QUALITY ASSESSMENT OF THE ORIENTED VOLUME OVER GROUND (OVOG) MODEL FOR POLINSAR RETRIEVAL ALGORITHMS APPLIED TO AGRICULTURAL CROPS

Emilio Ungría J. David Ballester-Berman Juan M. Lopez-Sanchez
Departamento de Física, Ing. Sistemas y Teoría de la Señal (DFISTS), University of Alicante
P.O. Box 99, E-03080 Alicante, Spain
Phone / Fax: +34965909597 / 9750, E-mail: juanma-lopez@ieee.org

ABSTRACT
A modification of the well-known random volume over ground (RVoG) model [1] has been recently proposed to accommodate its working principle to the case of oriented vegetation volumes, which are very common in agricultural crops [2]. The modified model, also called OVoG, can be applied to the retrieval of vegetation height and ground topography of agricultural crops, provided that the spatial distribution of the coherence values on the complex plane fit a narrow triangular shape.

The first stage of the inversion algorithm is identical to the original RVoG procedure: When the triangle is narrow enough, and the coherences are sufficiently separate in the complex plane, a line model can be fitted to them in order to extract the ground topography.

In this paper we analyze the quality of the least squares fit by computing a quality factor, in a way similar to [3]. In our case, the quality test should provide an indication about the narrowness of the triangle or, equivalently, the applicability of the OVoG model to the data. This quality test provides a measure about the error produced in the ground topography estimation and its influence on the rest of retrieved parameters.

This approach has been tested with indoor experimental data from the European Microwave Signature Laboratory (EMSL) at JRC-Ispira, Italy, with samples of maize and rice crops.

1 INTRODUCTION
The first step in the inversion of the Random Volume over Ground model [1] and the Oriented Volume over Ground model [2], [4], is the fitting of the coherence points by a line in the complex plane. The fitted line intersects the unit circle at two points: one of them is the interferometric phase corresponding to the ground topography, and the other is a false solution that should be rejected by the inversion algorithm. The ground interferometric phase biases all coherence points, so it must be estimated carefully in order to obtain correct estimates of the rest of biophysical parameters: vegetation height and extinction coefficient.

In this document, the performance of two line fitting methods and their effects over the ground interferometric phase estimation, and the subsequent inversion results, is studied. The two line fitting methods are presented in Section 2. Then, two quality factors are described in Section 3. In Section 4, the fitting methods are tested with real data obtained by a wide band fully polarimetric and interferometric radar. Finally, the conclusions of this work are summarized in Section 5.

2 LINE FITTING METHODS
Some line fitting methods were proposed in [1]. Among them, the least squares fitting method (LS) and the total least squares fitting method (TLS) are the simplest and require less computational burden, so they are the most suitable for elaborating large maps of topographic and vegetation heights because the inversion algorithm must be applied to each pixel of the map. Therefore, the performance of these two methods must be studied in order to obtain some knowledge about the reliability of the inverted parameters.

When the least squares line fitting is used, the imaginary parts of the coherence points are fitted as if they were a linear function of their real parts:

$$\Im(\gamma_j) = \tilde{m}\Re(\gamma_j) + \tilde{c}$$ (1)

where $\Re()$ and $\Im()$ denote the real and imaginary parts, respectively, $\gamma_j$ denotes a generic coherence point, and $\tilde{m}$, $\tilde{c}$ are the fitted line parameters. The resulting expressions of $\tilde{m}$, $\tilde{c}$ are shown in (2–3), where $N$ is the number of coherence points to be fitted. A detailed demonstration can be consulted in [5].
\[
S = \sum_{j=1}^{N} \frac{1}{\sigma_j^2} \quad S_x = \sum_{j=1}^{N} \frac{\Re(\gamma_j)}{\sigma_j^2} \quad S_y = \sum_{j=1}^{N} \frac{\Im(\gamma_j)}{\sigma_j^2} \quad S_{xy} = \sum_{j=1}^{N} \frac{\Re(\gamma_j)\Im(\gamma_j)}{\sigma_j^2} \quad S_{xx} = \sum_{j=1}^{N} \frac{\Re(\gamma_j)^2}{\sigma_j^2}
\]  

(2)

The parameter \(\sigma_j\) shown in (2) is the standard deviation of each coherence point. This parameter is used to weight the data with the aim of making the fitted line to pass closer to the points measured more exactly. To avoid underestimating the measurement errors, the real and imaginary parts of the coherence points must be considered in the computation of the standard deviation [3]. At this stage one could use any other kind of measurement error estimator instead of the standard deviation. Consequently, the performance of the fitting method will vary according to the exactitude of the error estimator. In the case of not having any estimator of the measurement errors, one can fix \(\sigma_j\) to 1 and compute a not weighted line.

The goodness of the least squares fitting has been already tested with real data in [3], but this method exhibits some limitations. The first problem arises when the slope of the fitted data, \(\tilde{m}\), is too high. As the slope increases, the performance of the fitting method gets worse. Accordingly, if the coherence points are aligned vertically in the complex plane, the representation of the fitted line by means of (1) is not possible. A second disadvantage is that the fitting method only minimizes errors along the vertical axis (imaginary axis) but the measurement errors affect both imaginary and real parts of the coherence.

The total least squares fitting [6] can avoid these limitations because it fits the real and the imaginary parts of the coherence points, as shown in (5).

\[
B = \begin{bmatrix}
\Re(\gamma_1) - w_x & \Im(\gamma_1) - w_y \\
\Re(\gamma_2) - w_x & \Im(\gamma_2) - w_y \\
\vdots & \vdots \\
\Re(\gamma_N) - w_x & \Im(\gamma_N) - w_y 
\end{bmatrix}
\]  

(6)

The matrix \(B\) can be decomposed into singular values, as expressed in the following:

\[
B_{N \times 2} = U_{N \times N} \cdot \Sigma_{N \times 2} \cdot V^T_{2 \times 2},
\]  

(7)

where the entries \(\Sigma(1,1) \geq \Sigma(2,2)\) are the singular values of \(B\), and the columns of matrix \(V\) contain the singular vectors of \(B\). The optimum value of \(r\) is the singular vector corresponding to the smallest singular value of the matrix \(B\). The relationship between \(r\) and \(r^\perp\) is expressed as:

\[
r = \begin{bmatrix} r_x \\ r_y \end{bmatrix} \rightarrow r^\perp = \begin{bmatrix} -r_y \\ r_x \end{bmatrix}.
\]

(8)

Note that a weighted version of the total least squares fitting can be computed if one divides each row of \(B\) by \(\sigma_j\).

3 QUALITY FACTORS

In this section we present, in first place, two quality estimators for a line fitted by means of the simple least squares fitting. Then, the two estimators are also adapted to the total least squares fitting method.
Finally, the new expression of the regression index for total least squares results in:

\[ \chi^2(\bar{m}, \bar{c}) = \sum_{j=1}^{N} \left( \frac{3(\gamma_j) - \bar{m} \Re(\gamma_j) - \bar{c}}{\sigma_j} \right)^2 \]  

(9)

The weighting factor is the measurement error estimator used in (2). The fitting error computed by the Chi square depends on the number of points to be fitted, \( N \). If one only wants to compare the quality of fittings carried out over the same number of points, one only has to compute their values of Chi square, and the best fit will be that with the lowest value of Chi square. Instead, if one wants to compare fittings made over different numbers of points, their corresponding quality factors have to be computed as shown in the following expression:

\[ Q(\chi^2|N - 2) = Q \left( \frac{N - 2}{2}, \frac{\chi^2}{2} \right), \]

(10)

where \( Q(\chi^2|\nu) \) is the probability that the computed value of Chi square for a fitting with \( \nu \) degrees of freedom will exceed the value \( \chi^2 \) by chance. If this probability is high, it means that it is very probable that any other fitting be worse than ours, so the optimum value of \( Q \) is 1. \( Q(a, x) \) is the complementary function of the incomplete gamma function but in many publications both functions are confused [5].

To compute the Chi square when coherence points have been fitted by means of the total least squares method, the measurement of the distance between the data and the fitted line, used in (9), must be replaced by the formula shown in (11), which measures the distance in the direction normal to the line [6]. The resulting expression can be decomposed into matrix products as shown in (11), where \( B \) is the matrix computed to perform the total least squares fitting already presented in (6). Once the Chi square has been computed, the quality factor \( Q \) can be calculated as shown in (10).

\[ \chi^2(r, w) = \sum_{j=1}^{N} \left( \frac{r^T(z_j - w)}{\sigma_j} \right)^2 = r^TB^TB, \quad \text{where} \quad z_j^T = (\Re(\gamma_j), \Im(\gamma_j)) \]

(11)

### 3.2 Regression index

The second quality estimator that we have employed is the regression index \( R \). This parameter is given by the following expression,

\[ R(l) = \frac{\sum_{j=1}^{N} l(\Re(\gamma_j)) - \bar{y})^2}{\sum_{j=1}^{N} (\Im(\gamma_j) - \bar{y})^2} \quad \text{where} \quad \bar{y} = \frac{1}{N} \left( \sum_{j=1}^{N} \Im(\gamma_j) \right) \]

(12)

where \( l \) represents the fitted line. By definition the value of \( R \) is between 0 and 1, and the optimum value is 1. To adapt the regression index to work with total least squares fitted lines, we have to use the notation shown in (13–15).

\[ z_j = (x_j, y_j) = (\Re(\gamma_j), \Im(\gamma_j)) \]

(13)

\[ \bar{z} = (\bar{x}, \bar{y}) = (\text{mean}(x_j), \text{mean}(y_j)) \]

(14)

\[ l(z_j) = (l_x(t_j), l_y(t_j)) \]

(15)

If the line fit is correct, the value of \( t_j \) can be obtained from (16–17). In order to reduce computing errors one must calculate \( t_j \) from the equation corresponding to the higher component of \( r^\perp \).

\[ x_j \simeq l_x(t_j) = w_x + t_j \cdot r_x^\perp \]

(16)

\[ y_j \simeq l_y(t_j) = w_y + t_j \cdot r_y^\perp \]

(17)

Finally, the new expression of the regression index for total least squares results in:

\[ R(l) = \frac{\sum_{j=1}^{N} (l_x(t_j) - \bar{x})^2 + (l_y(t_j) - \bar{y})^2}{\sum_{j=1}^{N} (x_j - \bar{x})^2 + (y_j - \bar{y})^2}. \]

(18)
Computer simulations have shown that $R$ provides information about how the width of the triangular region affects the performance of the fitting method, but the computed value also depends on the fitting method. The results obtained for the TLS method showed a low sensitivity to the triangle width, since the fitting was always correct (in absence of other error sources), and the measured error in the topographic phase estimation was always smaller than 5 degrees (considering a value of $k_z$ equal to 1.56 rad/m). The LS fitting method produced errors higher than 10 degrees for values of $R$ lower than 0.7. Note that the expression of $R$ shown in (12) fails when the fitted line is horizontal because the numerator of $R$ becomes zero.

4 APPLICATION TO REAL DATA

The fitting methods presented in previous sections and their quality estimators have been tested with some experiments carried out in the European Microwave Signature Laboratory (EMSL) at JRC-Ispira, Italy, with samples of maize and rice crops. The samples were observed at different frequencies, from 2 GHz to 9 GHz, and for incidence angles ranging from 44 to 45 degrees. For each incidence angle, the target was rotated in azimuth in order to obtain independent samples of its target vector and to perform the necessary average. From this set of data, coherence values were computed for two baselines, 0.25 and 0.5 degrees. For the 0.25 degrees baseline, 4 sets of coherence values were computed at 4 different central angles. These coherence values were averaged and its standard deviation computed as an estimation of $\sigma_j$, which is used in the weighted fittings. The inversion algorithms were tested with the averaged coherence values. For the 0.5 degrees baseline, only 3 sets of coherence values can be computed from the same set of data, so the averaged coherence values and its standard deviation are expected to be less reliable than the 0.25 degrees data. For this configuration at the EMSL, the vertical wavenumber for the 0.25 degrees baseline can be computed as $k_z = 0.26 \cdot f$ where $f$ is the frequency in GHz. For the 0.5 degrees baseline the resulting expression is $k_z = 0.52 \cdot f$.

4.1 Maize sample

The estimated topographic height and volume depth are shown in Figs. 1 and 2 as a function of the frequency. These parameters are obtained from different line fitting methods. The red traces correspond to vegetation heights estimated by the RVoG inversion algorithm assuming zero extinction, whereas the blue traces with circles correspond to the volume height estimated by the OVoG inversion algorithm, and their standard deviations are represented by the dashed blue lines. The estimated topographic heights are represented by solid black lines. The true values of topographic height and volume height are represented by dashed black lines.

In a first look to Fig.1 we can see that topographic height estimated by the least squares fitting method (LS) is very similar to the estimated with the total least squares fitting method (TLS). The difference between the two values are small when the coherence points are well aligned in the complex plane, provided that the resulting line is not vertically oriented. If the coherence points are not well aligned, the fitting procedure becomes harder and the estimated values result different. In that case, the TLS computes a better estimation than the LS. The topographic height estimates corresponding to the weighted versions of the same methods are only slightly more accurate, but noisier than the estimates of the not weighted procedures. This low improvement comes from the underestimation of the measurement error produced by the computed values of $\sigma_j$, as we anticipated in Section 2. An average of the target vectors is performed before the coherence values are computed, so the improvement obtained by weighting is small and shows bad statistical properties.

We also can observe a dip in the weighted estimates of topographic height, located between 8.5 and 9 GHz. This dip is caused because the bad statistical properties of the estimated values for $\sigma_j$ makes the weighted coherence points more difficult to fit, so both weighted fitting methods fail at that frequency. The descending peak in the trace of topographic height causes an ascending peak in the trace of the estimated volume height, because the topographic height has been underestimated and it has not been completely removed before computing the volume height. If this error is small enough, it can be masked by other error sources.

In a second careful look to both Figs. 1 and 2 we can observe that the estimated value of the vegetation height decreases as frequency increases, being more evident for the larger baseline. This trend is caused by the lower penetration capability of the radar signal as the frequency increases. As a result, the contribution from the lower particles in the volume to the backscatter decreases, and the visible depth of the volume is smaller. On the other hand, larger baselines are more sensitive to the vertical distribution of scattering centers and more suitable to measure heights accurately. Consequently, the estimates of topographic height are more accurate for the 0.5 degrees baseline than for the 0.25 degrees one, but the estimates of volume height are less reliable than the ones obtained by using the 0.25 degrees baseline.
The traces of the regression index, $R$, computed for both baselines, are shown in Fig. 3. As we expected, the higher values of $R$ are obtained with the total least squares fitting method, in both weighted and not weighted versions. The computed values of $Q$ depend strongly on the value of $\sigma_j$ used to compute the Chi square. The obtained values of $Q$ corresponding to the not weighted fitting methods, LS and TLS, are always very close to 1 because the fitting methods are equivalent to the weighted fitting methods using a value of $\sigma_j$ equal to 1 for all coherence points. This value of $\sigma_j$ overestimates enormously the measurement error so the computed value of $\frac{\chi^2}{2}$ is too small, and the obtained value of $Q$ is always optimist. On the other hand, the traces of $Q$ corresponding to the weighted fitting methods, WLS and WTLS, are noisy because the values of $\sigma_j$ used to compute the chi square underestimate the real measurement errors. The traces of $R$ and $Q$ for the 0.5 degrees baseline are very similar to the obtained for the 0.25 degrees one. The main difference corresponds to the traces of $Q$ for the weighted fitting methods, where the underestimation of the measurement errors causes very high values of $\frac{\chi^2}{2}$ and values of $Q$ which are mainly 0.

4.2 Rice sample

The measured sample of rice is very short and has a very low density, and, consequently, the backscatter of the above-ground volume is very weak. As a result, the ground-stem interaction is the dominant scattering mechanism due to the presence of the water (the ground is flooded). This mechanism prevails over the weak direct backscattering from the volume. Because of the strong ground-stem interaction, the coherence points are located in the complex plane forming a small cloud close to the unit circle. At low frequencies the coherence points are so close that the cloud seems to be a point and the line fitting has no sense. However, the computed values of $Q$ and $R$ are high because the fitted line is very close
to all coherence points. Hence, when the coherence points are too close, the quality factors fail.

The topographic height is always estimated accurately because of the proximity of the cloud to the unit circle, but the volume height only is correctly estimated when frequency increases. This behavior is a consequence of the larger vertical wavenumber, in conjunction with the stronger response of the canopy at these frequencies. The extreme cases are shown if Fig.4. The worst results are obtained when the 0.25 degrees baseline is used and the coherence points are fitted by the LS method, the best results correspond to the 0.5 degrees baseline and WTLS fitting method.

![Fig 4. Retrieved topographic height and volume height of the rice sample for a 0.25° baseline by using the LS fitting method (left) and for a 0.5° baseline by using the WTLS fitting method (right).](image)

5 CONCLUSIONS

The experimental results presented in this work have demonstrated that the line model is applicable to oriented volumes. The TLS fitting method, and its weighted version, have shown the best performance (higher values of $R$) among all the compared methods.

We have also seen that if we do not dispose of accurate estimations of the measurement errors, the value computed by the $Q$ function is highly biased and should not be used. Instead, the regression index adapted to the TLS fitting $R$ should be used because it only depends on the width of the coherence region and on the performance of the fitting method.

ACKNOWLEDGMENTS

This work has been supported by the Spanish Ministry of Science and Technology (MCYT) and FEDER, under Project TIC2002-04451-C02-02.

The authors would like to thank the personnel of the European Microwave Signature Laboratory (EMSL) for their support in performing the measurements and providing the experimental data.

REFERENCES


