 and indicated in figure 1. This means that the product between f and g can have a discontinuity at each position corresponding to a grid point of either f or g. Letting k be the calculation interval index, the integral in equation (22) is
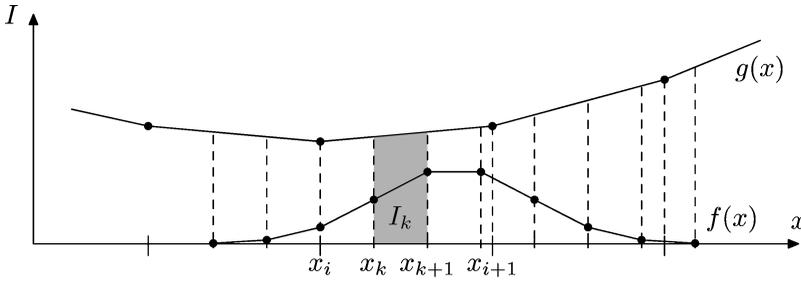
Figure 1. Schematic of antenna and backend calculations. The dashed lines show the end points of each integral interval. The shaded area corresponds to the interval considered in equation (23).

broken down as

$$\int_{x_k}^{x_{k+1}} f(x)g(x)\mathrm{d}x = \frac{x_{k+1}-x_k}{2}\left(f(x_k)g(x_k)+f(x_{k+1})g(x_{k+1})\right). \tag{23}$$

The value $g(x_k)$ is

$$g(x_k)=g(x_i)\frac{x_{i+1}-x_k}{x_{i+1}-x_i}+g(x_{i+1})\frac{x_k-x_i}{x_{i+1}-x_i}, \quad x_i \le x_k < x_{i+1}, \tag{24}$$

where i is grid index of g, and the range $[x_i, x_{i+1}]$ encompasses the range $[x_k, x_{k+1}]$. The value $g(x_{k+1})$ is determined likewise. The response function f is known and $f(x_k)$ and $f(x_{k+1})$ can be explicitly calculated (following equation (24)).

Inserting equation (24) in (23) yields the following expressions for the considered integral interval:

$$\int_{x_k}^{x_{k+1}} f(x)g(x)\mathrm{d}x =$$

$$\frac{x_{k+1}-x_k}{2}\left[\left(f(x_k)\frac{x_{i+1}-x_k}{x_{i+1}-x_i}+f(x_{k+1})\frac{x_{i+1}-x_{k+1}}{x_{i+1}-x_i}\right)g(x_i)+ \quad (25)\right.$$

$$\left.\left(f(x_k)\frac{x_k-x_i}{x_{i+1}-x_i}+f(x_{k+1})\frac{x_{k+1}-x_i}{x_{i+1}-x_i}\right)g(x_{i+1})\right],$$

and the weights to be added to element i and $i+1$ of $\mathbf{h}$ are then

$$\mathbf{h}_i=\mathbf{h}_i+\frac{x_{k+1}-x_k}{2}\left[f(x_k)\frac{x_{i+1}-x_k}{x_{i+1}-x_i}+f(x_{k+1})\frac{x_{i+1}-x_{k+1}}{x_{i+1}-x_i}\right] \tag{26}$$

$$\mathbf{h}_{i+1}=\mathbf{h}_{i+1}+\frac{x_{k+1}-x_k}{2}\left[f(x_k)\frac{x_k-x_i}{x_{i+1}-x_i}+f(x_{k+1})\frac{x_{k+1}-x_i}{x_{i+1}-x_i}\right]. \tag{27}$$

The vector $\mathbf{h}$ is initiated to hold zeros and the calculation procedure is iterated over k. To ensure energy conservation this vector should be normalized such that the sum of the elements equals one.

3.3.2 Considerations for multiple calculations. Valuable calculation time is saved by identifying when an obtained $\mathbf{h}$ vector is also valid for other frequencies or viewing directions. In the case of the antenna it can often be assumed that the

antenna response has no frequency dependency, and it is then only necessary to calculate **h** for a single frequency and apply the obtained weights for all frequencies. If a scanning procedure is performed, it is in general needed to calculate a new **h** vector for each viewing direction. An exception is if the pencil beam grid is equidistant and the sensor viewing directions throughout have the same relative position with respect to the pencil beam grid. In this special case the weights in **h** can be re-used, they just have to be moved to new positions inside the vector.

The calculations for the backend are simplified by the fact that its frequency response should seldomly be changed inside a single measurement sequence, such as a limb scan. This has the consequence that the same weighting vector can be used for all viewing directions. On the other hand, specific calculations are normally needed for each backend channel as an optimized monochromatic frequency grid is not likely to be equidistant (section 4).

3.4 *Sideband folding*

To model the sideband folding, a vector **h** representing the summation of the sideband filtered radio frequency (RF) signals into an intermediate frequency (IF) signal is to be determined. The translation of frequencies is given by equation (15). To preserve all spectral information it is needed that the IF grid includes all unique transformations from RF to IF. That is, the IF grid shall be constructed by the projection of the RF grid from both sidebands (with duplicates of IF values removed). This implies that each IF grid point corresponds directly to a RF of one of the sidebands, but will normally fall between RF grid points in the other band (figure 2), if not the RF grid is perfectly symmetric around the LO frequency.

Using the same notation as in equation (16), the elements of the response vector **h** can then be written as

$$\mathbf{h}_i = \frac{\mathbf{v}_{i+1} - v}{\mathbf{v}_{i+1} - \mathbf{v}_i} \frac{w_{\mathrm{sb}}(v)}{w_{\mathrm{sb}}(v) + w_{\mathrm{sb}}(v')}, \tag{28}$$

$$\mathbf{h}_{i+1} = \frac{v - \mathbf{v}_i}{\mathbf{v}_{i+1} - \mathbf{v}_i} \frac{w_{\mathrm{sb}}(v)}{w_{\mathrm{sb}}(v) + w_{\mathrm{sb}}(v')}, \tag{29}$$
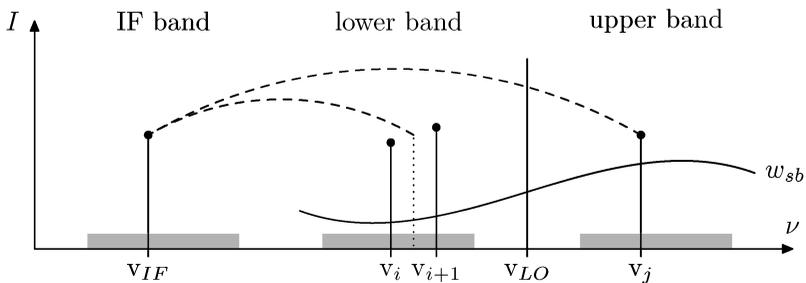
Figure 2. Schematic of sideband folding calculations. The figure shows a situation where the IF corresponds exactly to a grid position in the upper band, but not for the lower sideband. The vertical solid lines indicate spectral values, the dashed line gives the lower RF matching v_{IF}, the shaded areas show the extension of the RF and IF bands, and $w_{\mathrm{sb}}(v)$ represents the sideband filter response.

$$\mathbf{h}_j = \frac{\mathbf{v}_{j+1} - v'}{\mathbf{v}_{j+1} - \mathbf{v}_j} \frac{w_{\mathrm{sb}}(v')}{w_{\mathrm{sb}}(v) + w_{\mathrm{sb}}(v')}, \tag{30}$$

$$\mathbf{h}_{j+1} = \frac{v' - \mathbf{v}_j}{\mathbf{v}_{j+1} - \mathbf{v}_j} \frac{w_{\mathrm{sb}}(v')}{w_{\mathrm{sb}}(v) + w_{\mathrm{sb}}(v')}, \tag{31}$$

where $\mathbf{v}$ is the RF grid and all other elements of $\mathbf{h}$ are zero. Either $\mathbf{h}_{i+1}$ or $\mathbf{h}_{j+1}$ is unnecessary to compute, depending on if the IF corresponds directly to $\mathbf{h}_i$ or $\mathbf{h}_j$. The complete response matrix is computed by iterating over all IF.

The normal situation should be that the sideband folding and filtering are kept constant, and it is not needed to repeat the calculations even if the instrument is scanning. The weights of $\mathbf{h}$ from one viewing direction can for such cases be applied throughout the scan.

3.5 *Polarization*

Following equation (10), the matrix to incorporate the polarization response has the structure

$$\mathbf{H} = \frac{1}{2} \begin{bmatrix} \mathbf{pL}(\chi_1) & [0\,0\,0\,0] & \cdots & [0\,0\,0\,0] \\ [0\,0\,0\,0] & \mathbf{pL}(\chi_1) & \cdots & [0\,0\,0\,0] \\ & & \ddots & \\ \vdots & \vdots & \mathbf{pL}(\chi_2) & \\ & & & \ddots \\ [0\,0\,0\,0] & [0\,0\,0\,0] & & \mathbf{pL}(\chi_n) \end{bmatrix} \tag{32}$$

where n is the number of observation zenith angles, χ the associated rotations and the product $\mathbf{pL}(\chi_i)$ is repeated for all considered frequencies. An example on an instrument where the polarisation response varies inside the scanning procedure is AMSU.

Several sensor parts can have a polarisation varying response, but there is normally a single part that dominates the polarization response. The normal case for the instrument type considered here is that the mixer is only sensitive to a single polarization, and the natural time to apply equation (32) is then after antenna and sideband folding, but before the backend.

3.6 *Data reduction*

3.6.1 Data binning. Data binning means that neighbouring channels are combined by weighted averaging. The averaging weights are normally determined by the width of each channel, to minimize the thermal noise in the new data vector. If channels i_1 to i_2 of $\mathbf{y}'$ are combined to give element j of $\mathbf{y}$, the binning can be expressed as

$$\mathbf{y}_j = \frac{1}{\sum_{i=i_1}^{i_2} \Delta v_i} \sum_{i=i_1}^{i_2} \Delta v_i \, \mathbf{y}'_i, \tag{33}$$

where Δv_i is the (noise) width of channel i. From this expression the jth row of the

data reduction matrix, $\mathbf{H}_d$, is given as

$$\mathbf{h}^i = \frac{\Delta v^i}{\sum_{i=i_1}^{i_2} \Delta v^i}, \quad i_1 \leq i < i_2. \tag{34}$$

Other values of $\mathbf{h}$ are zero. The matrix for data binning is highly sparse.

3.6.2 Reduction by eigenvectors. A commonly used approach for reducing data sizes is to base the reduction on the eigenvectors of the covariance matrix expressing the variability of the measurements, $\mathbf{S_y}$. These empirical eigenvectors, $\mathbf{E}$, fulfils the relationship

$$\mathbf{S_y} = \mathbf{E}\Lambda\mathbf{E}^{\mathrm{T}}, \tag{35}$$

where Λ is a diagonal matrix of the eigenvalues. Using the j first eigenvectors the data reduction is performed as

$$\mathbf{y} = \mathbf{E}_j^{\mathrm{T}}\mathbf{y}, \tag{36}$$

that is

$$\mathbf{H}_d = \mathbf{E}_j^{\mathrm{T}}. \tag{37}$$

Different versions of this approach are discussed in Eriksson *et al.* (2002a).

4. Simple grid optimization

The calculation speed can depend critically on the length of the grids used to represent the involved functions. The importance to select shortest possible grid and for which quantities this is most crucial, differ depending on details of the radiative transfer problem to solve and how many times the calculations will be repeated. In the case of atmospheric measurements, the general situation is that most time can be saved by carefully selecting the grid for monochromatic frequencies. A dense monochromatic grid is needed around the positions of molecular transitions, while other parts of the spectrum is comparably flat and a coarse grid is sufficient. If an equidistant grid is used, the grid spacing must be selected in such way that spectra are properly represented around the sharp transitions and a very long grid will be obtained. A more carefully selected grid will be order of magnitudes shorter and it thus of interest to have a method to automatically select this and other grids.

The instrumental expressions are here solved without introducing any approximations beside that the involved functions are presented in a discrete manner. This means that the grid for a particular function can be selected by only considering the shape and variation of the corresponding function. The only relevant task of the grid is to represent its function with sufficient accuracy, it is not needed to consider how the calculations are performed when selecting the grids. This fact makes it easy to design algorithms and general functions for selecting the grids. We have adopted the following simple and straightforward algorithm.

1. Obtain a set of realizations for the function of interest. This for a grid sufficiently fine that discretisation errors can be neglected. The fine grid is here denoted as x_f, and function values as y_f.
2. Initiate output grid, x_c, to only include end points of fine grid.
3. Create coarse representation of the function by setting $y_c(x_c) = y_f(x_c)$.

4. Interpolate coarse representation to the fine grid and calculate difference to original values: $\Delta y(x_f) = y_c(x_f) - y_f(x_f)$.

5. If $\max(\Delta y)$ is smaller than set accuracy limit, return x_c. Otherwise, include the position where Δy is max in x_c and go to 3.

As commented below, it can be needed that y_f includes a set of function realizations. In this case, y_f can be seen as a matrix and the maximum of Δy is calculated globally for the matrix. The interpolation in step 4 shall correspond to the discretization strategy used. For example, a piece-wise linear treatment of the functions is here used and a linear interpolation shall then be applied. The accuracy limit (in step 5) can be defined in various ways, where either absolute or relative units can be used. The algorithm is defined likewise for multi-dimensional functions. A practical result of the algorithm is given in figure 3.

The algorithm above can, of course, only adapt the grid to features present in y_f. It is not always possible to incorporate all possible shapes of the function in a single realisation and such cases are handled by including several realisation in y_f. For example, using atmospheric limb sounding as example, a natural choice is to consider spectra for different tangent altitudes, to make sure that the grid is capable of representing spectra with different levels of pressure broadening (see figure 3).

5. An example

To illustrate the structure of the different sensor response matrices, a simplified configuration for the sub-mm receiver on-board the Odin satellite, Odin-SMR, is used. In short, Odin-SMR is a limb scanning instrument consisting of a parabolic antenna, devices for sideband filtering, four mixers and three spectrometers (Murtagh *et al.* 2002). Polarization response and data reduction are not considered here. Odin-SMR has several observation modes and for this example a band around 501.4 GHz (figure 3) from the stratospheric mode is selected. The limb sequence of this mode corresponds to tangent altitude distances of 1.5 km between adjacent spectra, where a scanning between 12 and 60 km is here assumed.

5.1 *Antenna*

The response of the Odin-SMR antenna is here modelled as a truncated Gaussian function. The structure of the antenna matrix, $\mathbf{H}_a$, is shown in figure 4. Each diagonal line in the inlay image corresponds to a pencil beam direction, and each row corresponds to a monochromatic frequency. It can be seen that in this simplified example only five pencil beam directions are used to estimate the angular variation of the radiance field over the range covered by the antenna

5.2 *Mixer*

The primary band of the receiver is 501.170–501.592 GHz and the LO frequency is here placed at 497.904 GHz. The matrix representing the sideband folding and filtering responses, $\mathbf{H}_m$, is shown in figure 5. The upper, and primary, band is covered by 174 monochromatic frequencies selected by the algorithm of section 4, with 0.005 K as accuracy limit. The image range lacks significant transitions and a low number of monochromatic frequencies would have been sufficient to represent the flat shape of the spectrum in that range, where it also shall be considered the relative contribution of this band is $<2\%$. An equidistant grid with 20 points was

here selected to let the image band appear clearly in figure 5. The mixer and sideband filtering responses are constant during the limb scan and the weights obtained for the first zenith angle could be applied for all viewing directions.

5.3 *Spectrometer*

This Odin-SMR mode uses only the two auto-correlation spectrometers (there is also an acousto-optic spectrometer). The spacing between channels for these spectrometers is 1.14 MHz. A Hamming window is normally applied for the transformation from auto-correlation values to the frequency spectrum, and this results in a poorer resolution. The response of each backend channel is here
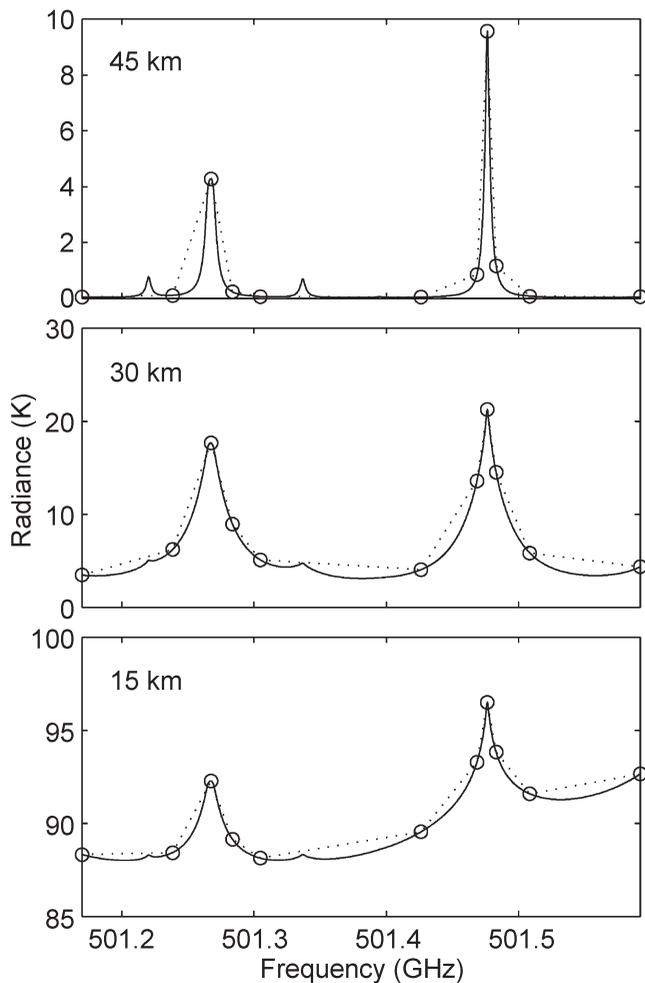
Figure 3. Monochromatic frequency grid selected by the algorithm described in section 4. Example valid for limb sounding using 501.17–501.592 GHz, where spectra for tangent altitudes 15, 30 and 45 km are considered in parallel. The accuracy limit is set to 3 K, a value not used in practice and only selected for pedagocial reasons. Solid lines are the fine grid calculations, circles give the spectral values at the selected grid positions and the dashed lines show the piece-wise linear representation of spectra for obtained grid.
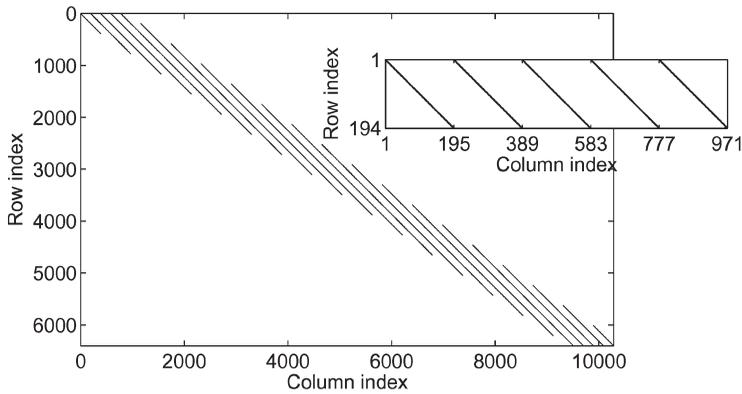
Figure 4. The structure of the response matrix for antenna pattern. The position of nonzero elements are marked. The inlay image is a magnifaction showing the structure for the first zenith angle.

approximated by a truncated Gaussian function, with a width of 2.0 MHz. The resulting backend response matrix, $\mathbf{H}_b$, is shown in figure 6. The flatter parts of the curve in the inlay image correspond to ranges where the grid for monochromatic frequencies is finer, to better capture the central parts of the two main spectral features of the band (figure 3). It is interesting to note that the automatic grid selection of section 4 allows that the number of monochromatic frequencies (192) is lower than the number of backend channels (400), despite that a high calculation accuracy was requested (section 5.2). This is the case as Odin-SMR has a low and constant spacing between the channels over the band, and that this part of the spectrum includes few spectral features. The backend response is not dependent on zenith angle.
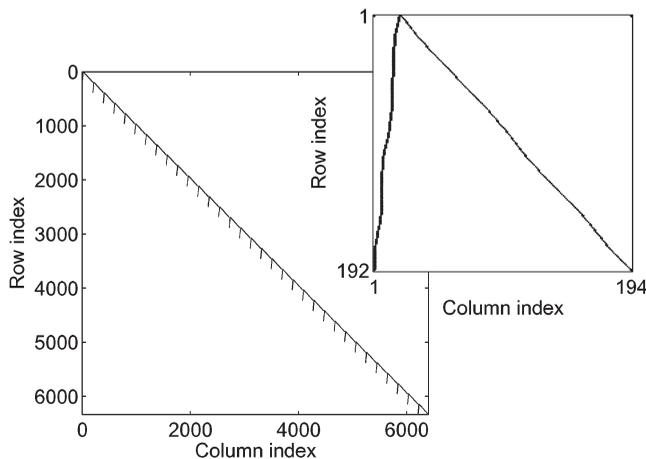
Figure 5. The structure of the response matrix for sideband folding and filtering. The position of nonzero elements are marked. The inlay image is a magnifaction showing the structure for the first zenith angle.
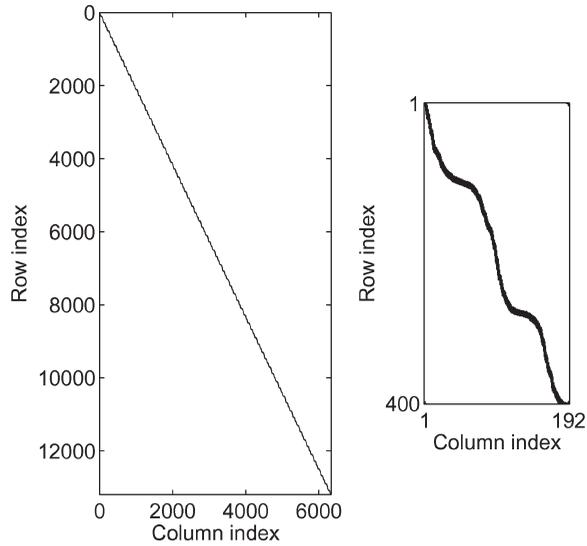
Figure 6.   The structure of the response matrix for backend. The position of nonzero elements are marked. The inlay image is a magnifaction showing the structure for the first zenith angle.

5.4   *Total sensor response*

The total response matrix is obtained by multiplying the individual response matrices:

$$\mathbf{H} = \mathbf{H}_b \mathbf{H}_m \mathbf{H}_a, \tag{38}$$

and its structure is shown in figure 7.
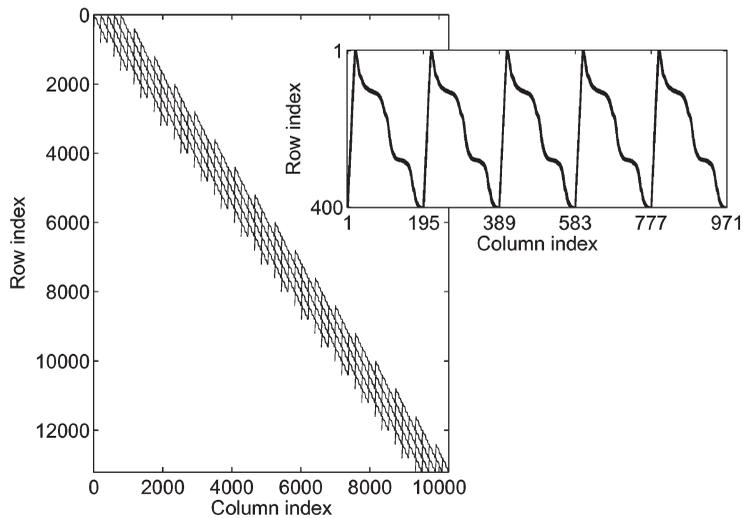
Figure 7.   The structure of the response matrix for the total instrument. The position of nonzero elements are marked. The inlay image is a magnifaction showing the structure for the first zenith angle.

6.   Conclusions

A highly efficient method for including sensor impacts in forward model calculations has been presented. The efficiency of the method depends on several factors, but it can be several orders of magnitude faster than traditional implementations. The matrix responses are converted to weights for each Stokes component of all considered monochromatic pencil beam radiances. The weights are stored in the form of a matrix and the sensor part of the forward model is simplified to a matrix multiplication. The complete sensor matrix is created by determining the response matrix for each sensor part, and multiplicating these individual matrices. Expressions are given to determine the response matrix for the antenna, sideband mixing and folding, polarization characteristics and spectrometer. The presented scheme can be extended to incorporate data reduction, and expressions for binning and eigenvector reduction are given.

The dimension of the weight matrices can be very large. For example, the size of the complete matrix for a single Odin-SMR (the sensor used in this paper as example) band is 12000×18000, or larger. However, the sensor response matrices are highly sparse and the usage of computer memory can be kept at a reasonable level, as long as a sparse matrix data type is available. Data reduction based on eigenvector expansion results in a matrix filled with non-zero values, but this is compensated by the fact that a much smaller matrix dimension is then obtained.

A piece-wise linear treatment of the involved functions has been assumed here, but the described approach can be extended regarding this respect, such as using a spline representation for all or some functions. A more advanced function treatment could allow that a shorter grid is used, but the grid must then be selected more carefully and the creation of the sensor matrices will require more computational time. Such an option is of highest interest for the pencil beam and monochromatic frequency grids as if these grids are made shorter, this results in fewer radiative transfer calculations.

The presented method has been implemented in two different versions around the Atmospheric Radiative Transfer System (ARTS) forward model. For the operational version, ARTS-1.0.x (Buehler *et al.* 2004), the sensor modelling is made by a set of Matlab functions, distributed together with ARTS. A new ARTS version with several major extensions is under development, and the sensor modelling has been ported to the C++ code as part of this development. More information and both ARTS versions can be obtained at www.sat.uni-bremen.de/arts.

References

BUEHLER, S.A., ERIKSSON, P., KUHN, T., VON ENGELN, A. and VERDES, C., 2005, ARTS, the Atmospheric Radiative Transfer Simulator. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **91**, pp. 65–93.

ERIKSSON, P., JIMÉNEZ, C., BÜHLER, S. and MURTAGH, D., 2002a, A Hotelling transformation approach for rapid inversion of atmospheric spectra. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **73**, pp. 529–543.

ERIKSSON, P., MERINO, F., MURTAGH, D., BARON, P., RICAUD, P. and DE LA NOË J., 2002b, Studies for the Odin sub-millimetre radiometer: 1. Radiative transfer and instrument simulation. *Canadian Journal of Physics*, **80**, pp. 321–340.

LIOU, K.N., 2002, *An Introduction to Atmospheric Radiation*, 2nd edn (London: Academic Press).

MURTAGH, D., FRISK, U., MERINO, F., RIDAL, M., JONSSON, A., STEGMAN, J., WITT, G., ERIKSSON, P., JIMÉNEZ, C., MEGIE, G., DE LA NOË J., RICAUD, P., BARON, P.,

PARDO, J.R., HAUCHCORNE, A., LLEWELLYN, E.J., DEGENSTEIN, D.A., GATTINGER, R.L., LLOYD, N.D., EVANS, W.F.J., McDADE, I.C., HALEY, C.S., SIORIS, C., VON SAVIGNY, C., SOLHEIM, B.H., MCCONNELL, J.C., STRONG, K., RICHARDSON, E.H., LEPPELMEIER, G.W., KYRË, E., AUVINEN, H. and OIKARINEN, L., 2002, An overview of the Odin atmospheric mission. *Canadian Journal of Physics*, **80**, pp. 309–319.

RODGERS, C.D., 2000, *Inverse Methods for Atmospheric Sounding: Theory and Practise* (Singapore: World Scientific Publishing).

SMILES: SCIENCE TEAM, 2001, JEM/SMILES Mission Plan, Version 1.0.

URBAN, J., BARON, P., LAUTIÉ, N., SCHNEIDER, N., DASSAS, K., RICAUD, P. and DE LA NOË, 2004, Moliere (v5): a versatile forward- and inversion model for the millimeter and sub-millimeter wavelength range. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **83**, pp. 529–554.

ZORN, S., VON CLARMANN, T., ECHLE, G., FUNCKE, B., HASE, F., HÖPFNER, M., KEMNITZER, H., KUNTZ, M. and STILLER, G.P., 2000, Analytical expressions for modeling of radiative transfer and instrumental effects in KOPRA. In *The Karlsruhe Optimized and Precise Radiative Transfer Algorithm (KOPRA)*, G.P. Stiller, (Ed.), Forschungszentrum Karlsruhe, Wissenschaftliche Berichte, Bericht Nr. 6487, pp. 9–28.