1. INTRODUCTION

Satellite systems such as the Tropical Rainfall Measuring Mission (TRMM) are complex and rely on the proper design and functioning of each of its many subsystems—satellite vehicle, ground validation (GV) sites, precipitation radar (PR), calibration, and retrieval algorithms—to ensure high quality data products. The retrieval algorithms employed in TRMM are based on mathematical models and assumptions about the microphysics of precipitation. Errors in the model assumptions, variables input to the model, and model parameters can lead to errors in the output, which in this case, is the retrieved rain rate.

In general, systems engineering consists of the methods and/or processes used to optimize the performance of a system given limited time, technology and/or resources. By analyzing the system, the designers are able to determine which parts are important and which parts contribute little to the overall success or outcome. Likewise, in the TRMM retrieval algorithm, some input variables, parameters and model assumptions are more important than others. Global Sensitivity analysis (SA) methods and techniques can be used to assist in systems engineering analysis by providing quantitative insight as to the importance of each input factor. Saltelli (2000, 2004) defined SA as the study of how the output of a model (numerical or mathematical) varies as a function of its input parameters and how to relate the output uncertainty (variance) back to the uncertainty in each of the input parameters.

There are at least two main techniques for implementing global sensitivity analysis: Monte Carlo (MC); and variance decomposition. Each of these techniques is sampling based, meaning that samples are drawn from the probability density functions (PDFs) of each of the factors and the model is executed once for each set of sampled values. A block diagram, adapted from Saltelli (1999), of the key components and flow of global SA using the variance decomposition technique is shown in Figure 1. The model being studied, represented by $Y = f(X)$, where $X = (X_1, X_2,...X_k)$, uses sampled values from the distribution space of each factor.

As shown in Figure 1, the primary output of SA (variance decomposition) can be best illustrated by a pie chart showing the relative importance of each of the factors in relation to the total unconditional output variance. The factor distributions are sampled, the model is executed, and the total variance is decomposed. A factor that has a large impact on the total variance will be shown in the pie chart with a large percentage while a factor that has minimal influence on the output variance will have a correspondingly small percentage of the pie chart. Once SA is completed, feedback can be provided to the model structure and to the underlying assumptions used in the factor distributions.

This paper describes a systems engineering analysis of a TRMM-like (TL) retrieval algorithm using the SA technique of variance decomposition to determine and quantify the relative importance of each of the model input factors. In this case, the model output is estimated rain rate. The term TL algorithm is used to convey the notion that the algorithm studied in this paper is based on the TRMM algorithm but deviates in select aspects. The paper is outline as follows:
section 2 provides a background and mathematical basis of variance decomposition; section 3 describes the TL retrieval algorithm that is studied and the assumptions made in the model; section 4 shows the SA and UA results from those analyses; section 5 examines the implications for GPM; and section 6 is a summary of this paper.

2. VARIANCE DECOMPOSITION

For a given model with \( k \) input factors and function \( f \), the output \( Y \) can be expressed as

\[
Y = f(X_1, X_2, \ldots, X_k)
\]

where \( X_1, X_2, \ldots, X_k \) are the input factors. Variance decomposition methods include the correlation ratio described by McKay (1998), the method of Sobol’ (2001), Sobol’ (2003), Saltelli (2002), and the Fourier Amplitude Sensitivity Test (FAST) developed by Cukier (1978).

The Method of Sobol’ is used to perform variance decomposition more computationally efficient than the correlation ratio while yielding more information, Chan (1997). It is based on the decomposition of total variance given as

\[
V = \sum_{i=1}^{k} V_i + \sum_{i<j}^{k} V_{ij} + \cdots + V_{12\ldots k}
\]

where \( V \) is the total unconditional variance, \( V_i \) is the variance due to each input factor \( X_i \) by itself. \( V_{ij} \) is the variance due to each pair of input factors, \( X_i \) and \( X_j \). Additionally, higher-order interaction terms are included all the way up to \( V_{12\ldots k} \) which is the variance due to the interaction of all the input factors. A measure of sensitivity \( S_i \) for a factor can be derived by dividing each term in (2) by the total unconditional output variance \( V \) giving,

\[
\sum_{i=1}^{k} S_i + \sum_{i<j}^{k} S_{ij} + \cdots + S_{12\ldots k} = 1
\]

with the useful property that the sum of all the sensitivity indices and interactions sum to one. The \( S_i \) variables are called the first-order sensitivity indices. The \( S_{ij} \) are called the second order indices. Likewise, \( S_{ijk} \) are called third order and so on up to the highest order, \( S_{12\ldots k} \), which accounts for the interaction of all \( k \) input factors. For a model that is additive,

\[
\sum_{i=1}^{k} S_i = 1
\]

meaning the sum of the first-order sensitivity indices is one. If the \( S_i \) do not sum to one, then higher-order interactions are present. Rather than directly compute all the first and higher-order sensitivity indices, it is often sufficient to compute only the \( S_i \) and total sensitivity indices, \( S_T \) for each input factor \( X_i \). The \( S_T \) indices measure the contribution to the total variance as a result of the first and higher-order interactions of each factor \( X_i \). For example, assuming \( k = 3 \), the total sensitivity index \( S_T \) for factor \( X_1 \) would be

\[
S_T^1 = S_1 + S_{12} + S_{13} + S_{123}
\]

which accounts for all the interactions of input factor \( X_1 \) with \( X_2 \) and \( X_3 \). Similar expressions can be generated for factors \( X_2 \) and \( X_3 \).

3. TL RETRIEVAL ALGORITHM

3.1. Background

The retrieval algorithm used to estimate the TRMM rainfall rate is designated as 2A25 in the TRMM data products and is thoroughly described in Iguchi (2000). The underlying assumption of the retrieval method is that specific attenuation \( k(r) \) in dB/km, at some range \( r \) from the satellite, can be modelled as

\[
k(r) = \alpha Z_e^\beta (r)
\]

and that the rainfall rate in mm/hr can be found from

\[
R(r) = a Z_e^b (r)
\]

where \( \alpha, \beta, a \) and \( b \) are empirically-derived coefficients and the true radar reflectivity factor is designated as \( Z_e(r) \). Because of attenuation, \( Z_e(r) \) is masked and must be estimated. The observed or measured radar reflectivity factor \( Z_m(r) \) at range \( r \) is related to \( Z_e(r) \) by the two-way attenuation factor \( A(r) \),

\[
Z_m(r) = Z_e(r) A(r)
\]

and \( A(r) \) is given by

\[
A(r) = \exp\left[-0.2 \ln(10) \int_0^r k(s) ds\right]
\]

Using the assumption of (6), (8) can be solved for \( Z_e(r) \) and written as

\[
Z_e(r) = \frac{Z_m(r)}{A_{HB}(r)}
\]

where \( A_{HB}(r) \) is the Hitschfeld-Bordan (HB) derived attenuation factor,

\[
A_{HB}(r) = \left[1 - q \beta \int_0^r s \frac{Z_m(s) ds}{n^\beta}\right]^{1/\beta}
\]

with \( q = 0.2 \ln10 \). Note that \( \log = \log_{10} \) and \( \ln = \log_e \). Using a hybrid of the HB and an independent surface-reference technique (SRT), (10) and (11) can be written as
$Z_v(r) = \frac{Z_m(r)}{[1 - \varepsilon q \beta \int_0^r \alpha(s)Z_m^\beta(s)ds]^{1/\beta}} \quad (12)$

The coefficient $\varepsilon$ is found from the path-integrated-attenuation from HB (PIA$_{HB}$) and SRT ($\Delta\sigma^o$) values so that $Z_v$ can be calculated at each range bin $r$. Meneghini (2000). In this work, $\varepsilon$ is calculated using the simplistic linear function from Iguchi (1994) instead of the more detailed maximum-likelihood method described by Iguchi (2000). Normally, after the final $\varepsilon$ value is calculated, the $a$ and $b$ terms in (7) are recalculated (adjusted) using second-order logarithmic polynomials to provide the final Z-R relationship,

$$R(r) = a'Z_v^b(r) \quad (13)$$

with $R(r)$ in mm/hr such that $\varepsilon$ is linked to the DSD via the $\alpha$-correction process, Kozu (2001). The adjustment for non-uniform beam filling (NUBF) is not implemented in this work.

3.2. Model

Using equations (6) through (13), a model for the TL algorithm can be formulated according to the sensitivity analysis guidelines of the form

$$R(r) = f(Z_m, \alpha, \beta, a, b, \Delta\sigma^0) \quad (14)$$

where the rain-rate $R(r)$ is a function of the listed input factors and range $r$. Although the input profile, $Z_m$, is many bins high, the sensitivity analysis is performed using only the bottom bin nearest the ground.

3.3. Input Factors

Each of the input factors listed in (14) must have a specified PDF that can be sampled as part of the SA process. For ocean retrieval, the TL-model factor distributions are shown in Table 1.

Table 1. List of the input factors and their distributions for the TL algorithm rain-rate sensitivity analyses over ocean. The nominal values for these factors were obtained from Iguchi (2000), assuming 20° C stratiform rain.

<table>
<thead>
<tr>
<th>Factor Name</th>
<th>Probability Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>U(2.851 e-4 ± 20% )</td>
</tr>
<tr>
<td>$\beta$</td>
<td>constant = 0.792 30</td>
</tr>
<tr>
<td>$a$</td>
<td>U(0.022 82 ± 0.014)</td>
</tr>
<tr>
<td>$b$</td>
<td>U(0.672 7 ± 0.036)</td>
</tr>
<tr>
<td>$\Delta\sigma^o$</td>
<td>N(Δ$\sigma^o_{true}$, 1 dB)</td>
</tr>
<tr>
<td>$Z_m$ (error in each bin)</td>
<td>N(0, 1 dB)</td>
</tr>
</tbody>
</table>

In this work, $\alpha$ is assumed to be uniform and vary ±20% about its nominal value specified for rain, 20° C, stratiform. $\beta$ is constant and is fixed at its defined value for rain, 20° C, stratiform, Iguchi (2000).

As this paper focuses on a TL algorithm, the calculation and usage of $a$ and $b$ are handled differently than the method in the TRMM algorithm. For low rainfall rates, where $\varepsilon$ is unity (or close to unity), and $\alpha$-correction is not performed, the Z-R relationship of (7) is used to calculate rain rate using estimated PDFs (shown in Table 1) for $a$ and $b$. For higher rainfall rates, the $a$ and $b$ method given in the TRMM algorithm (log polynomials based on $\alpha$-correction) is used to calculate rain rate instead of sampling the PDFs. The TL algorithm automatically transitions from one method to the other. The uniform PDFs, used in this paper, for $a$ and $b$ were derived from global disdrometer data from Bringi (2003) using a technique based on the work of Kwiatkowski (2002) which performs Z-R curve fits and generates multiple $a$, $b$ pairs, linked to $\alpha$-correction, which can then be statistically analyzed. Bringi (2001) showed that an alternative procedure is to keep $b$ fixed and relate all variabilities of Z-R through the concept of normalized DSD. At low rain rates, statistical variation (error) from the $a$, $b$ coefficients propagates through the model, and at higher rain rates, variation (error) in $Z_m$ and $\Delta\sigma^o$ propagate through the model and are implicit in the calculation of $\varepsilon$ and in the values of $a'$ and $b'$. Note that even small variations in $a$ and $b$ in (7) (or $a'$ and $b'$ in (13)) can cause large swings in estimated rainfall rate, and consequently, the distributions of $a$ and $b$ can be very important to accurate rainfall estimation.

The value of $\Delta\sigma^o$ depends on whether it is over land or ocean. For ocean, it is assumed to be 1-dB, standard deviation, mean value equal to true PIA from the simulated data sets.

The input factor, $Z_m$, is a vertical profile (vector) of reflectivity values. It is understood that there is measurement error in each bin, and that the error is normally distributed, zero mean, with a 1-dB standard deviation. Because the sensitivity and uncertainty analysis methods don’t allow for a direct mapping of an error vector in the function (14), the $Z_m$ profiles must first be mapped to another random variable called a trigger factor or trigger variable. Saisana (2005) showed that the trigger factor serves as an index to select specific $Z_m$ profiles for SA. In this way, the error in the $Z_m$ profiles can be integrated into the model evaluations.

4. TL-ALGORITHM OCEAN RESULTS

SA was performed on the TL function (14) for the DSD pairs listed in Table 2 for ocean conditions. Each profile is based on a vertical rain column, 3-km high assuming 12 bins, each with a radar range resolution of 0.25 km.

The first-order indices, $S_a$, $S_b$, $S_{\Delta\sigma^o}$, $S_{Z_m}$, versus rain rate are shown in Figure 2.
Table 2. List of the $N_w$, $D_o$ pairs and rain rate that were used to create $Z_m$ profiles.

<table>
<thead>
<tr>
<th>$D_o$ (mm)</th>
<th>$N_w$</th>
<th>Rain Rate (mm/hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.65</td>
<td>33,800</td>
<td>1.09</td>
</tr>
<tr>
<td>0.75</td>
<td>33,900</td>
<td>2.14</td>
</tr>
<tr>
<td>0.85</td>
<td>34,300</td>
<td>3.87</td>
</tr>
<tr>
<td>0.90</td>
<td>34,600</td>
<td>5.10</td>
</tr>
<tr>
<td>0.95</td>
<td>35,000</td>
<td>6.65</td>
</tr>
<tr>
<td>1.00</td>
<td>35,500</td>
<td>8.56</td>
</tr>
<tr>
<td>1.10</td>
<td>36,700</td>
<td>13.52</td>
</tr>
<tr>
<td>1.15</td>
<td>37,400</td>
<td>17.33</td>
</tr>
<tr>
<td>1.25</td>
<td>38,900</td>
<td>26.64</td>
</tr>
<tr>
<td>1.30</td>
<td>39,700</td>
<td>32.65</td>
</tr>
<tr>
<td>1.35</td>
<td>40,500</td>
<td>39.73</td>
</tr>
</tbody>
</table>

$\Delta \sigma^o$ declines with increasing rain rate because the true-$\Delta \sigma^o$ values increase and the error associated with this variable decreases as a percentage of its true value and hence its contribution to output variance decreases.

5. GPM IMPLICATIONS

SA methods have been applied to a GPM dual-frequency retrieval algorithm based on profile-optimization techniques, Rose (2005). This method assumes that the profiles for $D_o$ and $\log(N_w)$ are linear and that an optimization routine can find appropriate top and bottom $D_o$, $N_w$ values such that internally calculated $Z_{m1}$, $Z_{m2}$ profiles match the input profiles. This GPM retrieval algorithm has been adapted to the form,

$$R(r) = f(Z_{m1}, Z_{m2})$$  \hspace{1cm} (15)

where $Z_{m1}$ and $Z_{m2}$ are the input, measured-reflectivity profiles at 13.6 and 35.6 GHz, respectively, and $R(r)$ is the rain rate. In this case, the error associated with the $Z_{m1}$ and $Z_{m2}$ profiles is zero mean, 0.5-dBZ standard deviation, normally distributed, in each bin. From SA, the first and second-order sensitivity indices, $S_{z1m1}$, $S_{z2m2}$, $S_{z1m2}$, are shown in Figure 3.

The analysis shows that at low rain rates, a significant higher-order interaction between the input factors is present, and at higher rain rates, that the error associated with $Z_{m2}$ ($S_{z2m2}$) dominates. As with the TL algorithm, trigger variables were used with both $Z_{m1}$ and $Z_{m2}$ to map them to the SA process. The implication for GPM is that the error in $Z_{m2}$ dominates the output variance and more emphasis should be applied to its measurement and processing than to that of $Z_{m1}$.

6. SUMMARY

This paper has briefly shown how global sensitivity analysis can be applied to a TL algorithm indicating which input factors are most important in contributing to output variance in the rain-rate retrieval. We have also
shown preliminary SA results and implications for a GPM retrieval algorithm based on profile optimization.

**ACKNOWLEDGEMENT**

This research was jointly supported by the NASA Precipitation Program and Los Alamos National Laboratory.

**REFERENCES**


