P5R.1 **SYSTEMS ENGINEERING ANALYSIS OF A TRMM-LIKE RETRIEVAL ALGORITHM: IMPLICATION FOR GPM DESIGN**

C. R. Rose* and V. Chandrasekar
Colorado State University
Fort Collins, CO 80523

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 $Y = f(X_i)$, where $X_i = \{X_1, X_2,\ldots X_k\}$, uses sampled values from the distribution space of each factor.

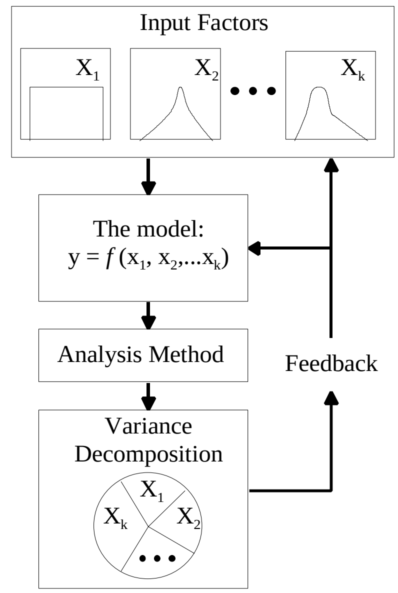

Figure 1. Block diagram of the important steps in global sensitivity analysis. Using SA the output variance can be decomposed and the relative importance of each of the input factors can be determined.

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 is as follows:

*Corresponding author address: C. R. Rose, Colorado State University
e-mail: chris.r.rose@colostate.edu

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

$$Y = f\left(X_1, X_2, ..., X_k\right) \tag{1}$$

where X_1, X_2, …, X_k are the input factors. Variance decomposition methods include the correlation ratio described by McKay (1998), the method of Sobol', 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

$$V = \sum_{i=1}^{k} V_i + \sum_{i}^{k} \sum_{j>i}^{k} V_{ij} + ... + V_{12...k} \tag{2}$$

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_{ij}. Additionally, higher-order interaction terms are included all the way up to $V_{12...k}$ 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,

$$\sum_{i=1}^{k} S_i + \sum_{i}^{k} \sum_{j>i}^{k} S_{ij} + ... + S_{12...k} = 1 \tag{3}$$

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_{ijm} are called third order and so on up to the highest order, $S_{12..k}$, which accounts for the interaction of all k input factors. For a model that is additive,

$$\sum_{i=1}^{k} S_i = 1 \tag{4}$$

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_{Ti} for each input factor X_i. The S_{Ti} 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_{T1} for factor X_1 would be

$$S_{T1} = S_1 + S_{12} + S_{13} + S_{123} \tag{5}$$

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

$$k(r) = \alpha\, Z_e^{\beta}(r) \tag{6}$$

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

$$R(r) = a\, Z_e^{b}(r) \tag{7}$$

where α, β, 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 $Z_m(r)$ ($mm^6\ m^{-3}$) at range r is related to $Z_e(r)$ by the two-way attenuation factor $A(r)$,

$$Z_m(r) = Z_e(r) A(r) \tag{8}$$

and $A(r)$ is given by

$$A(r) = \exp\left[-0.2 \ln(10) \int_0^r k(s) ds\right] \tag{9}$$

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

$$Z_e(r) = \frac{Z_m(r)}{A_{HB}(r)} \tag{10}$$

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

$$A_{HB}(r) = \left[1 - q\,\beta \int_0^r \alpha(s) Z_m^{\beta}(s) ds\right]^{1/\beta} \tag{11}$$

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

$$Z_e(r) = \frac{Z_m(r)}{\left[1 - \varepsilon\, q\, \beta \int_0^r \alpha(s) Z_m^{\beta}(s)\, ds\right]^{1/\beta}} \tag{12}$$

The coefficient ε is found from the path-integrated-attenuation from HB (PIA_{HB}) and SRT (Δσ°) values so that Z_e can be calculated at each range bin *r*, Meneghini (2000). In this work, ε 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 ε 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,

$$R(r) = a' Z_e^{b'}(r) \tag{13}$$

with *R*(*r*) in mm/hr such that ε is linked to the DSD via the α-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

$$R(r) = f\left(Z_m, \alpha, \beta, a, b, \Delta\sigma^0\right) \tag{14}$$

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 rainfall-rate sensitivity analyses over ocean. The nominal values for these factors were obtained from Iguchi (2000), assuming 20 ° C stratiform rain.

Factor Name	Probability Distribution
α	U{ 2.851 e-4 ± 20% }
β	constant = 0.792 30
a	U{ 0.022 82 ± 0.014 }
b	U{ 0.672 7 ± 0.036}
°σΔ	N{ $\Delta°\sigma_{true}$, 1 dB }
Z_m *(error in each bin)*	N{ 0, 1 dB}

In this work, α is assumed to be uniform and vary ± 20% about its nominal value specified for rain, 20° C, stratiform. β 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 ε is unity (or close to unity), and α-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 α-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 α-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 Δσ° propagate through the model and are implicit in the calculation of ε 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 Δσ° 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_α, S_a, S_b, $S_{\Delta\sigma°}$, S_{Zm}, 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.

D_O (mm)	N_W	Rain Rate (mm/hr)
0.65	33,800	1.09
0.75	33,900	2.14
0.85	34,300	3.87
0.90	34,600	5.10
0.95	35,000	6.65
1.00	35,500	8.56
1.10	36,700	13.52
1.15	37,400	17.33
1.25	38,900	26.64
1.30	39,700	32.65
1.35	40,500	39.73

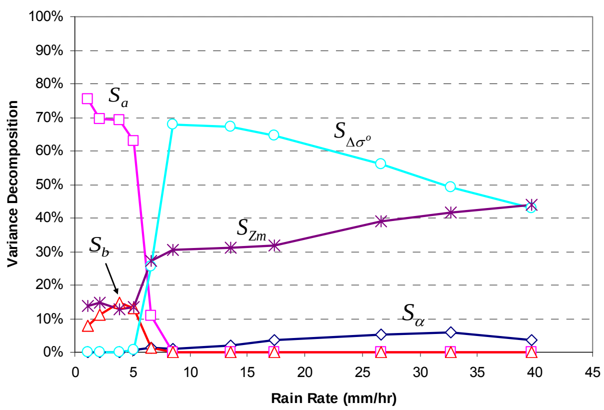

Figure 2. Graph showing the first-order variance decomposition of the important factors in the TRMM-like algorithm over ocean. The factors *a* and *b* have significant impact in the low rainfall region but diminish as the algorithm transitions to higher region where the log polynomials are used to calculate *a* and *b*. At high rain rates, the error in $\Delta\sigma°$ dominates.

From the SA results, there is a clear peak in the contribution of α to the total output variance around the 30 mm/hr range. At low rain rates, the S_a term dominates while the contribution from S_α and $S_{\Delta\sigma°}$ are negligible. In the low rain-rate region, the TL algorithm uses the HB solution which does not use the $\Delta\sigma°$ input values. In the higher rain-rate region, the $\Delta\sigma°$ input is used more significantly in the retrieval algorithm. A transition between HB and $\Delta\sigma°$ methods occurs between about 5-8 mm/hr. In both the lower and upper regions, the sum of the three, first-order sensitivity factors is approximately 100% at each rain-rate indicating that the TL model of (14) is additive. However, in the transition region, the model is not purely additive and higher-order sensitivity interactions occur. Above the transition, $S_{\Delta\sigma°}$ dominates the rain rate variance. As expected, the contribution from the error in $\Delta\sigma°$ declines with increasing rain rate because the true-$\Delta\sigma°$ 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,

$$R(r) = f\left(Z_{m1}, Z_{m2}\right) \quad (15)$$

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_{Zm1}, S_{Zm2}, S_{Zm12}, are shown in Figure 3.

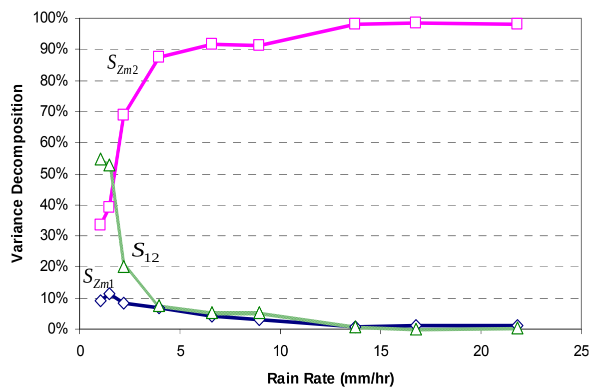

Figure 3. Graph showing the first and higher-order variance decomposition of the two input factors for the GPM profile-optimization method. At low rain rates, the model is not additive where significant higher-order interactions are present. At higher rain rates, the model is additive. The error in Z_{m2} is the primary contributor to total output variance throughout the entire rain rate region.

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_{ZM2}) 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

Bringi, V. N., V. Chandrasekar, 2001: *Polarimetric Doppler Weather Radar*, Cambridge University Press.

Bringi, V. N., V. Chandrasekar, J. Hubbert, 2003: Raindrop Size Distribution in Different Climatic Regimes from Disdrometer and Dual-Polarized Radar Analysis. *J. Atmos. Sci.*, Vol. 60, January, 354-365.

Chan, K., A. Saltelli, S. Tarantola, 1997: Sensitivity Analysis of Model Output: Variance-Based Methods Make the Difference. Proceedings Winter Simulation Conference.

Cukier, R. I., H. B. Levine, K. E. Shuler, 1978: Nonlinear Sensitivity Analysis of Multiparameter Model Systems. *Journal of Computational Physics*, 26, 1-42.

Iguchi, T., R. Meneghini, 1994: Intercomparison of Single-Frequency Methods for Retrieving a Vertical Rain Profile from Air-borne or Space borne Radar Data. *J. Atmos. Oceanic Technol.*, Vol. 11, 1507-1516.

Iguchi, T., T. Kozu, R. Meneghini, J. Awaka, K. Okamoto, 2000: Rain-Profiling Algorithm for the TRMM Precipitation Radar. *Journal of Applied Meteorology*, Vol. 39, December, 2038-2052.

Kozu, T, T. Iguchi, 2001: Drop Size Distribution (DSD) Models for TRMM 2A25. TRMM PR User's Manual.

Kwiatkowski, J., R. Meneghini, A. Tokay, 2002: On Incorporating Drop Size Measurements into the TRMM Satellite Radar Rain Retrieval Algorithm. *TRMM International Science Conference*, 22-26 July, Honolulu, Hawaii 2P.12 p. 23.

McKay, M. D., J. D. Morrison, S. C. Upton, 1998: Evaluating Prediction Uncertainty in Simulation Models. Los Alamos National Laboratory, LA-UR-98-1362.

Meneghini, R., T. Iguchi, T. Kozu, L. Liao, K. Okamoto, J. A. Jones, and J. Kwiatkowski, 2000: Use of the Surface Reference Technique for Path Attenuation Estimates from the TRMM Precipitation Radar. *J. Appl. Meteor.*, 39, 2053-2070.

Rose, C. R., V. Chandrasekar, 2005: A GPM Dual Frequency DSD Retrieval Method Based on Linear Model for DSD Vertical Profile. Invited Talk, *IGARSS'05*, July 25-29, Seoul, Korea.

Saisana, M., A. Saltelli, S. Tarantola, 2005: Uncertainty and Sensitivity Analysis Techniques as Tools for the Quality Assessment of Composite Indicators. *Journal of the Royal Statistical Society A*, Vol. 168, Issue 2, pp 307, March.

Saltelli, A., S. Tarantola, K. Chan, 1999: A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. *Technometrics*, Vol. 41, No. 1.

Saltelli, A., S. Tarantola, F. Campolongo, 2000: Sensitivity Analysis as an Ingredient of Modeling. *Statistical Science*, Vol. 15, No. 4, 377-395.

Saltelli, A., 2002: Making Best Use of Model Evaluations to Compute Sensitivity Indices. *Computer Physics Communications*, Vol. 145, pp.280-297.

Saltelli, A., S. Tarantola, F. Campolongo, M. Ratto, 2004: *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, J. Wiley & Sons.

Sobol', I., 1993: Sensitivity Analysis for Nonlinear Mathematical Models. *Math. Modelling Comput. Exp.* 1 407-414.

Sobol', I. M., 2001: Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates. *Mathematics and Computers in Simulation*, Vol. 55, Issue 103.