# Communications

## Stack Algorithm For Recursive Deconvolution of Bernoulli-Gaussian Processes

JÉRÔME IDIER AND YVES GOUSSARD

*Abstract*—**This paper deals with the deconvolution of pulse trains modeled as Bernoulli-Gaussian processes. The detector presented here implements a stack Viterbi algorithm with a parametric level of suboptimality. Simulations results are satisfactory, even in the difficult case of a poor spectral content of the wavelet. The recursive and parallel structure of the method allows fast data processing on modern architectures.**

### I. Introduction

This paper deals with the deconvolution of sparse spike trains modeled as Bernoulli-Gaussian (B-G) processes. This is a detection-estimation problem which cannot be solved optimally due to the size of the signals to be restored. Performances of deconvolution procedures are closely related to their level of suboptimality; i.e., to the manner in which possible realizations of the Bernoulli sequence are explored.

Here, we present a recursive approach to this problem based upon the maximization of a MAP criterion. The detector is called the *stack algorithm* by analogy to a suboptimal Viterbi algorithm (VA) described in [1]. The suboptimality level is parameterized and can be easily adjusted according to the characteristics of the problem at hand.

Other recursive approaches to B-G deconvolution have been proposed. The method described here is an efficient extension of the recursive detector recently introduced in [3]. A suboptimal Viterbi detector had already been proposed by Chi and Mendel in [2], but the stack algorithm presents better convergence properties (Section III-B), and the updating equations have been arranged so as to minimize the computational load (Section III-C). Finally, the intrinsically parallel structure of the algorithm makes it suitable for implementation on modern architectures.

### II. System Representation and Problem Formulation

#### A. Models and Notation

Models and notation are taken from [3]. The observed data are assumed to be generated by the discrete-time convolution equation,

$$z(k) = \sum_{i=0}^{n} h(i)x(k-i) + n(k) \qquad (k = 1, \cdots, N) \quad (1)$$

where $\boldsymbol{h} = [h(n), h(n-1), \cdots, h(0)]'$ denotes the finite-memory impulse response (IR) of the system (the prime sign denotes the transpose operator). The input $\boldsymbol{x} = [x(-n+1), \cdots, x(0), x(1), \cdots, x(N)]'$ is modeled as a B-G signal: $x(k)$, $(k = -n+1, \cdots, N)$ is a Gaussian random variable with zero mean and covariance $r^x t(k)$, where $t(k)$ is a white binary sequence defined by $P\{t(k) = 1\} = \lambda$ and $P\{t(k) = 0\} = 1 - \lambda$. The additive noise $n(k)$ is assumed to be a zero-mean white Gaussian process with variance $r^n$.

Blind or myopic deconvolution is beyond the scope of the present paper; the set of hyperparameters $(r^n, r^x, \lambda)$ and the IR $\boldsymbol{h}$ of the system are assumed to be known *a priori*.

#### B. The Maximum a posteriori (MAP) Approach

The problem of estimating the input $\boldsymbol{x}$ given the observed data $z$ is powerfully resolved by the maximization of a likelihood function. Here, we decide to maximize the *a posteriori* likelihood criterion,

$$J(\boldsymbol{t}, \boldsymbol{x}|z) \propto p(z|\boldsymbol{t}, \boldsymbol{x})p(\boldsymbol{x}|\boldsymbol{t})P(\boldsymbol{t}). \quad (2)$$

The choice between a MAP approach and other types of likelihood methods will not be discussed here. In fact, the algorithm presented here can cope with a *maximum likelihood* (ML) approach [3] equally well.

Since the above criterion is quadratic in $\boldsymbol{x}$ conditionally to the knowledge of $\boldsymbol{t}$, a two-step procedure is well suited to perform its maximization. Such an approach has been used by several authors [3], [4]. It reduces the problem to the determination of a binary sequence $\hat{\boldsymbol{t}}$ which maximizes $J(\boldsymbol{t}, \hat{\boldsymbol{x}}(\boldsymbol{t})|z)$ as defined in (2), where $\hat{\boldsymbol{x}}(\boldsymbol{t})$ represents the MAP estimate of $\boldsymbol{x}$ conditionally to $\boldsymbol{t}$:

$$\hat{\boldsymbol{x}}(\boldsymbol{t}) = \Pi \boldsymbol{H}'(\boldsymbol{H}\Pi\boldsymbol{H}' + r^n\boldsymbol{I})^{-1}z \quad (3)$$

$\Pi \triangleq r^x \operatorname{Diag}\{\boldsymbol{t}\}$ is the *a priori* covariance matrix of $\boldsymbol{x}$ conditionally to $\boldsymbol{t}$ and matrix $\boldsymbol{H}$ contains the shifted samples of the system IR. $J(\boldsymbol{t}, \hat{\boldsymbol{x}}(\boldsymbol{t})|z)$ must be maximized over the $2^N$ possible realizations of $\boldsymbol{t}$ (we assume that $t(k) = 0$ if $k \leq 0$) and $N$ is generally far too large to allow an exhaustive search of the solution; thus, we are led to suboptimal methods.

### III. The Stack Recursive Detector

#### A. Suboptimal Detection

Throughout the rest of the paper, the superscript * denotes the unknown sequence that maximizes the MAP criterion, and the subvector $[v(1), \cdots, v(k)]'$ of any vector $\boldsymbol{v}$ is noted $\boldsymbol{v}^k$.

Recursive detection consists of sequentially finding a sequence $\hat{\boldsymbol{t}}^k$ that (almost) maximizes the partial criterion $J_k(\boldsymbol{t}^k, \hat{\boldsymbol{x}}^k(\boldsymbol{t}^k)|z^k)$, thanks to the previous partial solution $\hat{\boldsymbol{t}}^{k-1}$. Most recursive detectors are based upon the following assumption: if $\hat{\boldsymbol{t}}^{k-1}$ is close to $\boldsymbol{t}^{k-1}*$, then at least one of the two sequences $\hat{\boldsymbol{t}}_0^k = [\hat{\boldsymbol{t}}^{k-1}|0]'$ and $\hat{\boldsymbol{t}}_1^k = [\hat{\boldsymbol{t}}^{k-1}|1]'$ is a *plausible* solution to the maximization of $J_k(\boldsymbol{t}^k, \hat{\boldsymbol{x}}^k(\boldsymbol{t}^k)|z^k)$. *Plausible* means that the corresponding value of the criterion is close to the optimal $J_k(\boldsymbol{t}^k*, \hat{\boldsymbol{x}}^k(\boldsymbol{t}^k*)|z^k)$. It should be kept in mind that the above assumption is merely qualitative and empiric. As a justification, let us observe that if the wavelet is a true spike, then $\boldsymbol{t}^k* = [\boldsymbol{t}^{k-1}*|\delta_k]'$ (where $\delta_k = 0$ or $\delta_k = 1$). When the IR has a limited spectral content and is therefore far from a spike, the above assumption is hardly fulfilled. This explains that most existing recursive methods exhibit some oscillations in this difficult situation. These unstabilities are typical of *decision-directed* procedures.

#### B. The Stack Algorithm

In order to improve the stability of the detector, we propose to build $L$ sequences at the same time, the final solution being the one that maximizes $J_N$ ($L$ is the parametric *stack length* to be chosen

Manuscript received October 2, 1989; revised March 7, 1990.

The authors are with the Laboratoire des Signaux et Systèmes (CNRS/ESE/UPS), Ecole Supérieure d'Electricité, 91192 Gif-sur-Yvette Cédex, France.

IEEE Log Number 9037448.

0196-2892/90/0900-0975$01.00 © 1990 IEEE

between 1 and $2^{N-1}$). Stepping from time $k-1$ to time $k$ gives rise to $2L$ plausible sequences $t^k$ from $L$ sequences $t^{k-1}$, so it is necessary to select half of those $2L$ available sequences before performing the next recursion. To do this, the $2L$ sequences are graded according to the value of their partial criterion $J_k(t^k, \hat{x}^k(t^k)|z^k)$ (at time $k$), and every sequence in the second half of the stack is rejected.

The procedure is initialized at time $k = 1$ with the only two possible sequences $t_0^1 = [0]$ and $t_1^1 = [1]$. Next, the stack memorizes the whole $2^k$ sequences of length $k$ until its parametric capacity $L$ is reached. Only then are the overflowing, less probable sequences rejected.

Let $\hat{x}(L)$ denote the best final sequence among the $L$ stacked solutions and $J(L)$ the corresponding value of the criterion. Two interesting properties are straightforward to bring out:

a) $\forall L_1, L_2, \quad L_2 > L_1 \Rightarrow J(L_2) \geq J(L_1)$

b) $\hat{x}(2^{N-1}) = x^*$.

Their conjunction stands for monotone convergence of $J(L)$ towards the optimal value $J^*$. Practically, such a property ensures that an increase of $L$ improves the performance of the method. Note that the numerical cost is increased accordingly. The choice of a value of $L$ clearly results from a compromise between the computational effort and the quality of the solution, and depends on the specific characteristics of the problem at hand. Practical experience shows that the signal-to-noise ratio and the spectral content of the wavelet are the main parameters which control the difficulty of the problem, so that admissible choices for $L$ will greatly depend on them. From various simulations performed with a 10 dB signal-to-noise ratio, it has been possible to infer the following empirical law which roughly estimates a minimal value of $L$ for a given wavelet:

$$\hat{L} \approx 10^{1/(5F)} \tag{4}$$

where $F$ denotes the normalized frequency such that 95% of the energy of the wavelet is contained in the frequency band $[0, F]$. $\hat{L}$ must be thought as an approximate minimal value of $L$ which provides acceptable results. The exponential relation between $1/F$ and $\hat{L}$ in (4) accounts for the high sensitivity of such a deconvolution method to the spectral content of the IR. It is not so surprising that simple recursive algorithms perform poorly when the wavelet lacks high frequencies.

Comparison of the procedure presented here to Chi and Mendel's detector [2] and to the optimal VA [1] shows the following:

1) The optimal VA would have a growing finite-state $t^k$ at time $k$ (with $2^k$ possible values).

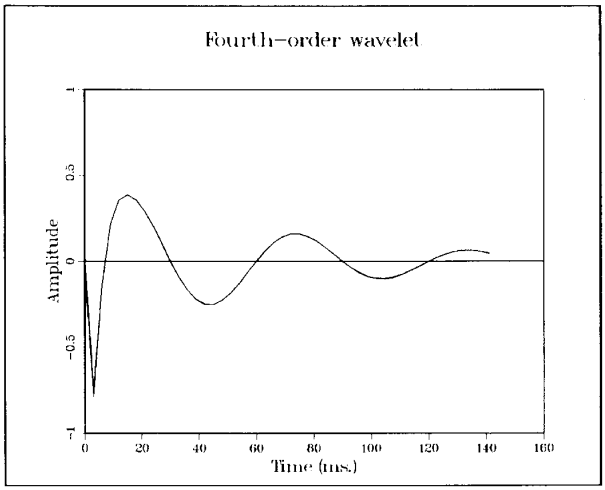
2) Chi and Mendel's detector is a suboptimal VA, with a reduced finite-state $T(k) = [t(k-l+1), \cdots, t(k-1), t(k)]'$ at time $k$.

3) On the other hand, the detector described here memorizes $L$ *complete* states among the $2^k$ possibilities (at time $k$), which are the most plausible among $2L$ trials.

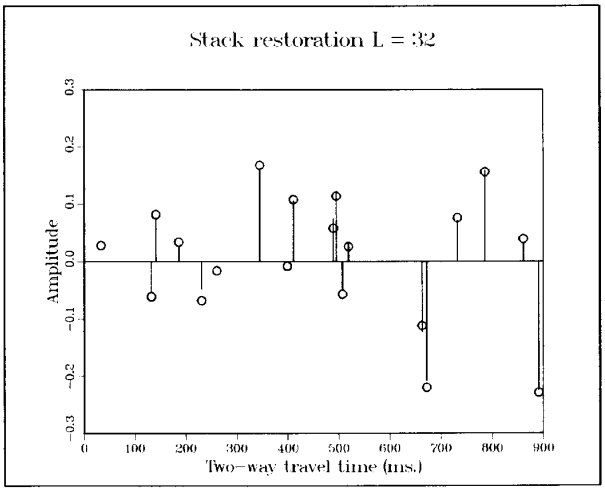
The two suboptimal methods exhibit strong similarities. However, *exhaustive* comparison of the criterion values is proposed here, whereas in [2], likelihood values are only compared two by two to determine the rejected sequences. Such a *partial* comparison is rather arbitrary, and it does not ensure the monotony property (property a), so that Chi and Mendel's detector does not provide monotone convergence to the true solution with respect to the complexity of exploration. From a practical standpoint, this means that increasing the stack size may deteriorate the quality of the results. The stack algorithm does not present such an awkward characteristic.

### C. Updating Equations

Given a Bernoulli sequence, the amplitude vector can be computed recursively by Kalman filtering, whose updating equations are given in [3]. The computation of the *a posteriori* log-likelihood can also be found elsewhere [2]–[4].




(a)




(b)

Fig. 1. Simulation results obtained in the favorable case of a broad-band wavelet. (a) Fourth-order source wavelet (from [7]). (b) Deconvolution results obtained with the stack algorithm detector for a stack size $L = 32$. The circles depict the actual reflectivity (from [5]) and the estimated signal is shown in solid lines.

At time $k$, the criterion must be evaluated for each of the $2L$ sequences generated by the stack algorithm, so as to keep the $L$ best sequences. The updating of covariance matrix $P$ can be performed subsequently to the sorting operation, so that it need not concern the $L$ rejected trials: it suffices to define and compute only $L$ matrices instead of $2L$. Moreover, elementary, although tedious, considerations about the structure of matrices $P$ show that only columns $k-n-1$ through $k$ of $P_{k/k-1}$ are involved in the $k$th recursion of the algorithm. In terms of storage capacity and computations, updating the covariance matrices is the main computational burden. The above considerations allow us to deal with only $L \times n \times N$ elements of matrices, instead of $2L \times N \times N$ for comparable algorithms (such as [2]).

The resulting algorithm is plainly parallel since $2L$ sequences of equations must be computed independently at each recursion: a massively parallel implementation of the detector makes it almost as fast as a unique Kalman filter.

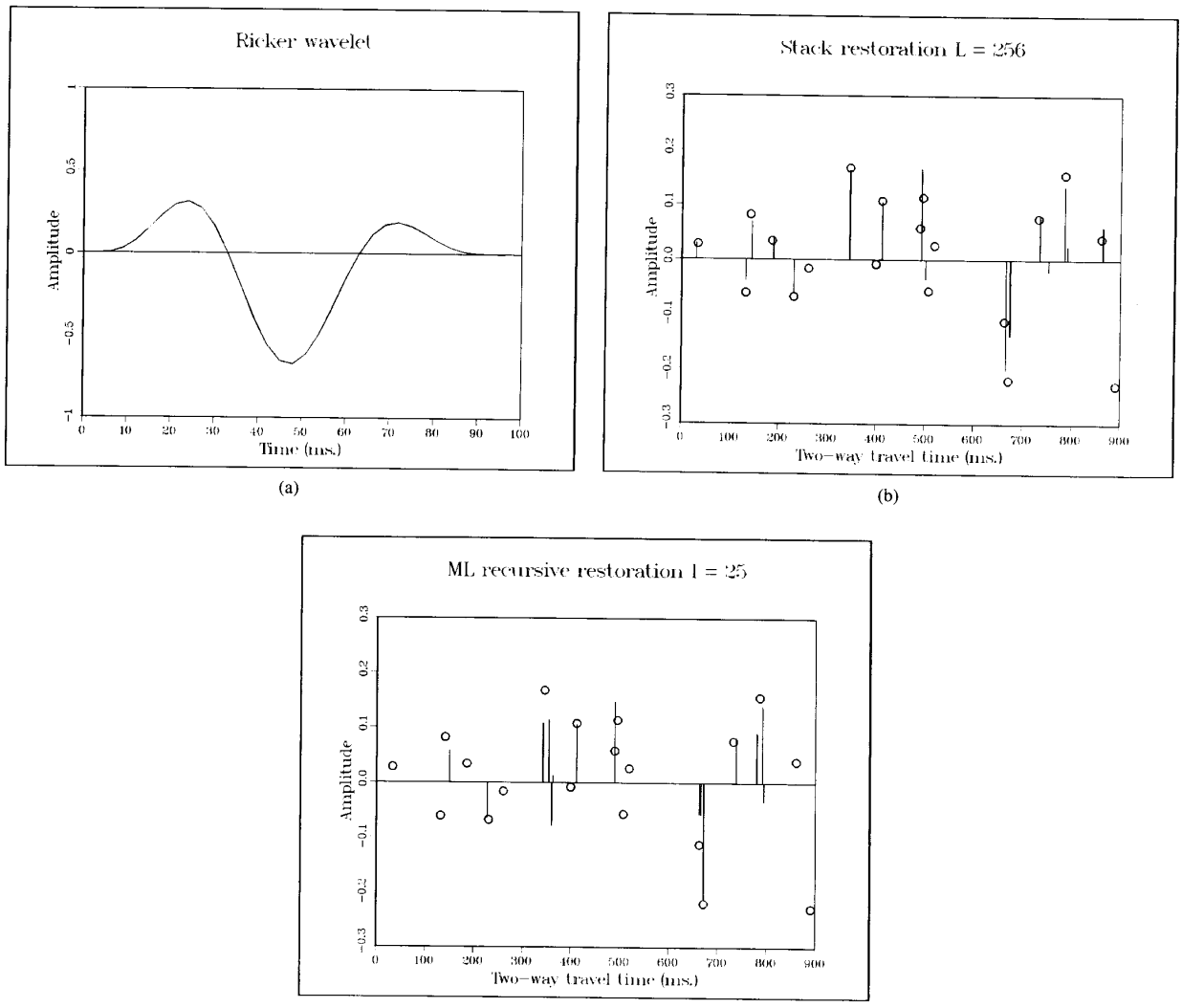



Fig. 2. Simulation results obtained in the difficult case of a narrow-band low-pass wavelet. (a) Ricker wavelet (from [5]). (b) Deconvolution results obtained with the stack algorithm detector for a stack size $L = 256$. (c) Deconvolution results obtained with the ML recursive detector for a 25-sample detection delay (from [3]). The circles depict the actual reflectivity (from [5]) and the estimated signal is shown in solid lines.

## IV. Simulation Results

Many tests were carried out on synthetic data in order to study the influence of stack size $L$ on the quality of the results and to compare them to previously published simulations. To make the comparison easier, the examples presented here were obtained with the reflectivity sequence defined by Mendel and Kormylo in [5], and used later by Kollias and Halkias [6]. This true sequence is depicted by circles in Fig. 1(b) and Fig. 2(b) and (c).

In both examples, the signal-to-noise ratio, defined as the ratio between the mean power of the noiseless observations and the variance of the observation noise $r^n$, was set to 10 dB. In the first test, the system IR was a fourth-order Kramer wavelet [7] (Fig. 1(a)). Stack deconvolution performs well with no need for high storage capacities: $L = 8$ gives a satisfactory solution, while a nearly global maximum of the *a posteriori* likelihood is reached for $L = 32$ (Fig. 1(b)). The second example deals with a smoother wavelet (Fig. 2(a)). In this case, the stack algorithm performs much better than the ML recursive method with a detection delay proposed in [3] (Fig. 2(b) compared to Fig. 2(c)). The price to pay for good quality of the results is an increase in the stack size: acceptable solutions were obtained for $L = 256$, while values of $L$ smaller than 64 provided poor results. These results are in agreement with the empirical formula of (4).

## V. Conclusion

A poor spectral content of the system impulse response decreases the quality of recursive deconvolution of Bernoulli-Gaussian processes. A recursive *stack algorithm* which can be interpreted as a suboptimal Viterbi algorithm was presented. It allows a more exhaustive exploration of plausible Bernoulli sequences, and its complexity can be adjusted through the *stack length* parameter $L$. A parallel implementation allows fast processing.

The stack algorithm has been successfully tested on synthetic data using a maximum *a posteriori* criterion. It was shown to solve this maximization problem quasi-optimally for any type of wavelet, provided that fitted processing means are available.

## References

[1] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268-278, 1973.

[2] C. Y. Chi and J. M. Mendel, "Viterbi algorithm detector for Bernoulli-Gaussian processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 511-519, 1985.

[3] Y. Goussard and G. Demoment, "Recursive deconvolution of Bernoulli-Gaussian processes using a MA representation," *IEEE Trans. Geosci. Remote Sensing*, vol. 27, pp. 384-394, 1989.

[4] J. M. Mendel, *Optimal Seismic Deconvolution*. New York: Academic, 1983.

[5] J. M. Mendel and J. Kormylo, "Single-channel white-noise estimators for deconvolution," *Geophys.*, vol. 43, pp. 102-124, 1978.

[6] S. D. Kollias and C. C. Halkias, "An instrumental variable approach to minimum-variance seismic deconvolution," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-23, pp. 778-788, 1985.

[7] F. J. Kramer, R. W. Peterson, and W. C. Walter, Eds., *Seismic Energy Sources 1968 Handbook, 38th Annual Meeting Society Exploration Geophys.*, Denver, CO, 1968.


# Attenuation of Soil Microwave Emission by Corn and Soybeans at 1.4 and 5 GHz

THOMAS J. JACKSON, AFFILIATE MEMBER, IEEE, AND PEGGY E. O'NEILL

*Abstract*—**Theory and experiments have shown that passive microwave radiometers can be used to measure soil moisture. However, the presence of a vegetative cover alters the measurement that might be obtained under bare conditions. Significant obstacles to the practical use of this approach are deterministically accounting for the effect of vegetation and developing algorithms for extracting soil moisture from observations of a vegetation-soil complex. The presence of a vegetation canopy reduces the sensitivity of passive microwave instruments to soil moisture variations. Data collected using truck-mounted microwave radiometers were used to examine the specific effects of corn and soybeans canopies.**

*Keywords*—**Radiometer, soil, moisture, vegetation, canopy, water content.**

## I. Introduction

Although attempts have been made to model the soil-vegetation system, even the most sophisticated approaches must utilize some approximations and parameterizations due to the complex nature of this target. Ultimately, these effects must be incorporated into a deterministic algorithm which utilizes parameters that can be readily measured, hopefully using remote sensing.

A simple model was proposed [1] to account for vegetation effects that utilized the vegetation wet biomass or water contents as the only canopy parameter. Additional studies demonstrated that an improvement could be made by including the single scattering albedo in the modeling approach [2], [3].

Further improvements in vegetation effect models are possible if the canopy structure can be incorporated into the model. One approach to solving this problem is to develop a physically based electromagnetic model of the system. This requires a detailed knowledge of the dielectric properties of the plant constituents [4] and a model to represent the complex and highly variable agrometrical distribution. This approach can work and is valuable in furthering our understanding [5].

An alternative approach to accounting for structure is to develop a parametric representation of the vegetation which could be readily implemented using exciting remotely sensed data. It is hypothesized here that canopy structure is basically a function of the structural crop type, and that for each crop a unique relationship between vegetation water content and attenuation can be established. Multispectral remote sensing could be used to perform crop classification and vegetation water content estimation. This hypothesis was examined using data collected by truck-mounted *C*- and *L* band radiometers over controlled-condition corn and soybean fields.

Manucript received October 16, 1989; revised March 8, 1990.

T. J. Jackson is with the USDA-ARS Hydrology Laboratory, Room 205, Building 265, BARC-East, Beltsville, MD 20705.

P. E. O'Neill is with the Hydrological Sciences Branch/924 NASA/Goddard Space Flight Center, Greenbelt, MD 20771.

IEEE Log Number 9036087.

## II. Microwave Emission From Vegetation-Covered Fields

Any attempt to model the effects of vegetation on microwave emission must make a number of assumptions and simplifications, since a vegetation canopy is an extremely physical system and is highly variable in all dimensions and parameters. On the scale of an agricultural field a canopy does have some degree of uniformity, and it is the vegetation effects at this scale that are of interest here.

Following the development by Ulaby *et al.* [6] for a uniform layer of vegetation at a given incidence angle and polarization,

$$\mathrm{TB}_c = (1 + R_B\alpha)(1 - \gamma)(1 - \alpha)T_v + (1 - R_B)\gamma T_B \quad (1)$$

where

- $\mathrm{TB}_c$ brightness temperature of the canopy observed by a radiometer (K),
- $T_v$ physical temperature of the vegetation canopy (K),
- $T_B$ physical temperature of the background media (K),
- $\gamma$ transmissivity of the vegetation layer,
- $R_B$ air-background reflectivity,
- $\alpha$ single-scattering albedo.

The transmissivity $\gamma$ is expressed as a function of the optical depth $\tau$ and incidence angle $\theta$ as follows:

$$\gamma = \exp(-\tau \sec\theta). \quad (2)$$

The model described by (1) has been successfully applied for a variety of crops and frequencies [2], [3], [6], [7].

Equation (1) can be simplified if $\alpha$ is assumed to be zero and that $T_v \approx T_B$ [1]. Under these assumptions,

$$e_c = \frac{\mathrm{TB}_c}{T_v} = 1 - R_B \exp(-2\tau) \quad (3)$$

where $e_c$ is the emissivity of the canopy.

Single-scattering albedo should be a function of plant agrometry, polarizaiton, and frequency and, therefore, should vary with the crop type, planting pattern, and stage of growth [2]. There is no established physical relationship between $\alpha$ and any of these parameters. Based on (1) and (2), when the attenutation by the canopy is small ($\mathrm{TB}_c \approx \mathrm{TB}_B$) and the background is cold ($\mathrm{TB}_B = 200$), the value used for $\alpha$ is not important. Sensitivity increases as the canopy temperature increases and diverges from the background temperature.

There are three approaches that can be used to deal with the single-scattering albedo: 1) Assume $\alpha = 0$; 2) use literature values; and 3) estimate through experimental conditions.

The first approach was used with some success in [1]. This approach presumes that the variability in $\alpha$ is small and that the effects of $\alpha$ can be incorporated in $\tau$.

With respect to the second approach, there is limited amount of data on $\alpha$ in the literature. Representative values for selected conditions considered in the current study are: Corn—1.4 GHz $\alpha$ =