Fast and Stable Solution Method for Angle-Resolved Light Scattering Simulation II – Model Enhancements

Per Edström
Marcus Lehto
Mid Sweden University

FSCN Report – ISSN 1650-5387 2003:17
Internal FSCN Report Number: R-03-42
April 2003
Contents

ABSTRACT ................................................................................................................... 3

1. INTRODUCTION ..................................................................................................... 4
   1.1. SOME NOTES AND REFERENCES ON RADIATIVE TRANSFER ......................... 4
   1.2. SOME COMMENTS ON THIS PRESENTATION ...................................................... 4

2. ENHANCEMENTS IN DORT2002, VERSION 2.0 ..................................................... 6
   2.1. GRAPHICAL USER INTERFACE AND MANUAL ..................................................... 6
   2.2. MULTILAYER STRUCTURES ................................................................................... 6
   2.3. THE δ - N METHOD ............................................................................................... 6
   2.4. INTENSITY CORRECTION PROCEDURES ............................................................. 6
   2.5. BREAKING THE AZIMUTHAL LOOP ..................................................................... 7
   2.6. AUTOMATIC CONTROL OF NEEDED NUMBER OF PHASE FUNCTION MOMENTS .... 7
   2.7. USER DEFINED PHASE FUNCTIONS ..................................................................... 7
   2.8. USER DEFINED ANGULAR RESOLUTION ............................................................... 7
   2.9. LOWER BOUNDARY CONDITIONS ...................................................................... 7
   2.10. OPTIMIZED CODE ............................................................................................. 8
   2.11. COMPUTATIONAL SHORTCUT ............................................................................ 8
   2.12. ENHANCED PRESENTATION OF RESULTS ......................................................... 8
   2.13. MORE OUTPUT VARIABLES .............................................................................. 9
   2.14. MORE PLOT OPTIONS ....................................................................................... 9

3. SOLUTION METHOD ............................................................................................... 10
   3.1. MULTILAYER SOLUTION ...................................................................................... 10
   3.2. MULTILAYER INTERPOLATION FORMULA ............................................................. 12
   3.3. THE δ - N METHOD ............................................................................................. 16
   3.4. INTENSITY CORRECTION PROCEDURES ............................................................. 17
      3.4.1. The TMS Method ........................................................................................... 18
      3.4.2. The IMS Method .......................................................................................... 20
   3.5. BREAKING THE AZIMUTHAL LOOP ..................................................................... 24

4. SUGGESTIONS FOR FUTURE WORK ..................................................................... 28
   4.1. MODEL DEVELOPMENT ...................................................................................... 28
   4.2. INVERSE PROBLEM ........................................................................................... 29

5. DISCUSSION .......................................................................................................... 30
   5.1. ASPECTS OF SCIENTIFIC COMPUTING ............................................................... 30
   5.2 THE IMPACT OF DORT2002 ON THE PAPER AND PRINTING INDUSTRIES ........ 30

6. ACKNOWLEDGEMENTS ....................................................................................... 33

APPENDIX 1 – A SIMPLIFICATION ........................................................................ 34
APPENDIX 2 – A SIMPLIFYING APPROXIMATION .............................................. 37
APPENDIX 3 – A SIMPLIFYING APPROXIMATION .............................................. 38
APPENDIX 4 – SOME RADIOMETRIC QUANTITIES ............................................. 40

REFERENCES ......................................................................................................... 44
Fast and Stable Solution Method for Angle-Resolved Light Scattering Simulation II – Model Enhancements

Per Edström
Marcus Lehto
Mid Sweden University, S-871 88 Härnösand

Abstract

This report continues previous work to extensively treat a solution method to the radiative transfer problem. Several enhancements have been made, including handling of multilayer structures, handling of sharply peaked phase functions, and code optimization. The resulting model is referred to as DORT2002, now in version 2.0, and has been implemented in MATLAB using a discrete ordinate model geometry. The introduction of a graphical user interface makes the model easy to use, and any desired simulation is just a button-click away.

The main steps to get a numerically stable multilayer solution procedure include the preconditioning of the system of equations for the boundary and continuity conditions, and the avoidance of over- and underflow in the multilayer solution and interpolation formulas. Important are also the recognition of potential divide-by-zero situations, and reformulation of those.

Several measures are taken to make the code fast. This includes exploitation of the sparse structure of the system of equations for the boundary and continuity conditions. Several features allow high speed by maintaining accuracy at significantly lower number of channels than would otherwise be needed, or by automatically stopping calculations earlier when certain convergence criteria have been met.

Comments are given on possible applications in the paper and printing industries. It is suggested that the Kubelka-Munk model should be replaced with DORT2002 in most applications.
1. Introduction

Models for calculating the light intensity inside an illuminated turbid medium, e.g. paper, involve several numerical challenges, and are crucial for the paper and printing industries. The general problem, with several different applications, is known as the radiative transfer problem. The problem was for long considered intractable because of numerical difficulties. This report continues to describe a fast and numerically stable solution procedure, which has been implemented in MATLAB.

This report is not meant to be read on its own. It is, as the title suggests, the direct continuation of the report Fast and Stable Solution Method for Angle-Resolved Light Scattering Simulation, by P. Edström [27]. The solution model DORT2002 presented in the previous report has been substantially enhanced, and that work is reported here.

Section 1 begins with some notes and references, in section 2 the enhancements in version 2.0 of the DORT2002 model are reviewed, and in section 3 the corresponding parts of the solution method are worked out with all important numerical steps. Suggestions for future work are given in section 4, and section 5 discusses some aspects of scientific computing and the impact on the paper and printing industries, part of which is repeated from [27] for completeness. Some algebraic details are given in appendices, together with relations between some radiometric quantities.

1.1. Some Notes and References on Radiative Transfer

Radiative Transfer theory describes the interaction of radiation with scattering and absorbing media. Solution methods for radiative transfer problems have been studied during the last century [1-10]. In the beginning most radiative transfer problems were considered intractable. Therefore coarse approximations were used, and methods developed slowly due to the lack of mathematical tools. Among the solution methods in use today are Discrete Ordinate methods.

Stamnes et al [12, 13] reported on a stable Discrete Ordinate algorithm, and later provided a complete software, DISORT. Thomas and Stamnes [14] also wrote a textbook on radiative transfer in the atmosphere.

1.2. Some Comments on This Presentation

This report, as the previous, is very mathematical. It is so because the model development involves a large amount of mathematical deduction and many ideas from scientific computing. Work is currently in progress that will present the DORT2002 tool to a broader audience, e.g. the paper and printing industry. This includes a manual to the DORT2002 Graphical User Interface, a tool developed to provide users with a fast and easy way of performing DORT2002 simulations. It works as a shell that encapsulates the parameters and the function calls, and offers powerful simulations through a mouse click.

This presentation owes much to that of Thomas and Stamnes, and uses mostly the same notation. However, some adaptation from atmospheric applications to the field of paper and print has been done.
Since it is in both cases a radiative transfer problem, the basic problem is identical in the two areas, so the same outline, notation and methods have been used where applicable. On the other hand, some phenomena in the atmosphere are uninteresting in paper, and have been omitted, and other phenomena have been (or are prepared to be) added. Also, a number of errors of Thomas and Stamnes have been corrected.

The implementation in MATLAB is entirely new. It uses the same textbook ideas in many cases, but has been implemented from scratch, and many parts are entirely differently implemented. Numerical stability has been improved by reformulation of certain formulas. Some effort has been made to use matrix formulations, not only for the parts naturally expressed as matrices. This has been done to make use of MATLAB’s excellent matrix handling. MATLAB 6.5 incorporates LAPACK, which is state-of-the-art today, and these routines have been used for standard procedures. MATLAB 6.5 also has a JIT accelerator that makes it as fast as compiled languages like Fortran or C.

All necessary definitions are given in the previous report [27]. Also, frequent references are done to equations in that previous report. In fact, these two reports should be seen, and will probably be used, as one unit. Therefore, the equation numbering in this report starts from where the previous report ended. This facilitates easy references within this report, as well as references to these two reports from other sources. Of the same reason, the reference sections in the two reports are identical, although most references are only used in the first report. The introduction and problem statement sections of the previous report are applicable to this report as well, and this report essentially continues the solution method section of the previous report.
2. Enhancements in DORT2002, Version 2.0

In version 2.0 of DORT2002 several enhancements have been made, both through introduction of several new features, and through improvements of existing ones. Together these enhancements extend the applicability of DORT2002 considerably. They also reduce the computation time for all problems. Certain moderate problems are solved 10 – 1000 times faster, and for large problems the improvement is even far bigger. In this section the most important enhancements are reviewed.

2.1. Graphical User Interface and Manual

A graphical user interface has been developed to facilitate easy use of DORT2002, and to give a convenient overview over model parameters. It allows easy parameter manipulation, including saving and loading parameters to and from files. It makes it easy to configure multilayer structures, to choose which results to calculate and which graphs to plot, and to save the results to files. It is still possible to run the program from the command prompt, or to embed it in a larger program.

A manual has been worked out to make the program easy to use. The manual describes how the graphical user interface works, and how to call DORT2002 from the command prompt or as an embedded component. It also gives references to the reports describing the theory behind DORT2002.

2.2. Multilayer Structures

DORT2002 has been enhanced to include multilayer structures. This makes it possible to handle discrete multilayer structures, but also vertically inhomogeneous media by approximation with a sufficiently large number of adjacent homogeneous layers. This is one of the major improvements of DORT2002, version 2.0, and the theory behind it is described in a separate section in this report.

2.3. The δ - N Method

A transformation procedure, the δ-N method, has been implemented to allow handling of strongly forward peaked phase functions (g close to 1) with maintained accuracy without a tremendously increased computational burden. The δ-N method also gives maintained accuracy for all g for significantly lower N than otherwise needed. However, the closer g is to zero, the smaller N is needed anyway, so the savings in computation time diminishes with decreasing g. The overhead introduced by the method is insignificant compared to the core calculations. This is one of the major improvements of DORT2002, version 2.0, and the theory behind it is described in a separate section in this report.

2.4. Intensity Correction Procedures

Intensity correction procedures have been implemented to further enhance the handling of strongly forward peaked phase functions beyond the capabilities of the δ-N method. These procedures also give maintained accuracy for all g for significantly lower N than otherwise needed. However, the closer g is to zero, the smaller N is needed anyway, so
at some point the possible savings in computation time are smaller than the overhead introduced by the correction procedure. The intensity correction procedures are therefore automatically turned off when they are not needed, in order to save computation time. This is one of the major improvements of DORT2002, version 2.0, and the theory behind it is described in a separate section in this report.

2.5. Breaking the Azimuthal Loop
A method has been implemented that breaks the azimuthal loop when a convergence criterion has been met. This saves a tremendous amount of computation time in the vast majority of cases, since in most cases the intensity converges well before the azimuthal loop has ended. Since it is the outermost loop, much is gained if it can be terminated earlier. This is one of the major improvements of DORT2002, version 2.0, and the theory behind it is described in a separate section in this report.

2.6. Automatic Control of Needed Number of Phase Function Moments
The intensity correction procedures use a larger number of phase function moments than the core calculations. That larger number of phase function moments can be chosen by the user, but it can also be automatically controlled. The automatic control chooses the number of phase function moments to achieve a predefined high accuracy.

2.7. User Defined Phase Functions
It is possible for the user to define own phase functions in addition to the built-in Henyey-Greenstein phase function. This is done through their phase moments, i.e. their coefficients in Legendre polynomial expansion. This extends the applicability of the model considerably, since the angular dependence of light scattering can deviate significantly from the Henyey-Greenstein case. Phase moments from many media are readily available in tabulated form. Phase moments can also be computed in advance for any case that has a mathematical model for the scattering process. The graphical user interface makes it possible to use libraries of phase functions, stored as vectors of moments in mat-files.

2.8. User Defined Angular Resolution
It is possible for the user to define output azimuthal and polar angles as well as depths. It is worth noting that the user supplied output polar and azimuthal angles are entirely decoupled from the predefined channels in the core calculations, and a high angular resolution does not require a large number of channels. In fact, it is one of the main features of DORT2002 to offer high accuracy and resolution at a small number of channels, thus giving a large decrease in computation time.

2.9. Lower Boundary Conditions
Two types of lower boundary conditions are supported in DORT2002. One is a perfectly diffuse underlying surface with variable normal reflectance, and the other is no surface at all. The previously mentioned Minnaert and Lommel-Seeliger formulas
are not relevant here, and using Henyey-Greenstein for the bottom surface with \( g \neq 0 \) has proven not to fulfill normalization conditions.

Note that the lower boundary conditions do not describe the lower surface of the medium, but the surface of any underlying medium. This is closely related to \( R_s \) in the Kubelka-Munk theory.

### 2.10. Optimized Code

The DORT2002 code has been optimized in many ways, and various small bugs have been removed. Computational shortcuts have been added, program statements have been rewritten in a more efficient way, the code is vectorized where possible, preallocation of large variables has been introduced, and variable types and memory handling has been looked over. A great effort has been put into handling and exploiting the sparse structure of the systems of equations for the boundary and continuity conditions.

### 2.11. Computational Shortcut

A computational shortcut has been implemented to allow for much faster calculation of variables that depend only on the azimuthally averaged intensity. These variables are total reflectance, total transmittance, total absorptance, BRDF and BTDF. If only variables from this group are required, DORT2002 breaks the azimuthal loop after the first time instead of fulfilling the prescribed \( 2^N \) times, thus giving a large decrease in computation time.

### 2.12. Enhanced Presentation of Results

The results are available in three different forms from the graphical user interface. They can be given as a struct in the global variable `dort2002_output` for immediate use in MATLAB. The results can be saved in a mat-file for use in a later MATLAB session. The results can also be saved in a formatted text-file, which is useful for reading and for written presentations.

If DORT2002 is used without the graphical user interface, the output is returned in a variable as from any MATLAB function: `result = dort2002(parameters)`, which is especially useful if DORT2002 is embedded in a larger program.
2.13. More Output Variables
Some new output variables have been added, and DORT2002 now offers calculation of:
- total reflectance;
- total transmittance;
- total absorptance;
- Azimuthally averaged BSDF at any polar angles;
- BSDF at any polar and azimuthal angles;
- Fourier components of the intensity at any polar angles at any depths;
- azimuthal components of the intensity at any polar and azimuthal angles at any depths;
- remainder of the incident beam at any depths.

2.14. More Plot Options
Some new plot options have been added, and DORT2002 now offers the following plots:
- line plot of azimuthally averaged BSDF at any polar angles;
- 3D plot of BSDF at any polar and azimuthal angles;
- mesh plot of Fourier components of the intensity at any polar angles at any depths;
- mesh plot of azimuthal components of the intensity at any polar and azimuthal angles at any depths;
- 3D plot of azimuthal components of the intensity at any polar and azimuthal angles at any depths.
3. Solution Method

3.1. Multilayer Solution

The solutions found so far pertain to a single, vertically homogenous layer. There are at least two reasons for considering multilayer structures. One is that the medium might in fact be constructed as several discrete and vertically homogenous layers placed on top of each other. Another is that an inhomogeneous medium can be approximated with a (sufficiently large) number of homogeneous layers.

If the optical properties of the medium are a function of optical depth, the medium can be divided into a number of adjacent layers, each of which having constant optical properties. The continuous variation of the optical properties of the medium can thus be approximated as closely as desired by choosing a sufficiently large number of layers.

Since each of the layers in the multilayer structure is homogeneous – whether it is a real discrete structure or an approximation of a continuously varying one – the previously derived solutions can be used. In analogy with equation (125), the solution for the \( p \)th layer can be written as

\[
I_p^\pm (\tau, \mu) = \sum_{j=1}^{N} C_{jp} g_{jp} (\pm \mu_j) e^{-k_{zp} \tau} + C_{jp} g_{-jp} (\pm \mu_j) e^{+k_{zp} \tau} + U_p^\pm (\tau, \mu),
\]

where \( L \) is the number of layers. The only difference from equation (125) is the addition of the layer index \( p \). In analogy with the single layer case, \( k_{jp} > 0 \) and \( k_{-jp} = -k_{jp} \).

The multilayer solution contains \( 2N \times L \) constants to be determined. In addition to the previous boundary conditions, the intensity must now be required to be continuous across layer interfaces. This gives the following system of equations to determine the unknown constants:

\[
\begin{aligned}
I_i^m (0, -\mu_i) &= \Upsilon^m (-\mu_i), \quad i = 1, \ldots, N, \\
I_p^m (\tau_p, \mu_i) &= I_{p+1}^m (\tau_p, \mu_i), \quad i = \pm 1, \ldots, \pm N, \quad p = 1, \ldots, L - 1, \\
I_L^m (\tau_L, +\mu_i) &= (1 + \delta_m) \sum_{j=1}^{N} \omega_j \mu_j \rho_d^m (-\mu_j, \mu_i) I_{L-1}^m (\tau_L, -\mu_j) + \\
&\quad + \frac{\mu_0}{\pi} \rho_d^m (-\mu_0, \mu_i) I_{L-1} e^{-\tau_L / \mu_0}, \quad i = 1, \ldots, N
\end{aligned}
\]

As before, \( \Upsilon^m (-\mu_i) \) is the incident radiation at the top boundary. For perfectly diffuse incident radiation, \( \Upsilon^m (-\mu_i) \) is constant for \( m = 0 \), and zero for \( m \neq 0 \).

Inserting the multilayer solution into this system of equations gives (now omitting the superscript \( m \))
\[
\sum_{j=1}^{N} \left( C_{j+1} g_{j+1}(-\mu_i) + C_{j-1} g_{j-1}(-\mu_i) \right) = \gamma(-\mu_i) - U_j(0, -\mu_i), \quad i = 1, \ldots, N
\]
\[
\sum_{j=1}^{N} \left( C_{p+1} g_{p+1}(-\mu_i) + C_{p-1} g_{p-1}(-\mu_i) \right) =
\begin{align*}
&= \left( C_{p+1} g_{p+1}(-\mu_i) + C_{p-1} g_{p-1}(-\mu_i) \right) = \Gamma(\tau_L, \mu_i), \quad i = 1, \ldots, N
\end{align*}
\]

where
\[
r_j(\mu_i) = g_{jL}(\mu_i) - (1 + \delta_{m0}) \sum_{n=1}^{N} \rho_d(-\mu_n, \mu_i) \omega_n \mu_n g_{jL}(-\mu_n)
\]
and
\[
\Gamma(\tau_L, \mu_i) = -U_L^+(\tau_L, \mu_i) + (1 + \delta_{m0}) \sum_{j=1}^{N} \rho_d(-\mu_j, \mu_i) \omega_j \mu_j U_L^-(\tau_L, \mu_j) +
\]
\[
\frac{H_0}{\pi} \rho_d(-\mu_0, \mu_i) I_{00} e^{\tau_L/\mu_0}
\]

The boundary and continuity conditions constitute a \((2N \times L) \times (2N \times L)\) system of equations for the \(2N \times L\) unknown coefficients \(C_{pL}, \quad j = \pm 1, \pm 2, \ldots, \pm N, \quad p = 1, \ldots, L\). The coefficient matrix is sparse and block diagonal, with \(6N-1\) diagonals. The blocks on the diagonal lead with a \(3N \times 2N\) block, then follow \(L-2\) blocks of size \(4N \times 2N\), and the diagonal ends with a \(3N \times 2N\) block.

As in the single layer case, the equations are ill conditioned due to the exponentials with positive arguments. The ill conditioning can be removed, however, with the scaling transformation
\[
C'_{pL} = C_{pL} e^{k_p \tau_L} \quad \text{and} \quad C'_{-pL} = C_{-pL} e^{-k_p \tau_L}.
\]

The scaled system of equations for the coefficients \(C'_{pL}\) then becomes (with \(\tau_0\) as the optical depth at the top).
\[
\sum_{j=1}^{N} \left( C'_{j,i} g_{j,1}(-\mu_j) + C'_{j,i} g_{j,-1}(-\mu_j)e^{-k_\mu (r_n - r_1)} \right) = \Upsilon(-\mu_i) - U_1(\tau_0, -\mu_i) \quad i = 1, \ldots, N
\]
\[
\sum_{j=1}^{N} \left( C'_{j,p} g_{j,p}(\mu_j) e^{-k_\mu (r_p - r_{p-1})} + C'_{j,p} g_{j,-p}(\mu_j) \right) = U_{p+1}(\tau_{p+1}, \mu_i) - U_p(\tau_p, \mu_i) \quad i = \pm 1, \ldots, \pm N, \quad p = 1, \ldots, L - 1
\]
\[
\sum_{j=1}^{N} \left( C'_{j,1} r_{j,1}(\mu_j) e^{-k_\mu (r_1 - r_{n-1})} + C'_{j,1} r_{j,-1}(\mu_j) \right) = \Gamma(\tau_L, \mu_i) \quad i = 1, \ldots, N
\]

Since \( k_p > 0 \) and \( \tau_p > \tau_{p-1} \), \( p = 1, \ldots, L \), all exponentials in the system of equations for the coefficients \( C'_{jp} \) have negative arguments. Thus, the ill conditioning is avoided, and the problem of solving for the \( C'_{jp} \) is unconditionally stable.

There is a risk for overflow in the solution for the \( p \)th layer, but by using the same scaling, that solution becomes
\[
I^+_p(\tau, \mu_i) = \sum_{j=1}^{N} \left( C'_{j,p} g_{j,p}(\pm \mu_j) e^{-k_\mu (r_p - r_{p-1})} + C'_{j,p} g_{j,-p}(\pm \mu_j) e^{-k_\mu (r_p - r_1)} \right) + U^+_p(\tau, \mu_i), \quad p = 1, 2, \ldots, L.
\]

Since \( k_p > 0 \) and \( \tau_{p-1} < \tau < \tau_p \), all exponentials have negative arguments, and the risk of overflow is avoided.

### 3.2. Multilayer Interpolation Formula

As in the single layer case, an interpolation formula for the intensity can be worked out, which makes it possible to calculate the intensity in an arbitrary direction, and not only in the quadrature points. Thus, the first equation in (133) is integrated layer by layer from \( \tau \) to \( \tau_L \), using \( e^{-t/\mu} \) as integrating factor, and the second equation in (133) is integrated layer by layer from \( \tau_0 \) to \( \tau \), using \( e^{\tau/\mu} \) as integrating factor. This gives
\[
\begin{align*}
I^+_p(\tau, \mu) &= I^+_1(\tau_L, \mu) e^{-(r_1 - r_p)/\mu} + \frac{1}{\mu} \sum_{\mu \neq \mu_p \tau_{n-p}}^{t} \int S^+_{n}(t, \mu) e^{-(t-r)/\mu} dt \\
I^-_p(\tau, \mu) &= I^-_0(\tau_0, \mu) e^{-(r_n - r_p)/\mu} + \frac{1}{\mu} \sum_{n=1}^{n=p} \int S^-_{n}(t, \mu) e^{-(t-r)/\mu} dt
\end{align*}
\]
with \( \tau_{n-1} \) replaced by \( \tau \) for \( n = p \) in the first equation and \( \tau_n \) replaced by \( \tau \) for \( n = p \) in the second equation. In both equations \( \tau_{p-1} \leq \tau \leq \tau_p \) and \( \mu > 0 \).

As in the single layer case, the source functions (135) can be used, but here with the proper layer indexing. Substituting the multilayer solution into that expression for the source functions yields.
\[ S_p^\pm (\tau, \mu) = \sum_{j=1}^{N} C_{j-p} \tilde{g}_{j-p} (\pm \mu) e^{\frac{-\lambda_j \tau}{\mu}} + \sum_{j=1}^{N} C_{j-p} \tilde{g}_{j-p} (\pm \mu) e^{\frac{-\lambda_j \tau}{\mu}} + \tilde{Z}_{0p}^\pm (\mu) e^{-\tau / \mu_0}, \]  

(163)

where

\[ \tilde{g}_{j-p} (\pm \mu) = \frac{a}{2} \sum_{i=1}^{N}(\omega_i p(-\mu_i, \pm \mu)g_{j-p} (-\mu_i) + \omega_i p(+\mu_i, \pm \mu)g_{j-p} (+\mu_i)) \]  

(164)

and

\[ \tilde{Z}_{0p}^\pm (\mu) = \frac{a}{2} \sum_{i=1}^{N}(\omega_i p(-\mu_i, \pm \mu)Z_{0p} (-\mu_i) + \omega_i p(+\mu_i, \pm \mu)Z_{0p} (+\mu_i) + X_{0p} (\pm \mu). \]  

(165)

These are analytical interpolation formulas for the source function for each layer, expressed in the solutions of the eigenvalue problem for the respective layer.

Using these interpolation formulas for the source functions gives interpolation formulas for the intensity as well, thus making it possible to calculate the intensity at any depth and at any angle. The interpolation formulas for the intensity then become

\[
\begin{align*}
I_p^+(\tau, \mu) &= I_L^+(\tau_L, \mu) e^{-\frac{(\tau L - \tau)}{\mu}} + \\
&\quad + \sum_{n=p}^{L} \sum_{j=1}^{N} C_{j-n} \tilde{g}_{j-n} (\pm \mu) \left( e^{\frac{k_{n+1} \tau L - \tau}{\mu}} - e^{\frac{k_{n+1} \tau L - \tau}{\mu}} \right)
\end{align*}
\]

(166)

\[
\begin{align*}
I_p^-(\tau, \mu) &= I_0^-(\tau_0, \mu) e^{-\frac{(\tau_0 - \tau)}{\mu}} + \\
&\quad + \sum_{n=p}^{N} \sum_{j=1}^{N} C_{j-n} \tilde{g}_{j-n} (\pm \mu) \left( e^{\frac{k_{n+1} \tau_0 - \tau}{\mu}} - e^{\frac{k_{n+1} \tau_0 - \tau}{\mu}} \right)
\end{align*}
\]

where

\[ C_{0p} \tilde{g}_{0p} (\pm \mu) = \tilde{Z}_{0p} (\pm \mu), \]  

(167)

\[ k_{0p} \equiv \frac{1}{\mu_0} \quad \text{and (of course) } k_{-jp} = -k_{jp}. \]  

(168)

As above, \( \tau_{n-1} \) should be replaced by \( \tau \) for \( n = p \) in the first equation and \( \tau_n \) should be replaced by \( \tau \) for \( n = p \) in the second equation.

In the interpolation formulas everything is known except \( I_0^-(\tau_0, \mu) \) and \( I_L^+(\tau_L, \mu) \). \( I_0^-(\tau_0, \mu) \) can be determined from the incident radiation at the top boundary. Then \( I_L^+(\tau_L, \mu) \) is calculated from the interpolation formula, and using the boundary conditions \( I_L^-(\tau_L, \mu) \) can be found.

There is a risk for overflow in the interpolation formulas for the intensity, but by using the same scaling as above, the interpolation formulas for the intensity become as follows.
\[ I_p'(\tau, \mu) = I_p'(\tau_0, \mu) e^{-\tau_0/\mu} + \sum_{n=p} \left\{ \bar{Z}_{0n}(+\mu) \left( e^{-(\mu \tau_{n-1} + \tau_{n-1})/\mu} - e^{-(\mu \tau_n + \tau_n)/\mu} \right) + \sum_{j=1}^N C^p_j e^{-(\mu \tau_{n-1} + \tau_{n-1})/\mu} - e^{-(\mu \tau_n + \tau_n)/\mu} \right\} \]

with \( \tau_{n-1} \) replaced by \( \tau \) and the exponentials in the second sum replaced by

\[ e^{-k_p(\tau - \tau_{n-1})} - e^{-k_p(\tau - \tau_{n-1})/\mu} \]

for \( n = p \).

\[ I_p'(\tau, \mu) = I_p'(\tau_0, \mu) e^{-\tau_0/\mu} + \sum_{n=0}^p \left\{ \bar{Z}_{0n}(-\mu) \left( e^{-(\mu \tau_{n-1} + \tau_{n-1})/\mu} - e^{-(\mu \tau_n + \tau_n)/\mu} \right) + \sum_{j=1}^N C^p_j e^{-(\mu \tau_{n-1} + \tau_{n-1})/\mu} - e^{-(\mu \tau_n + \tau_n)/\mu} \right\} \]

with \( \tau_{n} \) replaced by \( \tau \) and the exponentials in the third sum replaced by

\[ e^{-k_p(\tau - \tau_n)} - e^{-k_p(\tau - \tau_n)/\mu} \]

for \( n = p \). Since all \( k_{jn} > 0 \) (and especially \( k_{jp} > 0 \) ) and \( \tau_{p-1} < \tau < \tau_p \), all exponentials have negative arguments, and the risk of overflow is avoided.

As can be seen, there is also a risk that the denominators \( 1 - \mu/\mu_0 \) and \( 1 - k_{jn} \mu \), can be close to zero. This risk can be entirely eliminated, as in the single layer case, by noting that when they are close to zero, there is in fact an exponential with argument close to zero in the integral in the step before. An exponential with zero argument is a constant, and the corresponding anti-derivative does not have this denominator at all. If a denominator is close to zero, the corresponding term in the interpolation formula is simply substituted as described below.
In $I_p^+(\tau, \mu)$,

$$\frac{1}{\mu} C'_{-j_n} \tilde{g}_{-j_n}(+\mu) e^{-k_{j_n}(r_n-r_{n+1})(\tau-r)/\mu} - e^{-(r_n-r)/\mu}$$

for

$$\frac{1}{\mu} C'_{-j_n} \tilde{g}_{-j_n}(+\mu) e^{-k_{j_n}(r_n-r_{n+1})}(\tau_n - r_{n+1})$$

if $1 - k_{j_n} \mu$ is close to zero

(173)

(use $\frac{1}{\mu} C'_{-j_p} \tilde{g}_{-j_p}(+\mu) e^{-k_{j_p}(r_p-r)}(\tau_p - \tau)$ when $n = p$).

In $I_p^-(\tau, \mu)$,

$$\frac{1}{\mu} C'_{j_n} \tilde{g}_{j_n}(-\mu) e^{-k_{j_n}(r_n-r_{n+1})}(\tau_n - r_{n+1})$$

if $1 - k_{j_n} \mu$ is close to zero

(174)

(use $\frac{1}{\mu} C'_{j_p} \tilde{g}_{j_p}(-\mu) e^{-k_{j_p}(r_p-r_{p+1})}(\tau - \tau_{p+1})$ when $n = p$),

$$\frac{1}{\mu_0} \tilde{Z}_{0_p}(-\mu) e^{-k_{0_p}(r_{p+1})}(\tau_n - r_{n+1})$$

if $1 - \mu / \mu_0$ is close to zero

(175)

(use $\frac{1}{\mu_0} \tilde{Z}_{0_p}(-\mu) e^{-\tau_{p+1}/\mu_0}(\tau - \tau_{p+1})$ when $n = p$).

These terms are found by setting the corresponding exponential argument to zero and integrating as in the original interpolation formula. This can in fact be seen as an application of l’Hospita1’s rules.

It is worth stating once again that this is the solution for one Fourier component of the diffuse intensity. The complete azimuthal dependence can be assembled through the Fourier cosine series expansion for the diffuse intensity (151), as stated earlier. The total intensity is the sum of the diffuse and beam components, where the diffuse component has just been calculated, and the beam component is given by (152). The diffuse component includes reflection from the bottom surface. The beam component is therefore only present in downward directions, and the final expression for the total intensity is given by (153).
3.3. The $\delta$-N Method

If the scattering is strongly forward-peaked, an accurate expansion of the phase function needs a large number, up to several hundreds or thousands, of terms. To maintain the accuracy throughout the solution, a comparable number of terms are needed in the numerical quadrature used to approximate the integrals. This quickly gives very large eigenvalue problems and systems of equations, and since the computation time for these grows roughly as the third power of the size, the problem soon becomes intractable. To avoid this, a transformation proposed by Wiscombe [22] can be applied to get a problem with a less peaked phase function.

The idea is to consider the beams scattered through the small angles within the sharp forward peak as unscattered, and truncate this peak from the phase function. The phase function is separated into the sum of a Dirac delta function in the forward direction and a truncated phase function, which is expanded in a series of Legendre polynomials with a much smaller number of terms, preferably equal to the number of quadrature points, i.e. $2N$.

On one hand, the phase function is directly expanded in Legendre polynomials as

$$p(\cos \Theta) = \sum_{l=0}^{N_{\text{max}}} (2l+1) \chi_l P_l(\cos \Theta).$$

(176)

On the other hand the delta peak is first removed and then the remainder is expanded as

$$p(\cos \Theta) = fp\delta(\cos \Theta) + (1-f)p'(\cos \Theta) \approx$$

$$\approx f\delta(1-\cos \Theta) + (1-f)\sum_{l=0}^{2N-1} (2l+1) \tilde{\chi}_l P_l(\cos \Theta) = \tilde{p}_{\delta-N}(\cos \Theta),$$

(177)

where $f$ is a dimensionless parameter between 0 and 1 ($f$ thus denotes the fraction of the phase function that is contained in the separated delta peak). Demanding that the coefficients for Legendre polynomial expansion be the same for $p$ and $\tilde{p}_{\delta-N}$, as long as they have common terms, yields

$$\chi_l = f + (1-f)\tilde{\chi}_l, \quad \text{or} \quad \tilde{\chi}_l = \frac{\chi_l - f}{1-f}, \quad l = 0,1,...,2N-1.$$

(178)

The expansion for $\tilde{p}_{\delta-N}$ is truncated by demanding

$$\tilde{\chi}_{2N} = 0,$$

(179)

which gives

$$f = \chi_{2N}.$$

(180)

Starting with the radiative transfer equation

$$\frac{dI(\mu,\varphi)}{d\tau} = I(\mu,\varphi) - \frac{a}{4\pi} \int p(\mu',\varphi';\mu,\varphi)I(\mu',\varphi')d\omega',$$

(181)
and replacing $p$ with $\hat{p}_{\delta-N}$, gives

$$\mu \frac{dI(\mu, \varphi)}{d\tau} = I(\mu, \varphi) - \frac{a}{4\pi} \int_0^{4\pi} \hat{p}_{\delta-N}(\mu', \varphi'; \mu, \varphi) I(\mu', \varphi') d\omega' =$$

$$= I(\mu, \varphi) - \frac{a}{4\pi} \int_0^{4\pi} \left( f\delta(1 - \cos \Theta) + (1 - f)p'(\mu', \varphi'; \mu, \varphi) \right) I(\mu', \varphi') d\omega' = . \quad (182)$$

$$= I(\mu, \varphi) - afI(\mu, \varphi) - \frac{a}{4\pi} \int_0^{4\pi} (1 - f)p'(\mu', \varphi'; \mu, \varphi) I(\mu', \varphi') d\omega'$$

Some restructuring yields

$$\mu \frac{dI(\mu, \varphi)}{(1 - af)d\tau} = I(\mu, \varphi) - \frac{a}{4\pi} \frac{1 - f}{1 - af} \int_0^{4\pi} p'(\mu', \varphi'; \mu, \varphi) I(\mu', \varphi') d\omega' , \quad (183)$$

which, introducing

$$\tau' = (1 - af)\tau \quad (184)$$

and

$$a' = \frac{1 - f}{1 - af} \quad (185)$$

becomes

$$\mu \frac{dI(\mu, \varphi)}{d\tau'} = I(\mu, \varphi) - \frac{a'}{4\pi} \int_0^{4\pi} p'(\mu', \varphi'; \mu, \varphi) I(\mu', \varphi') d\omega' . \quad (186)$$

Hence, the $\delta-N$ method does not change the mathematical form of the radiative transfer equation. It only changes the optical properties of the medium to make it appear less anisotropic.

### 3.4. Intensity Correction Procedures

The accuracy of the intensity computation is generally improved by the use of the $\delta-N$ method except in the direction of the forward peak, but the $\delta-N$ method also introduces minor errors in other directions. However, combining the $\delta-N$ method with exact computation of low orders of scattering can reduce the error considerably. The purpose of this is to achieve high accuracy with small $N$, to speed up calculations. The TMS and IMS methods of Nakajima and Tanaka [23] serve to correct for single scattering and secondary and higher orders of scattering respectively. These methods are outlined below.

The TMS and IMS methods have no effect if there is no beam source, if the scattering coefficient is zero throughout the medium, or if the separated fraction $f'$ is very small, and the methods are therefore automatically shut off in these cases to save computation time.
3.4.1. The TMS Method

The phase function resulting from the δ-N method oscillates around the original phase function with a magnitude depending on the parameter \( f \). This gives the computed intensities an oscillating behavior, which becomes more apparent the more peaked the phase function is. Since single scattering resembles the phase function, it should be a good idea to compute the single scattering exactly to account for errors due to the δ-N method.

Exact solutions for the single-scattered intensity are easy to derive. Using equations (36) without the multiple scattering terms, and allowing for the optical properties to vary between layers, gives

\[
\begin{align*}
\frac{dI(\tau, +\mu, \varphi)}{d\tau} &= I(\tau, +\mu, \varphi) - \frac{a(\tau)}{4\pi} p(\tau, -\mu_0, \varphi_0; +\mu, \varphi) I_{0b} e^{-\tau/\mu_0}, \\
-\frac{dI(\tau, -\mu, \varphi)}{d\tau} &= I(\tau, -\mu, \varphi) - \frac{a(\tau)}{4\pi} p(\tau, -\mu_0, \varphi_0; -\mu, \varphi) I_{0b} e^{-\tau/\mu_0}.
\end{align*}
\]

These are elementary first order differential equations, which, using integrating factors, can be rewritten as

\[
\begin{align*}
\frac{d}{d\tau} \left( I(\tau, +\mu, \varphi) e^{-\tau/\mu} \right) &= -\frac{I_{0b}}{4\pi\mu} a(\tau) p(\tau, -\mu_0, \varphi_0; +\mu, \varphi) e^{\tau(-1/\mu-1/\mu_0)} \\
\frac{d}{d\tau} \left( I(\tau, -\mu, \varphi) e^{\tau/\mu} \right) &= \frac{I_{0b}}{4\pi\mu} a(\tau) p(\tau, -\mu_0, \varphi_0; -\mu, \varphi) e^{\tau(1/\mu-1/\mu_0)}.
\end{align*}
\]

Integrating the first equation from \( \tau_L \) to \( \tau \), assuming \( I(\tau_L, +\mu, \varphi) = 0 \), gives

\[
I(\tau, +\mu, \varphi) e^{-\tau/\mu} = -\frac{I_{0b}}{4\pi\mu} \int_{\tau_L}^{\tau} a(t) p(t, -\mu_0, \varphi_0; +\mu, \varphi) e^{t(-1/\mu-1/\mu_0)} dt = \frac{I_{0b}}{4\pi(1 + \mu / \mu_0)} \sum_{n=\rho}^{L} a_n p_n (-\mu_0, \varphi_0; +\mu, \varphi) \left( e^{s_n(-1/\mu-1/\mu_0)} - e^{s_n(-1/\mu-1/\mu_0)} \right),
\]

and integrating the second equation from \( \tau_0 \) to \( \tau \), assuming \( I(\tau_0, -\mu, \varphi) = 0 \), gives

\[
I(\tau, -\mu, \varphi) e^{\tau/\mu} = \frac{I_{0b}}{4\pi\mu} \int_{\tau_0}^{\tau} a(t) p(t, -\mu_0, \varphi_0; -\mu, \varphi) e^{t(1/\mu-1/\mu_0)} dt = \frac{I_{0b}}{4\pi(1 - \mu / \mu_0)} \sum_{n=1}^{\rho} a_n p_n (-\mu_0, \varphi_0; -\mu, \varphi) \left( e^{s_n(1/\mu-1/\mu_0)} - e^{s_n(1/\mu-1/\mu_0)} \right).
\]

The assumption of zero reflected intensity at the lower boundary is not a problem, since this is only the calculation of the single-scattered intensity, and the diffusely scattered intensity at the lower surface can be considered as multiply scattered, and is handled by the boundary conditions in the overall problem. It could be a problem, however, if the lower surface were specularly reflecting. The assumption of zero incident intensity, except for the beam, at the upper boundary is not a problem, since this is only the
calculation of the single-scattered intensity, and the diffuse incident intensity at the upper surface is handled by the boundary conditions in the overall problem.

After restructuring, the exact single scattering solution becomes

\[
I(\tau, \mu, \varphi) = \frac{I_{0b}}{4\pi(1 + \mu / \mu_0)} \sum_{n=0}^{l} a_n P_n \left( -\mu_0, \varphi_0; \pm \mu, \varphi \right) \left( e^{-\frac{\tau - \tau_n}{\mu} - \frac{\tau_n}{\mu_0}} - e^{-\frac{\tau - \tau_n}{\mu} - \frac{\tau_n}{\mu_0}} \right), \tag{191}
\]

with \( \tau_{n-1} \) replaced by \( \tau \) for \( n = p \), and

\[
I(\tau, -\mu, \varphi) = \frac{I_{0b}}{4\pi(1 - \mu / \mu_0)} \sum_{n=0}^{p} a_n P_n \left( -\mu_0, \varphi_0; -\mu, \varphi \right) \left( e^{-\frac{\tau - \tau_n}{\mu} - \frac{\tau_n}{\mu_0}} \right), \tag{192}
\]

with \( \tau_n \) replaced by \( \tau \) for \( n = p \). For the special case that \( \mu = \mu_0 \), the last solution changes, since the integral includes an exponential with zero argument:

\[
I(\tau, -\mu_0, \varphi) e^{\tau / \mu_0} = \frac{I_{0b}}{4\pi\mu_0} \int_{\tau}^{\infty} a(t) p(t, -\mu_0, \varphi_0; -\mu_0, \varphi) e^{\tau_0} dt = \frac{I_{0b}}{4\pi\mu_0} \sum_{n=1}^{p} a_n P_n \left( -\mu_0, \varphi_0; -\mu_0, \varphi \right) (\tau_n - \tau_{n-1}) \tag{193}
\]

The solution then becomes

\[
I(\tau, -\mu_0, \varphi) = \frac{I_{0b}}{4\pi\mu_0} e^{-\tau / \mu_0} \sum_{n=1}^{p} a_n P_n \left( -\mu_0, \varphi_0; -\mu_0, \varphi \right) (\tau_n - \tau_{n-1}), \tag{194}
\]

with \( \tau_n \) replaced by \( \tau \) for \( n = p \). As can be seen, all exponentials in the solution have negative arguments, so there is no risk of overflow.

The TMS method subtracts the erroneous single-scattered intensity from the \( \delta-N \) method, and adds back the exactly calculated single-scattered intensity. Specifically, the TMS method subtracts the single-scattered intensity obtained by using the scaled \( \tau' \), the scaled \( a' \) and the phase function \( p'(\cos \Theta) = \sum_{l=0}^{2N-1} (2l+1) \chi_l P_l(\cos \Theta) \) from the \( \delta-N \) method, and adds back the single-scattered intensity obtained by using the scaled \( \tau' \), \( a' \) (where the denominator is a consequence of the scaled \( \tau' \)) and the exact phase function \( p(\cos \Theta) = \sum_{l=0}^{N_{max}} (2l+1) \chi_l P_l(\cos \Theta) \) with all available terms. This can be denoted

\[
I_{TMS} = I' + \Delta I_{TMS} = I' - I_{ss} + I_{ss}^{corr}, \tag{195}
\]

where \( I' \) is the intensity computed by using the \( \delta-N \) method, and \( I_{ss} \) and \( I_{ss}^{corr} \) are the single-scattered intensities described above.

The TMS method gives a substantial improvement for the computed intensity, and the oscillations are suppressed. Only in the direction of the forward peak an error remains.
This can be corrected by accounting for secondary and higher orders of scattering, which is done in the IMS method.

### 3.4.2. The IMS Method

The purpose of the IMS method is the same as of the TMS method, except that it corrects for secondary and higher orders of scattering. Of course, exact solutions cannot be found for these corrections, since that would be to actually solve the overall problem. Instead, an exact solution can be derived symbolically, and then clever approximations need to be made to make it possible to use it in the IMS method. Reaching the final expression for the IMS method requires a substantial amount of algebra, part of which is done in appendix.

The IMS method only corrects the intensity inside a cone centered on the forward peak direction, and thus only affects the downward intensity. Therefore, in this section, some simplifying notation can be used. All intensity variables \( I \) implicitly mean \( I(\tau, -\mu, \phi) \) and all angular integrals \( \frac{1}{4\pi} \int d\omega \) implicitly mean
\[
\frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi p(\tau, \mu', \phi'; -\mu, \phi) I(\tau, \mu', \phi') d\mu' d\phi'.
\]

The optical properties, \( a, f \) and \( p \cdot I_{0b} \) implicitly mean \( a(\tau), f(\tau) \) and \( p(\tau, -\mu_0, \phi_0; -\mu, \phi) \cdot I_{0b} \), respectively.

Using the notation
\[
I_{\text{true}} = I_{\text{TMS}} - \Delta I_{\text{IMS}},
\]
where \( I_{\text{true}} \) is the solution to the exact radiative transfer equation, what is left to be found is an expression for the IMS correction term
\[
\Delta I_{\text{IMS}} = I_{\text{TMS}} - I_{\text{true}}.
\]

Differentiating this and using the definitions of \( \tau' \) and \( I_{\text{TMS}} \) gives
\[
-\mu \frac{d}{d\tau} (\Delta I_{\text{IMS}}) = \mu \frac{d}{d\tau} (I_{\text{true}}) - \mu (1 - af) \frac{d}{d\tau'} (I' + I_{\text{corr}}^\text{ss} - I_{\text{ss}}').
\]

Using the definitions of \( I_{\text{true}} \) and \( I_{\text{TMS}} \), \( p'' = \frac{1}{f} (p - (1 - f)p') \) from the \( \delta\)-N method, defining the \( \delta\)-N multiple-scattered intensity as \( I_{\text{mult}}' = I' - I_{\text{ss}}' \) and algebraically rearranging, gives (see appendix 1 for details)
\[- \mu \frac{d}{d\tau} (\Delta I_{\text{IMS}}) = \Delta I_{\text{IMS}} - \frac{a}{4\pi} \int p \cdot \Delta I_{\text{IMS}} d\omega' + \]

\[-a I_{\text{mult}}' + \frac{a}{4\pi} \int p'' \cdot I_{\text{mult}} d\omega' + \]

\[-a I_{\text{ss}}' + \frac{a}{4\pi} \int p'' \cdot I_{\text{ss}}^\text{corr} d\omega' + \]

\[+ \frac{a}{4\pi} e^{-\tau/\mu_0} p \cdot I_{0b} - \frac{a}{4\pi} e^{-\tau/\mu_0} p \cdot I_{0b} + \frac{a}{4\pi} (1 - f) \int p' \cdot (I_{\text{ss}}^\text{corr} - I_{\text{ss}}') d\omega' \]

Introducing

\[Q_1 = af \left( I_{\text{mult}}' - \frac{1}{4\pi} \int p'' \cdot I_{\text{mult}} d\omega' \right), \quad (200)\]

\[Q_2 = af \left( I_{\text{ss}}' - \frac{1}{4\pi} \int p'' \cdot I_{\text{ss}}^\text{corr} d\omega' \right) \quad (201)\]

and

\[Q_3 = \frac{a}{4\pi} p \cdot I_{0b} \left( e^{-\tau/\mu_0} - e^{-\tau/\mu_0} \right) - \frac{a}{4\pi} (1 - f) \int p' \cdot (I_{\text{ss}}^\text{corr} - I_{\text{ss}}') d\omega' , \quad (202)\]

this can finally be written as

\[- \mu \frac{d}{d\tau} (\Delta I_{\text{IMS}}) = \Delta I_{\text{IMS}} - \frac{a}{4\pi} \int p \cdot \Delta I_{\text{IMS}} d\omega' - (Q_1 + Q_2 + Q_3) . \quad (203)\]

This exact equation for the IMS correction term $\Delta I_{\text{IMS}}$ is more complicated than the original radiative transfer equation, so several approximations need to be made in order to make the IMS method practically useful. The approximations are as follows.

- \[Q_1 \approx 0, \quad (204)\]

in the equation for $\Delta I_{\text{IMS}}$, since its contribution to the narrow forward peak, where the IMS correction method is used, is negligible.

- \[Q_1 \approx 0, \quad (205)\]

since its contribution to the narrow forward peak, where the IMS correction method is used, is negligible.

- \[Q_3 \approx \frac{I_{0b}}{4\pi} \frac{(af)^2}{1 - af} e^{-\tau/\mu_0} \left( p'' - \frac{1}{4\pi} \int p'' \cdot p' d\omega' \right) \quad (206)\]

see appendix 2 for details.
• 

\[ Q_3 \approx \frac{I_{ob}}{4\pi} \left( af \right)^2 \frac{e^{-\tau'/\mu_0}}{1 - af \tau' \cdot p^n}, \]  

(207)

see appendix 3 for details.

• The IMS method uses vertically averaged optical properties:

\[ \bar{a} = \frac{\sum_{n=1}^{p} a_n \tau_n}{\sum_{n=1}^{p} \tau_n}, \]  

(208)

\[ \bar{f} = \frac{\sum_{n=1}^{p} f_n a_n \tau_n}{\sum_{n=1}^{p} a_n \tau_n}, \]  

(209)

and

\[ p^n(\cos \Theta) = \sum_{l=0}^{N_{max}} \left( 2l + 1 \right) \bar{x}_l \bar{P}_l(\cos \Theta), \]  

(210)

where

\[ \bar{x}_l = \frac{\sum_{n=1}^{p} \chi'_{l,n} a_n \tau_n}{\sum_{n=1}^{p} f_n a_n \tau_n} \quad \text{and} \quad \chi'_{l,n} = \begin{cases} f_n, & l \leq 2N - 1 \\ \chi_{l,n}, & l > 2N - 1. \end{cases} \]  

(211)

In order to express \( Q_2 \) and \( Q_3 \) in \( \tau \) instead of \( \tau' \),

\[ \tau = \frac{1}{1 - \bar{a} \bar{f}} \tau' \]  

(212)

and

\[ \mu'_0 \equiv \frac{1}{1 - \bar{a} \bar{f}} \mu_0 \]  

(213)

can be used. This gives

\[ Q_2 + Q_3 \approx \frac{I_{ob}}{4\pi} \left( \frac{a f'}{\bar{a} \bar{f}} \right)^2 e^{-\tau'/\mu'_0} \frac{\tau(2p^n - p^{*2})}{\mu_0' \mu_0}, \]  

(214)

where

\[ p^{*2} = \frac{1}{4\pi} \int p^n \cdot p^n d\omega'. \]  

(215)
The equation for the IMS correction term $\Delta I_{\text{IMS}}$ then becomes

$$-\mu \frac{d}{d\tau} (\Delta I_{\text{IMS}}) = \Delta I_{\text{IMS}} - (Q_2 + Q_3), \quad (216)$$

a first order differential equation that can be solved by integrating from 0 to $\tau$, using $e^{\tau/\mu}$ as integrating factor. This yields

$$\Delta I_{\text{IMS}} = e^{-\tau/\mu} \int_0^\tau \frac{e^{\tau'/\mu}}{\mu} (Q_2 + Q_3) dt =$$

$$= e^{-\tau/\mu} \frac{I_{\text{ob}}}{4\pi} \left( \frac{\bar{a}_f}{1 - \bar{a}_f} \right)^2 (2p^* - p^{*2}) \int_0^\mu \frac{e^{\tau'/\mu}}{\mu} \frac{e^{-\tau'/\mu}}{\mu_0} d\mu d\mu' = . \quad (217)$$

$$= \frac{I_{\text{ob}}}{4\pi} \left( \frac{\bar{a}_f}{1 - \bar{a}_f} \right)^2 (2p^* - p^{*2}) \int_0^\mu \frac{e^{\tau'/\mu}}{\mu_0} \int_0^{(1/\mu - 1/\mu_0)} e^{\tau'/\mu} t dt d\mu' d\mu.$$ 

The evaluation of $p^{*2}$ gives, due to the orthogonality of the Legendre polynomials,

$$\frac{1}{4\pi} \int_0^\pi p^* \cdot p^* d\omega = \frac{1}{4\pi} \int_0^{2\pi} \int_0^1 \int_0^1 p^*(\tau, \mu', \varphi' ; - \mu, \varphi) p^*(\tau, \mu_0, \varphi_0, \mu', \varphi) d\mu' d\varphi' =$$

$$= \sum_{l=0}^{N_{\text{leg}}} (2l + 1) \bar{X}_l^2 P_l (\cos \Theta) \quad . \quad (218)$$

This gives an expression for the IMS correction term as

$$\Delta I_{\text{IMS}} = \frac{I_{\text{ob}}}{4\pi} \left( \frac{\bar{a}_f}{1 - \bar{a}_f} \right)^2 \sum_{l=0}^{N_{\text{leg}}} (2l + 1) \bar{X}_l^2 P_l (\cos \Theta) e^{-\tau/\mu} \int_0^\mu \frac{e^{\tau'/\mu}}{\mu_0} \int_0^{(1/\mu - 1/\mu_0)} e^{\tau'/\mu} t dt . \quad (219)$$

The evaluation of the integral $\frac{e^{-\tau/\mu}}{\mu_0} \int_0^\mu e^{(1/\mu - 1/\mu_0)} t dt$ must be separated into two cases.

The case $\mu \neq \mu_0'$ gives, using partial integration,

$$\frac{e^{-\tau/\mu}}{\mu_0} \int_0^\mu e^{(1/\mu - 1/\mu_0)} t dt =$$

$$= \frac{1}{\mu_0} \left( \frac{1}{1/\mu - 1/\mu_0} \right) \left( e^{-\tau/\mu_0} \left( \frac{1}{1/\mu - 1/\mu_0} \right) + (1/\mu - 1/\mu_0) \right) . \quad (220)$$

The case $\mu = \mu_0'$ gives

$$\frac{e^{-\tau/\mu}}{\mu_0} \int_0^\mu e^{(1/\mu - 1/\mu_0)} t dt = \frac{e^{-\tau/\mu}}{\mu_0} \int_0^\mu t dt = \frac{e^{-\tau/\mu}}{\mu_0^2} \frac{\tau^2}{2} . \quad (221)$$
A Fourier cosine series expansion gives the final expression to be used in the IMS method:

\[
\Delta I_{\text{IMS}} = \frac{I_{0b}}{4\pi} \left( \frac{\vec{a}}{1 - \vec{a}^2} \right)^2 \frac{e^{-\tau/\mu}}{\mu \mu_0'} \int_0^\tau \left( 2 - \delta_{0m} \right) \sum_{m=0}^{N_{\text{max}}} (2 - \delta_{0m}) p_{\text{IMS}}^m (-\mu_0', -\mu) \cos(m(\varphi_0 - \varphi)) ,
\]

where

\[
p_{\text{IMS}}^m (-\mu_0', -\mu) = \sum_{l=m}^{N_{\text{max}}} (21 + 1)(2\vec{x}_l - \vec{x}_l^2) \Lambda_l^m (-\mu_0') \Lambda_l^m (-\mu).
\]

Thus a single Fourier component becomes

\[
\Delta f_{\text{IMS}}^m = \frac{I_{0b}}{4\pi} \left( \frac{\vec{a}}{1 - \vec{a}^2} \right)^2 (2 - \delta_{0m}) \frac{p_{\text{IMS}}^m (-\mu_0', -\mu)}{\mu \mu_0'} \int_0^\tau \frac{e^{-\tau/\mu}}{\mu \mu_0'} \ t dt .
\]

### 3.5. Breaking the Azimuthal Loop

In many cases the intensity converges well before the azimuthal loop has ended. Since it is the outermost loop, much is gained if it can be terminated earlier. A method has been implemented that breaks the azimuthal loop when a convergence criterion has been met. This saves a large amount of computation time in the vast majority of cases.

The method is an engineering method based on an ad hoc assumption that the Fourier components of the intensity are approximately exponentially decreasing. Experimental numerical studies of the Fourier components indicate that this assumption is valid in most cases. The method makes DORT2002 far more efficient, but it does not stand on a solid scientific ground. The algorithm has been carefully tested under a wide range of conditions and has proven to be valid for all tested single layer cases, which is not a guarantee that it will work for all possible single layer cases. There is no way to assure that the assumption of exponentially decreasing terms is valid or even reasonable in the infinity of combinations of optical properties that are possible to create. Some multilayer cases give bad results. The user should be aware of this, and it is also possible to turn off this feature if there is any doubt that it yields bad results in a particular case.

The convergence criterion for breaking the azimuthal loop is based on the ad hoc assumption of an exponential behavior of the Fourier components of the intensity. An exception is made for the 0th component, since it shows a different behavior due to its direct coupling to any diffuse incident intensity. The assumption is

\[
I_m(\mu_i) \leq \begin{cases} I_0(\mu_i), & m = 0 \\ \alpha(\mu_i) e^{-m\beta(\mu_i)}, & m > 0 \end{cases},
\]

where \( m \) is the number of the Fourier component. As can be seen, the behavior is allowed to be different for each polar angle. The total intensity is then
\[ I(\mu_i) = I_0(\mu_i) + \sum_{m=1}^{\infty} \alpha(\mu_i) e^{-m\beta(\mu_i)} = I_0(\mu_i) + \frac{\alpha(\mu_i)e^{-\beta(\mu_i)}}{1-e^{-\beta(\mu_i)}}. \] (226)

The \( \alpha \) and \( \beta \) are unknown, and the \( I_m(\mu_i) \) are calculated by DORT2002 for \( m = 0,1,\ldots,2N-1 \).

The objective is to stop the calculations well before \( m = 2N-1 \), based on the assumed knowledge that the total contribution from the omitted terms is less than a user supplied limit.

The error when stopping after term \( k \), \( 0 < k \leq 2N-1 \), is thus
\[ E_k(\mu_i) = \sum_{m=k+1}^{\infty} \alpha(\mu_i) e^{-m\beta(\mu_i)} = \frac{\alpha(\mu_i)e^{-(k+1)\beta(\mu_i)}}{1-e^{-\beta(\mu_i)}}. \] (227)

which makes the relative error
\[ \varepsilon_k(\mu_i) = \frac{E_k(\mu_i)}{I(\mu_i)} = \frac{\alpha(\mu_i)e^{-(k+1)\beta(\mu_i)}}{I_0 + \frac{\alpha(\mu_i)e^{-\beta(\mu_i)}}{1-e^{-\beta(\mu_i)}}}. \] (228)

Since
\[ e^{-\beta(\mu_i)} = \frac{I_2(\mu_i)}{I_1(\mu_i)} \] (229)
and
\[ \alpha(\mu_i) = I_1(\mu_i)e^{\beta(\mu_i)}, \] (230)
this becomes
\[ \varepsilon_k(\mu_i) = \frac{I_1(\mu_i)(I_2(\mu_i)/I_1(\mu_i))^k}{I_0(\mu_i)(1-I_2(\mu_i)/I_1(\mu_i)) + I_1(\mu_i)}. \] (231)

With \( \varepsilon_{max} \) as an upper limit of the relative error, the needed number of terms is given by
\[ k = \max_{\mu_i} \frac{\ln \left( \frac{\varepsilon_{max}(I_0(\mu_i)(1-I_2(\mu_i)/I_1(\mu_i)) + I_1(\mu_i))}{I_1(\mu_i)} \right)}{\ln \left( \frac{I_2(\mu_i)/I_1(\mu_i)}{I_1(\mu_i)} \right)}. \] (232)

Already after computing \( I_0, I_1 \) and \( I_2 \) it is thus possible to determine how many terms in the Fourier expansion of the intensity that are needed. Of course, if several depths are calculated simultaneously, the maximum \( k \) over all depths must be used.

However, tests show that the Fourier components may be fluctuating, primarily when the asymmetry factor \( g \) is negative, and the odd terms may also have oscillating signs. It
appears in these cases that the assumption of exponentially decreasing Fourier components is valid for components with even and odd $m$ separately, and not for all components together. It tends to be either the even or the odd components that dominate, and hence determine what number of terms is needed. A modified approach is therefore needed, which interpolates $\tilde{T}_1(\mu_i)$ from $I_2(\mu_i)$ and $I_4(\mu_i)$ if the even terms dominate, and interpolates $\tilde{T}_2(\mu_i)$ from $I_1(\mu_i)$ and $I_3(\mu_i)$ if the odd terms dominate. The interpolation is done linearly instead of exponentially, to increase the safety margin. The interpolation formulas are

$$\begin{align*}
\tilde{T}_1(\mu_i) &= \begin{cases} 
I_2(\mu_i), & |I_1(\mu_i)| \leq I_2(\mu_i) \\
\frac{I_2(\mu_i) - I_4(\mu_i)}{2} + I_1(\mu_i), & |I_1(\mu_i)| > I_2(\mu_i)
\end{cases}, \\
\tilde{T}_2(\mu_i) &= \begin{cases} 
I_1(\mu_i), & |I_1(\mu_i)| \leq I_2(\mu_i) \\
\frac{|I_1(\mu_i)| - |I_3(\mu_i)|}{2} + I_1(\mu_i), & |I_1(\mu_i)| > I_2(\mu_i)
\end{cases}.
\end{align*}$$

This gives the needed number of terms as

$$\bar{k} = \max_{\mu_i} \ln \left( \frac{\varepsilon_{\text{max}}(I_0(\mu_i)(1 - \tilde{T}_3(\mu_i)/\tilde{T}_1(\mu_i)) + \tilde{T}_1(\mu_i))}{\ln(\tilde{T}_2(\mu_i)/\tilde{T}_1(\mu_i))} \right).$$

Further tests show that the error is larger than the user supplied upper limit in cases where the ratio $I_0/I_1$ is small. To overcome this problem, a scaling function is introduced to ensure that the upper limit for the error is not exceeded. The scaling function is a simple polynomial, and the coefficients are found by investigating for a number of cases with different $I_0/I_1$ ratio how many terms are really needed for the error to stay within allowed limits. The coefficients are chosen so that the error is below the upper limit for the investigated cases.

By defining

$$\eta(\mu_i) = \frac{I_0(\mu_i)}{I_1(\mu_i)},$$

and

$$\overline{\eta}(\mu_i) = \frac{I_0(\mu_i)}{I_1(\mu_i)},$$

the polynomial scaling function is
\[ \kappa(x) = \begin{cases} \alpha x^2 + b, & x \leq 4 \\ \kappa(4), & x > 4 \end{cases} \quad (238) \]

where
\[
\begin{align*}
\alpha &= 0.0551 \\
b &= 0.0162
\end{align*} \quad (239)
\]

and where \( x \) is replaced by \( \eta(\mu_i) \) or \( \bar{\eta}(\mu_i) \) as appropriate. The number of terms to be used then becomes

\[
\ln \left( \frac{\kappa(\eta(\mu_i)) \varepsilon_{\text{max}} \left( I_0(\mu_i) \left( 1 - I_2(\mu_i) / I_1(\mu_i) \right) + I_1(\mu_i) \right)}{I_1(\mu_i) / I_2(\mu_i)} \right) \quad (240)
\]

for non-fluctuating terms and

\[
\ln \left( \frac{\kappa(\bar{\eta}(\mu_i)) \varepsilon_{\text{max}} \left( \bar{I}_0(\mu_i) \left( 1 - \bar{I}_2(\mu_i) / \bar{I}_1(\mu_i) \right) + \bar{I}_1(\mu_i) \right)}{\bar{I}_1(\mu_i) / \bar{I}_2(\mu_i)} \right) \quad (241)
\]

for fluctuating terms.

Under the assumption of decreasing terms, the quotient \( I_2 / I_1 \) in the argument of the logarithm in the denominators above fulfills \( 0 < I_2 / I_1 < 1 \). For the expressions for \( k \) and \( \bar{k} \) to be valid, the argument of the logarithm in the numerator must also be between 0 and 1. Since \( 0 < (1 - I_2 / I_1) < 1 \), this condition becomes

\[
0 < \left( \kappa(\eta) \varepsilon_{\text{max}} \left( I_0 / I_1 + 1 \right) \right) < 1. \quad (242)
\]

Here everything is positive, and \( \eta = I_0 / I_1 \), so this gives a limit for \( \varepsilon_{\text{max}} \) as

\[
\varepsilon_{\text{max}} < \frac{1}{\max \left( \kappa(\eta)(\eta + 1) \right)}. \quad (243)
\]

From the above definition of the polynomial scaling function this is given by \( \eta = 4 \), so

\[
\varepsilon_{\text{max}} < 0.2 \quad (244)
\]

is needed. It is also highly unlikely that anyone would want a result with a relative error larger than 20%.
4. Suggestions for Future Work

4.1. Model Development

This section shortly describes planned and suggested developments of DORT2002.

- Work is currently in progress to present the DORT2002 tool to a broader audience, e.g. the paper and printing industry. This includes a manual to the DORT2002 Graphical User Interface, a tool developed to provide users with a fast and easy way of performing DORT2002 simulations. It works as a shell that encapsulates the parameters and the function calls, and offers powerful simulations through a mouse click.

- Another work is in progress to evaluate the performance and application of version 2.0 of DORT2002. The results will be reported, and will possibly generate further improvements of the model.

- A method to handle layers with different indices of refraction has been described [21], and can be included in the model if desired. This includes refraction and total reflection at the boundaries, described by Snell’s law and Fresnel’s formulas.

- The model is prepared to include fluorescence. All that is needed is to add a corresponding term in the integro-differential equation. This term is easy to fit into the discretization and the solution procedure. An outer loop over wavelengths will be needed to perform the coupling of the intensities of the different wavelengths associated with the fluorescence.

- An interesting development would be to investigate better surface modeling, including gloss and scattering from rough surfaces, which in turn can be divided into diffraction from small surface irregularities, and geometrical scattering from the distribution of surface facets. A possible approach would be to combine DORT2002 with a specialized surface model.

- The method for breaking the azimuthal loop is based on an ad hoc assumption. It has been worked out with engineering skill and some trial and error. It works well in most tested cases, and it has a large positive impact on the computation time, but a closer study of the terms in the Fourier expansion of the intensity could give an even more efficient algorithm, and knowledge of when it is guaranteed to work.

- One bottleneck that remains is the generation of the sparse matrix in the system of equations for the boundary and continuity conditions. Although the values and the indices of the nonzero elements are known, the assigning of these values to the sparse matrix is unsatisfactorily time consuming in MATLAB, to the extent that this purely administrative part of the code consumes a significant part of the execution time. Since all computational parts of the code already are so optimized, this item is the first candidate to improve the speed of the code. This problem remains in spite of several different implementation strategies and consultation of experts at MathWorks (MATLAB is their product), and any suggestions to a solution would be welcome.
4.2. Inverse Problem

The inverse problem for this model will be studied. This includes the study and development of fast and numerically stable algorithms for parameter estimation. The problem involves several numerical difficulties, so it will require some effort.

The parameter estimation will be done to fit model simulations to angle-resolved light scattering measurements or to desired angle-resolved light scattering patterns. Also, studies of error estimation and sensitivity of perturbations will be performed.

In addition, it would be interesting to see studies of inverse Monte-Carlo methods. They should be evaluated with respect to accuracy and speed, and compared with the inverse model outlined above. It should also be investigated whether an inverse DORT method and an inverse Monte-Carlo method could be used together, one doing a coarse fitting, and the other doing fine-tuning.

An inverse light scattering model opens for several research activities in the paper and printing industries. This topic is covered in section 5.2.
5. Discussion

5.1. Aspects of Scientific Computing

This report continues an extensive treatment of a radiative transfer solution method, including all necessary steps to get a fast and numerically stable solution procedure, using a discrete ordinate model geometry. The resulting method is referred to as DORT2002, now in version 2.0.

The main steps to get a numerically stable multilayer solution procedure include the preconditioning of the system of equations for the boundary and continuity conditions, and the avoidance of over- and underflow in the multilayer solution and interpolation formulas. Important are also the recognition of potential divide-by-zero situations, and reformulation of those.

Several measures are taken to make the code fast. This includes exploitation of the sparse structure of the system of equations for the boundary and continuity conditions. Several features allow high speed by maintaining accuracy at significantly lower number of channels than would otherwise be needed, or by automatically stopping calculations earlier when certain convergence criteria have been met.

The model has been implemented in MATLAB from scratch. Some effort has been made to use matrix formulations, not only for the parts naturally expressed as matrices. This has been done to make use of MATLAB’s excellent matrix handling. MATLAB 6.5 incorporates LAPACK, which is state-of-the-art today, and these routines have been used for standard procedures. It also contains a JIT accelerator, which, according to MathWorks, makes MATLAB as fast as compiled languages such as Fortran or C.

5.2 The Impact of DORT2002 on the Paper and Printing Industries

There are different areas within the paper industry where one could find use for light scattering models [24]. Among them are fine-tuning the papermaking process, troubleshooting in the papermaking process and designing new paper qualities. Today the Kubelka-Munk model (or extended models thereof) is most widely used to cover these applications.

From the view of the applied user, DORT2002 has several advantages compared to the Kubelka-Munk model. The angular distribution of reflection and transmission is modeled, as well as different scattering asymmetries of the bulk. DORT2002 allows the use of collimated light to analyze the optical response of a sample, and is not limited to diffuse light as the Kubelka-Munk model. Since DORT2002 handles any illumination and detection conditions, the interior of instruments otherwise closed for inspection can be simulated, and the influence of instrument geometry on measurements can be evaluated. This makes it possible to suggest measurement corrections for deviations due to instrument geometry, and to make calibration and measurements with different instrument geometries comparable. Furthermore, DORT2002 is consistent for translucent and highly absorbing media, and it is prepared to be combined with a surface model to handle gloss. It is also prepared for a future implementation of
fluorescence, which will allow the effect of OBA (optical brightening agents) in paper and print to be modeled. The whiteness and brightness of paper cannot be designed with the Kubelka-Munk model since fluorescence phenomena are not explicitly included. Other possibilities include building DORT2002 into applications such as automated process and quality control, and getting some insight in ink penetration and its influence on light scattering. With DORT2002 the speed and accuracy for color matching for multicolor prints would possibly increase.

These improvements are important for a number of reasons. Paper can be translucent, glossy and strongly absorbing, e.g. low opacity paper, calendared paper or heavily dyed paper. This is also the case for full tone print. The standardized measurement geometries (d/0 and 45/0) for brightness of paper can give different ranking, which cannot be explained with the Kubelka-Munk model, but is readily given by DORT2002. Moreover, there is experimental evidence that the reflection and transmission of paper and print deviate from the Kubelka-Munk model description [24], which can be interpreted more accurately with DORT2002.

DORT2002 is far more powerful than Kubelka-Munk, and still almost as fast. A Monte-Carlo model such as Grace [25] allows more detailed simulations, but can be very time consuming and includes a large number of parameters that must be determined. These arguments result in the following suggestions: replace Kubelka-Munk with DORT2002 for most applications, and use a combination of Grace and DORT2002 for accurate modeling of paper and print.

In version 2.0 of DORT2002 several enhancements have been made that will considerably extend the applicability of DORT2002 in the paper and printing industries. A printed paper is often built up of a base paper layer with fillers, coating layers and ink layers, and simulating this will benefit from the new multilayer capabilities in DORT2002. A fast and numerically stable optimized code gives good performance, and computation time is reduced for all problems. The introduction of a graphical user interface makes the model easy to use, and after a short introduction any desired simulation is just a button-click away.

While a direct model with given parameters can describe the light scattering in paper or print, an inverse model works the other way around. Given a desired light scattering pattern, an inverse model can calculate the parameters that give that pattern. It can also be used to determine parameters by fitting the model to angle-resolved light scattering measurements. An inverse model is a crucial link in the chain from customer needs and market demands to new printed paper products.

One opportunity would be the development of tailor made products, new paper grades or new properties in paper and print. This could be done faster, with less trial-and-error and at lower costs, since a fairly accurate estimate of the “recipe” could be retrieved from the inverse model, and the number of test runs needed in the development would decrease noticeably. It would also be possible to couple the optical result to chemical and mechanical processes in the production.

Another opportunity would be to use angle-resolved light scattering measurements and spectral reflectance factor measurements to investigate the parameter space for paper
applications. This includes finding parameter values such as scattering and absorption coefficients and asymmetry factor for different samples, and to relate those parameter values to the optical response of the samples. This could be used to identify the parts of the parameter space that are interesting and relevant for paper applications, and to investigate parameter dependencies.
6. Acknowledgements
This report was financially supported by T2F (“TryckTeknisk Forskning, a Swedish printing research program), which is gratefully acknowledged. Hjalmar Granberg is thanked for valuable discussions concerning applications and for commenting on the manuscript.
Appendix 1 – A Simplification

Using the fact that $I_{true}$ is the solution to the exact radiative transfer equation yields

$$-\mu \frac{d}{d\tau} \left( \Delta I_{\text{IMS}} \right) = -I_{true} + \frac{a}{4\pi} \int p \cdot I_{true} d\omega' + \frac{a}{4\pi} e^{-\tau'|/\mu_0} p \cdot I_{0b} +$$

$$-\mu(1-af) \frac{d}{d\tau'} \left( I' + I_{ss}^{corr} - I_{ss}' \right),$$

(245)

which, after eliminating $I_{true}$, becomes

$$-\mu \frac{d}{d\tau} \left( \Delta I_{\text{IMS}} \right) = -I_{TMS} + \Delta I_{\text{IMS}} + \frac{a}{4\pi} \int p \cdot I_{TMS} d\omega' - \frac{a}{4\pi} \int p \cdot \Delta I_{\text{IMS}} d\omega' +$$

$$+ \frac{a}{4\pi} e^{-\tau'/\mu_0} p \cdot I_{0b} - \mu(1-af) \frac{d}{d\tau'} \left( I' + I_{ss}^{corr} - I_{ss}' \right).$$

(246)

Again using the definition of $I_{TMS}$ turns this into

$$-\mu \frac{d}{d\tau} \left( \Delta I_{\text{IMS}} \right) = \Delta I_{\text{IMS}} - \frac{a}{4\pi} \int p \cdot \Delta I_{\text{IMS}} d\omega' +$$

$$- \left( I' + I_{ss}^{corr} - I_{ss}' \right) + \frac{a}{4\pi} \int p \cdot \left( I' + I_{ss}^{corr} - I_{ss}' \right) d\omega' +.$$

(247)

$$+ \frac{a}{4\pi} e^{-\tau'/\mu_0} p \cdot I_{0b} - \mu(1-af) \frac{d}{d\tau'} \left( I' + I_{ss}^{corr} - I_{ss}' \right)$$

Using

$$p'' = \frac{1}{f}(p - (1-f)p'),$$

(248)

from the $\delta$-$N$ method, and defining the $\delta$-$N$ multiple-scattered intensity as

$I'_{mult} = I' - I_{ss}'$,  

(249)

gives

$$-\mu \frac{d}{d\tau} \left( \Delta I_{\text{IMS}} \right) = \Delta I_{\text{IMS}} - \frac{a}{4\pi} \int p \cdot \Delta I_{\text{IMS}} d\omega' +$$

$$- I'_{mult} + I_{ss}^{corr} + \frac{a}{4\pi} \int \left( f p'' + (1-f)p' \right)(I' + I_{ss}^{corr} - I_{ss}') d\omega' +,$$

(250)

$$+ \frac{a}{4\pi} e^{-\tau'/\mu_0} p \cdot I_{0b} - \mu(1-af) \frac{d}{d\tau} \left( I'_{mult} + I_{ss}^{corr} \right)$$
or, rearranging,
The two last terms need to be examined separately. The first of the two terms becomes

\[-\mu(1-\alpha_f) \frac{d}{d\tau'} (I_{ss}^{corr}) = (1-\alpha_f)(-\mu) \frac{d}{d\tau'} (I_{ss}^{corr}) =\]

\[= (1-\alpha_f) \left( I_{ss}^{corr} \frac{a}{4\pi} 1 - \alpha_f e^{-\tau' / \mu_0} p \cdot I_{0b} \right) = \]

\[= (1-\alpha_f) I_{ss}^{corr} \frac{a}{4\pi} e^{-\tau' / \mu_0} p \cdot I_{0b} \]  \hspace{1cm} (252)

The second term is a little more elaborate:

\[-\mu(1-\alpha_f) \frac{d}{d\tau'} (I'_{mult}) = (1-\alpha_f)(-\mu) \frac{d}{d\tau'} (I'_{mult}) =\]

\[= (1-\alpha_f) \left( I' - I_{ss} \frac{a'}{4\pi} \int p' \cdot I'd\omega' - \frac{a'}{4\pi} e^{-\tau' / \mu_0} p' \cdot I_{0b} + \frac{a'}{4\pi} e^{-\tau' / \mu_0} p' \cdot I_{0b} \right) =\]

\[= (1-\alpha_f) \left( I' - I_{ss} \frac{1-f}{1-\alpha_f} \frac{a}{4\pi} \int p' \cdot I'd\omega' \right) =\]

\[= (1-\alpha_f) \left( I' - I_{ss} \right) \frac{a}{4\pi} \int p' \cdot I'd\omega' =\]

\[= (1-\alpha_f) \left( I_{mult}' + I_{ss}' - I_{ss}' \right) \frac{a}{4\pi} \int p' \cdot I'd\omega' =\]

\[= (1-\alpha_f) I_{mult}' \frac{a}{4\pi} (1-f) \int p' \cdot I'd\omega'\]

Inserting this gives
\[-\mu \frac{d}{d\tau} (\Delta I_{\text{IMS}}) = \Delta I_{\text{IMS}} + \frac{a}{4\pi} \int p \cdot \Delta I_{\text{IMS}} d\omega' + \]

\[-I_{\text{mult}}' - I_{ss}^{\text{corr}} + \frac{a}{4\pi} f \int p'' \cdot I_{\text{mult}}' d\omega' + \frac{a}{4\pi} f \int p'' \cdot I_{ss}^{\text{corr}} d\omega' + \]

\[+ \frac{a}{4\pi} (1 - f) \int p' \cdot I' d\omega' + \frac{a}{4\pi} (1 - f) \int p' \cdot (I_{ss}^{\text{corr}} - I_{ss}') d\omega' \]

(254)

which after rearranging becomes

\[-\mu \frac{d}{d\tau} (\Delta I_{\text{IMS}}) = \Delta I_{\text{IMS}} + \frac{a}{4\pi} \int p \cdot \Delta I_{\text{IMS}} d\omega' + \]

\[-I_{\text{mult}}' + (1 - af) I_{ss}^{\text{corr}} + \frac{a}{4\pi} f \int p'' \cdot I_{\text{mult}}' d\omega' + \]

\[-I_{ss}^{\text{corr}} + (1 - af) I_{ss}^{\text{corr}} + \frac{a}{4\pi} f \int p'' \cdot I_{ss}^{\text{corr}} d\omega' \]

(255)

\[+ \frac{a}{4\pi} e^{-r'/\mu_0} p \cdot I_{0h} - \frac{a}{4\pi} e^{-r'/\mu_0} p \cdot I_{0b} + \frac{a}{4\pi} (1 - f) \int p' \cdot (I_{ss}^{\text{corr}} - I_{ss}') d\omega' \]
Appendix 2 – A Simplifying Approximation

Using the single-scattered intensity

\[ I_{ss}^{corr} = \frac{I_{0b}}{4\pi} \frac{a e^{-t'/\mu}}{1 - af} \int_0^\infty e^{t/(1 - \mu - \mu_0)} dt \]  

(256)

from previous calculations, \( Q_2 \) can be rewritten as

\[ Q_2 = \left( 1 - af \right) \frac{I_{0b}}{4\pi} \frac{a e^{-t'/\mu}}{1 - af} \int_0^\infty e^{t/(1 - \mu - \mu_0)} dt \]

\[ - \left( 1 - af \right) \frac{I_{0b}}{4\pi} \frac{a e^{-t'/\mu}}{1 - af} \int_0^\infty \int_0^{\pi/2} e^{t/(1 - \mu - \mu_0)} d\omega' \]

\[ \frac{e^{-t'/\mu}}{\mu} \int_0^\infty e^{t/(1 - \mu - \mu_0)} dt \] is approximately independent of \( \mu \) for \( \mu \) close to \( \mu_0 \). It can therefore be moved outside of the angular integral by setting \( \mu = \mu_0 \). This gives

\[ Q_2 \approx \left( 1 - af \right) \frac{I_{0b}}{4\pi} \frac{a e^{-t'/\mu_0}}{1 - af} \int_0^\infty \int_0^{\pi/2} p'' \cdot p d\omega' \]

\[ \approx \left( 1 - af \right) \frac{I_{0b}}{4\pi} \frac{a e^{-t'/\mu_0}}{1 - af} \int_0^\infty \int_0^{\pi/2} \left( p'' + (1 - f) p' \right) d\omega' \]

\[ = \left( 1 - af \right) \frac{I_{0b}}{4\pi} \frac{a e^{-t'/\mu_0}}{1 - af} \int_0^\infty \int_0^{\pi/2} \left( p'' + (1 - f) p' \right) d\omega' \]

\[ = \left( 1 - af \right) \frac{I_{0b}}{4\pi} \frac{a e^{-t'/\mu_0}}{1 - af} \int_0^\infty \int_0^{\pi/2} \left( p'' - \frac{1}{4\pi} \int_0^{\pi/2} p'' d\omega' \right) \]

(257)

where the approximation

\[ \frac{1}{4\pi} \int_0^{2\pi} p''(\tau, \mu', \varphi', -\mu, \varphi) p'(\tau, -\mu_0, \varphi_0, \mu', \varphi') d\mu' d\varphi' \approx \]

\[ \approx p'(\tau, -\mu_0, \varphi_0, -\mu, \varphi) \]

(258)

has been used, since \( p'' \) is an approximation of a Dirac delta-function.
Appendix 3 – A Simplifying Approximation

$e^{-\tau' / \mu_0} - e^{-\tau / \mu_0}$ can be rewritten as a product, and the second factor can then be expanded in a Taylor series. If, as an approximation, only the first term in the expansion is used, this yields

$$e^{-\tau' / \mu_0} - e^{-\tau / \mu_0} = e^{-\tau' / \mu_0} (1 - e^{(\tau' - \tau) / \mu_0}) = e^{-\tau' / \mu_0} (1 - e^{((1 - af) \tau - \tau) / \mu_0}) =$$

$$= e^{-\tau' / \mu_0} (1 - e^{-af \tau / \mu_0}) = -e^{-\tau' / \mu_0} \sum_{n=1}^{\infty} \frac{1}{n!} \left( \frac{-af \tau}{\mu_0} \right)^n \approx -e^{-\tau' / \mu_0} \frac{-af \tau}{\mu_0} = .$$

(260)

Using this approximation, and also using the single-scattered intensities

$$I_{ss}^{corr} = I_{ss} \frac{a}{4\pi} \frac{e^{-\tau' / \mu_0}}{1 - af} \int_0^{\pi} e^{(1 - \mu / \mu_0)} dt$$

(261)

and

$$I_{ss}' = I_{ss} \frac{1 - f}{4\pi} \frac{af}{1 - af} \int_0^{\pi} e^{(1 - \mu / \mu_0)} dt$$

(262)

from previous calculations, $Q_3$ can be rewritten as

$$Q_3 = I_{ss} \frac{a}{4\pi} \frac{af}{1 - af} \int_0^{\pi} e^{(1 - \mu / \mu_0)} dt$$

$$- \frac{a}{4\pi} (1 - f) \int_0^{\pi} e^{(1 - \mu / \mu_0)} dt$$

(263)

Again, $e^{-\tau' / \mu_0} \int_0^{\pi} e^{(1 - \mu / \mu_0)} dt$ is approximately independent of $\mu$ for $\mu$ close to $\mu_0$. It can therefore be moved outside of the angular integrals by setting $\mu = \mu_0$. This gives
\[ Q_3 \approx \]
\[ = \frac{I_{0b}}{4\pi} \frac{a}{1 - af} \frac{e^{-\tau'/\mu_0}}{\mu_0} \tau' \left( afp - \frac{a}{4\pi} (1 - f) \int \frac{p'}{4\pi} \left( fp^\prime + (1 - f)p' - (1 - f)p'\right) d\omega' \right) = \]
\[ = \frac{I_{0b}}{4\pi} \frac{a}{1 - af} \frac{e^{-\tau'/\mu_0}}{\mu_0} \tau' \left( afp - \frac{af}{1 - f} \frac{1}{4\pi} \int p' \cdot p'' d\omega' \right) \approx \]
\[ \approx I_{0b} \frac{a^2 f}{4\pi} \frac{\tau}{\mu_0} \frac{e^{-\tau'/\mu_0}}{4\pi} \tau' \left( fp^\prime + (1 - f)p' - (1 - f)p' \right) = \]
\[ = \frac{I_{0b}}{4\pi} \frac{a^2 f^2}{1 - af} \frac{e^{-\tau'/\mu_0}}{\mu_0} \tau' p^* \]
Appendix 4 – Some Radiometric Quantities

The quantity used in the core calculations in DORT2002 is intensity. This appendix gives the general relations between the intensity and the output variables that are derived from the intensity. The relations given here are independent of illumination and instrument geometry. Specific instances for these quantities exist in several different standardized forms with different illumination and detection conditions.

**Intensity and Flux**

Intensity, $I$, is always considered to be positive, propagating along a direction $\hat{\Omega}$. The flux $F$ through a surface depends on the angle $\theta$ between the surface normal $\hat{n}$ and the propagation direction. The flux is connected to the intensity through the relation

$$F = \int_{4\pi} I \cos \theta d\omega,$$

where $\cos \theta = \hat{n} \cdot \hat{\Omega}$ and $I$ is the intensity in an element of solid angle $d\omega$ centered on the propagation direction $\hat{\Omega}$. If $\hat{n}$ and $\hat{\Omega}$ are in opposite hemispheres, then $\cos \theta = \hat{n} \cdot \hat{\Omega}$ is negative. The integral is over all solid angles.

Incident and reflected flux for a surface are defined by

$$F_{\text{in}} = \int_{-} I_{\text{in}} \cos \theta d\omega,$$

$$F_{\text{r}} = \int_{+} I_{\text{r}} \cos \theta d\omega,$$

where - and + show that the integral is only over downward and upward directions respectively. The subscripts $\text{in}$ and $\text{r}$ denote incident and reflected flux and intensity respectively.

Transmitted flux for a surface is defined by

$$F_{\text{t}} = \int_{-} I_{\text{t}} \cos \theta d\omega,$$

but note that the intensity here is at the lower boundary of the medium, which is not the same as at the upper boundary. The subscript $\text{t}$ denotes transmitted flux and intensity.

The contribution to the flux through a surface from a finite solid angle $\psi$ is defined by

$$F_{\psi} = \int_{\psi} I \cos \theta d\omega,$$

where the integral is over the solid angle $\psi$. 
**Total Reflectance, Total Transmittance and Total Absorptance**

Total reflectance, $\rho$, total transmittance, $\tau$, and total absorptance, $\alpha$, are scalars with values between 0 and 1. They are sometimes called just reflectance, transmittance and absorptance. They are defined by

$$\rho = \frac{F_r}{F_{in}}, \quad \tau = \frac{F_t}{F_{in}}, \quad \alpha = 1 - \rho - \tau,$$

where $F_{in}$, $F_r$ and $F_t$ are incident, reflected and transmitted flux for a surface.

**Reflectance Factor and Transmittance Factor**

Reflectance factor, $R$, and transmittance factor, $T$, are functions of angle of the reflected and transmitted light respectively. For a perfect diffusor with 100% reflectance the reflectance factor is 1, or 100%, independent of angle.

The reflected intensity from a perfect diffusor, $I_{r,d}$, is constant. Therefore

$$F_{r,d} = \int I_{r,d} \cos \omega \, d\omega = I_{r,d} \int \cos \omega \, d\omega = I_{r,d} \int_{0}^{\pi} \frac{1}{2} \sin 2\omega \, d\omega = \pi \cdot I_{r,d},$$

where $F_{r,d}$ is the reflected flux for a perfect diffusor.

In addition

$$F_{r,d} = F_{in},$$

since the reflectance is 100%. This gives

$$I_{r,d} = \frac{1}{\pi} F_{r,d} = \frac{1}{\pi} F_{in},$$

The contribution to the reflected flux from a surface from a finite solid angle $\psi$ for a perfect diffusor is thus

$$F_{r,\psi,d} = \int I_{r,d} \cos \omega \, d\omega = I_{r,d} \int_{\psi} \cos \omega \, d\omega = \frac{1}{\pi} F_{in} \int_{\psi} \cos \omega \, d\omega.$$

The contribution to the reflected flux from a surface from a finite solid angle $\psi$ for a general surface is given by

$$F_{\psi} = \int I_{r} \cos \omega \, d\omega.$$
The reflectance factor is defined as the quotient of the contribution to reflected flux from a surface from a finite solid angle $\psi$, and the same quantity for a perfect diffusor. The definition is thus

$$ R = \frac{F'_\psi}{F'_{\psi,d}} = \frac{\int_{\psi} I_r \cos \theta d\omega}{\frac{1}{\pi} \int_{\psi} \cos \theta d\omega} \cdot F_{in}.$$  \hspace{1cm} (278)

If it is assumed that the intensity is approximately constant over the solid angle, i.e. $\psi$ is small or $I$ varies little, this yields

$$ R = \frac{F'_\psi}{F'_{\psi,d}} \approx \frac{I^\prime_{\psi}}{\frac{\pi}{F_{in}}} \frac{\int_{\psi} \cos \theta d\omega}{\cos \theta d\omega} = \frac{I^\prime_{\psi}}{F_{in}} \approx \pi \cdot \frac{I^\prime_{\psi}}{F_{in}}, \hspace{1cm} (279)$$

where $I^\prime_{\psi}$ is the reflected intensity propagating along a direction within the solid angle $\psi$. Then $R$ is a function of $\theta$ and $\phi$, just as $I$.

Analogously the transmittance factor is defined by

$$ T = \frac{F'_\psi}{F'_{\psi,d}} = \frac{\int_{\psi} I_t \cos \theta d\omega}{\frac{1}{\pi} \int_{\psi} \cos \theta d\omega} \cdot F_{in}.$$  \hspace{1cm} (280)

or

$$ T = \frac{F'_\psi}{F'_{\psi,d}} \approx \frac{I^\prime_{\psi}}{\frac{\pi}{F_{in}}} \frac{\int_{\psi} \cos \theta d\omega}{\cos \theta d\omega} = \frac{I^\prime_{\psi}}{F_{in}} \approx \pi \cdot \frac{I^\prime_{\psi}}{F_{in}}, \hspace{1cm} (281)$$

but here $I^\prime_{\psi}$ is the transmitted intensity propagating along a direction within the solid angle $\psi$. Then $T$ is also a function of $\theta$ and $\phi$, just as $R$ and $I$.

**BSDF, BRDF and BTDF**

**BSDF, BRDF and BTDF** are short for **Bidirectional Scattering Distribution Function, Bidirectional Reflectance Distribution Function** and **Bidirectional Transmittance Distribution Function** respectively. **BRDF and BTDF** are simply defined as

$$ BRDF = \frac{R}{\pi} \hspace{1cm} (282) $$
and

\[ BTDF = \frac{T}{\pi}, \]  \hspace{1cm} (283)

and can be said to be normalizations of \( R \) and \( T \). \( BSDF \) is used collectively for both \( BRDF \) and \( BTDF \) together.
References


See also: [ftp://climate.gsfc.nasa.gov/pub/wiscombe/Multiple_Scatt/](ftp://climate.gsfc.nasa.gov/pub/wiscombe/Multiple_Scatt/)


An excellent reference on numerical linear algebra in general is the following, which has been used throughout the model development, without explicit reference in this report.


This report is the direct continuation of a previous report. The formal reference to that report is given here.