

SLC SAR SPECKLE FILTERING USING HOMOSKEDASTIC FEATURES IN LOGARITHMIC TRANSFORMED DOMAIN

Thanh-Hai Le^a, Ian Vince McLoughlin^b, Ken-Yoong Lee^c and Timo Bretschneider^d

^a PhD Student, ^b Assoc Prof,
School of Computer Engineering (SCE),
Block N4, Nanyang Avenue, Nanyang Technological
University, Singapore.
leth0011@ntu.edu.sg, mcloughlin@ntu.edu.sg.

^c Scientist, ^d Senior Expert,
European Aeronautic Defence and Space (EADS)
Innovation Works,
41 Science Park Road, Science Park 2, Singapore.
ken-yoong.lee@eads.net, timo.bretschneider@eads.net.

KEY WORDS: Synthetic Aperture Radar, Speckle Filter, Heteroskedasticity.

ABSTRACT: This paper presents a novel speckle filter for Single Look Complex (SLC) Synthetic Aperture Radar (SAR) data. Whilst significant work has been performed in the area of speckle filtering, it is the underlying heteroskedasticity of the original SAR data that may pose the most intractable difficulties in designing such filters. These issues are explored through statistical analysis in the current paper, and extended toward a consideration of the comparative characteristics of log-transformed domain data. Subsequently, a novel speckle filter is derived that exploit homoskedastic effects, allied with a consistent sense of distance, contrast and variance. The proposed k-means based maximum likelihood filter is evaluated and validated both quantitatively over simulated scenarios and qualitatively over captured SAR data.

1. INTRODUCTION

The speckle phenomena in SAR imagery is explained as the interference of many coherent but de-phased backscattering components, each reflecting from different and distributed elementary scatterers (Leith, 1971). These interferences can be considered as a random walk on the 2D complex plane (Goodman, 1976). Assuming the Central Limit Theorem is applicable, the real part A_r as well as the imaginary part A_i of the observed SAR signal A , are considered as random variables from uncorrelated Gaussian distributed stochastic processes with zero mean and identical variance $\sigma^2/2$ (Lee, 2009). Their probability density function (PDF) is given as:

$$pdf(A_r) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{A_r^2}{\sigma^2}\right) \quad (1).$$

It then can be proven that the measurable amplitude and intensity are respectively random variables of Rayleigh and negative exponential distributed random processes as shown in Equations (2) and (3), respectively:

$$pdf(A) = \frac{2A}{\sigma^2} \exp\left(-\frac{A^2}{\sigma^2}\right) \quad (2),$$

$$pdf(I) = \frac{1}{\sigma^2} \exp\left(-\frac{I}{\sigma^2}\right) \quad (3).$$

If spatial homogeneity is defined as image regions having the same backscattering coefficient, then over such areas, the measured values can be considered as samples from a single stochastic process. Consequently, the expected mean and variance of the distributions are given in Table 1. Evidently, in the original domain, amplitude as well as intensity SAR data suffers from heteroskedastic phenomena, which is defined as the dependence of conditional expected variance on the conditional expectation of mean.

SAR speckle-filtering can be positioned within the context of estimation theory (Touzi, 2002). The statistical framework attempts to estimate unknown statistical parameters from observed data. This involves statistical modelling, estimator development and performance evaluation of estimated output. Statistical models are used to describe how the underlying backscattering coefficient (σ) affects the distribution of measured SAR data. Even though the statistics within individual resolution cells is well established, the applicability of its corresponding estimators is restricted to homogenous areas. Real-world images, however, are in-homogeneous. Nonetheless, it is this spatial variation that is of particular interest, thus raising the question of how to detect an analysis area as being heterogeneous, and correspondingly how to handle heterogeneity.

Several statistical models for heterogeneous areas have been proposed and a detailed review can found in Touzi, 2002. Unfortunately, while most highlight the multiplicative nature of original SAR data in homogenous regions, in extending the model to heterogeneous images, virtually none have noted that spatial variation also causes heteroskedasticity phenomena. As noted in our previous works (Le *et al.*, 2010), the condition of inconsistent

skedasticity has posed various challenges in all stages of designing a SAR speckle filter.

In this paper, we aim to neutralize these various impacts. The paper is organized as follows. Section 2 is devoted to the discussion of the logarithmic transformation and its homoskedastic effects, i.e. the consistent sense of distance, contrast and variance is demonstrated. Section 3 presents a novel speckle filtering algorithm, while Section 4 presents an evaluation of our estimator both quantitatively through simulated experiments and qualitatively on real-life images.

2. HOMOSKEDAST EFFECTS OF LOG-TRANSFORMED DOMAIN

2.1. Statistical Homoskedasticity of Logarithmic Transformation

We propose to consider the binary logarithmic transformation of SAR original random variables, i.e. $L_A = \log_2(A)$ and $L_I = \log_2(I)$. Bearing the relationship among the random variables in mind, it is then trivial to give the probability distribution of these log-transformed variables as:

$$pdf(L_A) = 2 \cdot \ln 2 \cdot \frac{2^{2L_A - 2\ln \sigma}}{\exp\left\{2^{2L_A - 2\ln \sigma}\right\}} \quad (4)$$

$$pdf(L_I) = \ln 2 \cdot \frac{2^{L_I - \ln \sigma}}{\exp\left\{2^{L_I - \ln \sigma}\right\}} \quad (5)$$

Noting that these distributions belong to the Fisher-Tippet family, the population expected mean and variances are given in Table 1, with γ being the Euler-Mascheroni constant. This result is consistent with findings by Arsenault, 1976. The main difference is the use of the binary logarithm which is preferred here for faster computation. Table 1 clearly confirms the condition of homoskedasticity, which is defined as the independence of conditional expected variance on the conditional expectation of mean.

RV	Variance (skedasticity)	Mean (bias)
A	Heteroskedastic $var(A) = \frac{11 - \pi^2}{6} \cdot \sigma^2$	Unbiased $avg(A) = \frac{\sqrt{\pi}}{2} \cdot \sigma$
I	Heteroskedastic $var(I) = \sigma^4$	Unbiased $avg(I) = \sigma^2$
L_A	Homoskedastic $var(L_A) = \frac{\pi^2}{24} \cdot \frac{1}{(\ln 2)^2}$	Biased $avg(L_A) = \frac{\gamma}{2} \cdot \frac{1}{\ln 2} - \log_2 \sigma$
L_I	Homoskedastic $var(L_I) = \frac{\pi^2}{6} \cdot \frac{1}{(\ln 2)^2}$	Biased $avg(L_I) = \gamma \cdot \frac{1}{\ln 2} + 2 \log_2 \sigma$

Table 1: SAR Statistical Models

2.2. Consistent Sense of Distance, Contrast and Variance

The consistent sense is illustrated in this section from two different perspectives. First, assume the backscattering coefficient (σ) is known *a priori*. Consider the random variable defined as the distance between an observable sample and its expected value, i.e. $D = L_I - \text{avg}(L_I)$. Noting the results from the previous section, the PDF for this variable is given as:

$$pdf(D) = \ln 2 \cdot \frac{2^{D + \ln \sigma}}{\exp\left\{2^{D + \ln \sigma}\right\}} \quad (6)$$

From the second perspective, given two adjacent resolution cells known to have the same but unknown backscattering coefficient σ , consider the random variable defined as the contrast between two measured samples, i.e. $C = L_1 - L_2$. Then the analytical PDF for this contrast variable is given in (7). The variance, in the case of two samples, is $V = C^2$ and the CDF can be calculated as in (8).

$$pdf(C) = \frac{2^{C-1}}{1 + 2^{C-1}} \ln 2 \quad (7)$$

$$cdf(V) = \frac{2^{V/2}}{1 + 2^{V/2}} - \frac{2^{-V/2}}{1 - 2^{-V/2}} \quad (8)$$

Evidently, the sense of distance, contrast and variance are all independent to the underlying backscattering coefficient σ in log-transformed domain. This can be extended to larger number of pixels via the analysis of observable sample variance. Such analysis can be carried out through Monte Carlo simulations, a plot of which is given in Fig. 1. The result here is confirmed experimentally by inspecting measured data from a known homogenous area. Fig. 1 plots the

histogram of observable data within a homogenous area imaged by RadarSat-2 against the modelled PDF responses. The model is validated by the noticeable agreements between analytical distribution plots and captured data histograms.

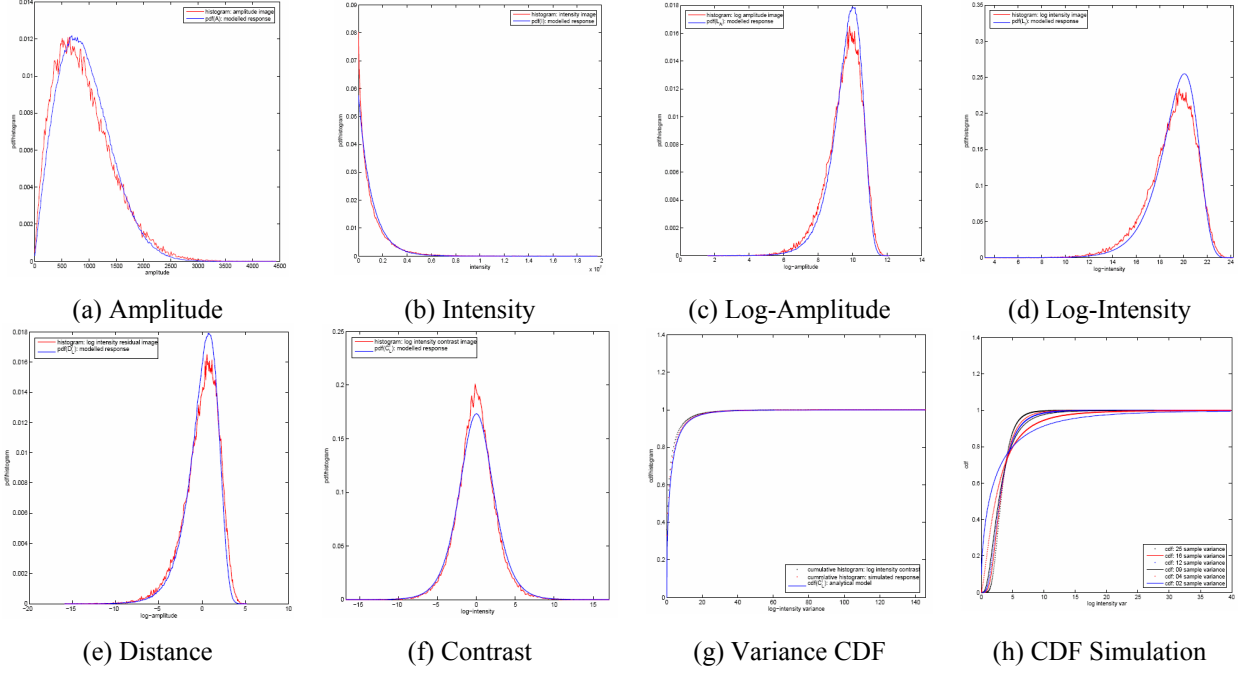


Figure 1: Validation of models against real-data

2.3. Sample Variance as Statistical Measure of Homogeneity

The previous analysis is based upon the assumption that a homogenous area arises from a single stochastic process, however, in real SAR images, the true condition is unknown. Still, the principle of SAR speckle filtering is to group pixels into homogenous areas, thus stochastic components can be removed. In order to do so, a methodology is required to assert a partitioned area as being homogenous. Fortunately, homoskedastic properties, manifested in the sampling's variance distribution, provide a consistent method for testing this homo-heterogeneity condition.

Under the null hypothesis that the given analysis area is homogenous, the theoretical distribution of sample variance is as given in the previous section. Using sample variance of log-transformed SAR data, we can put the null hypothesis to test. The statistical inference rule based upon a threshold value is that: if the observable sample variance is larger than pre-calculated threshold, the null hypothesis can be rejected. The threshold is calculated using Monte-Carlo simulations, a plot of which is given in Fig. 1.

3. Proposed SLC SAR Speckle filter using Homoskedastic Feature

Given an area of analysis, its non-homogeneity can be detected by the statistical hypothesis test described in Section 2.3. Upon such detection, a clustering algorithm is applied to partition the area into smaller and more homogenous regions. The testing-partitioning process can then be applied recursively until all partitions are considered to be homogenous (accepting the null hypothesis).

3.1. Homogenous Maximum Likelihood Estimation

In a window of analysis, an intensity dataset $X^I = \{x_1^I, x_2^I, \dots, x_m^I\}$ follows the exponential distribution given in Section 1. Assume the knowledge of homogeneity, i.e. a single σ , is available, the aim then is to estimate σ . It is already given that the derivative dataset $X^L = \log_2(X^I)$ follows a Fisher-Tippet distribution, in (5). Thus, the problem of estimating σ , in the log-transformed domain, is equal to the problem of estimating μ . Since log-transformed variables are biased, a Maximum Likelihood Estimation (MLE) approach is preferred:

$$\hat{x}^p = \frac{1}{n} \cdot \sum_{i=1}^n x_i^p \quad (9),$$

3.2. Heterogeneous k-MLE Clustering

In order to apply MLE in a given analysis window, a partitioning of the area into c groups of homogenous pixels (G) is needed. Assuming for now that the number of homogenous groups ($c = \text{size}(G)$) is known *a-priori*, a basic k-means clustering algorithm can be used – hereafter named k-MLE clustering.

1. Initialize the cluster centres
2. Assign each sample to a centre C_i that has the highest probability $P(x|C_i)$
3. Recalculate cluster centres for each C_i , based on Eqn. (9)
4. Iterate the above two steps until no more changes in cluster centres' locations occurs

In comparison to the standard k-means algorithm, the following differences are evident:

1. In k-means, data points are grouped into cluster with smallest distance. However, in the depicted situation, log-transformed data points are assigned to the cluster that gives them highest likelihood probability, as the dispersion PDF is non-symmetric around the cluster centre.
2. The calculation of new cluster centre is different as well. In k-means, new cluster centres are estimated as the average of all assigned data points. Here, due to the PDF bias, maximum likelihood estimations are preferred.

3.3. Determining the Number of Clusters

In the previous section, the algorithm design made the assumption that the number of clusters is known *a-priori*. In practice, this may not be true, requiring an adjustment to the algorithm. The basic idea is to start with a “single-cluster whole-area” homogenous assumption, which is then tested using the hypothesis testing procedure described above. If the assumption appears to be violated, the area is broken down by introducing a new partition. Clustering is carried out and the new clusters again are assumed to be homogenous. The process iterates until all partitions are confirmed to be homogenous. The pseudo code is given in Listing 1:

```
function centres = partition_area_into_homogenous_groups(l, threshold_vector)
    c_init = maximum_likelihood_of(l);
    %check if all current partitioning homogenous? If not, further partition is needed
    [further_partition_needed, c_new] = check_homogeneity_status(l, c_init, threshold_vector);

    while (further_partition_needed) do
        c_prev = c_new;
        c_new = k_mle_cluster(l, c_prev);
        [further_partition_needed, c_new] = check_homogeneity_status(l, c_prev, threshold_vector);
    end while
    %after this: all partitions are homogenous

    return centres = c_new;
end function
```

Listing 1: k-MLE pseudo code

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. k-MLE Filter Discussion

The decision rule for the statistical test is that if the sample variance is lower than the threshold, the null hypothesis is accepted. It should be noted that there is a difference between accepting the null hypothesis and simply failing to reject it. However, the following analysis shows that for the purpose of speckle filtering the difference is negligible. Consider some extreme threshold values. In the case of a very high value (towards 100%), the given analyzing area would almost never be broken down, and, hence, the k-MLE effect is equivalent to a box-car filter. Conversely, for a very low value (towards 0%), the given area would almost always be broken down to single point values and the k-MLE filtering effect would then be negligible. Figure 2 shows various degrees of speckle filtering controllable via

appropriate setting of the threshold value.

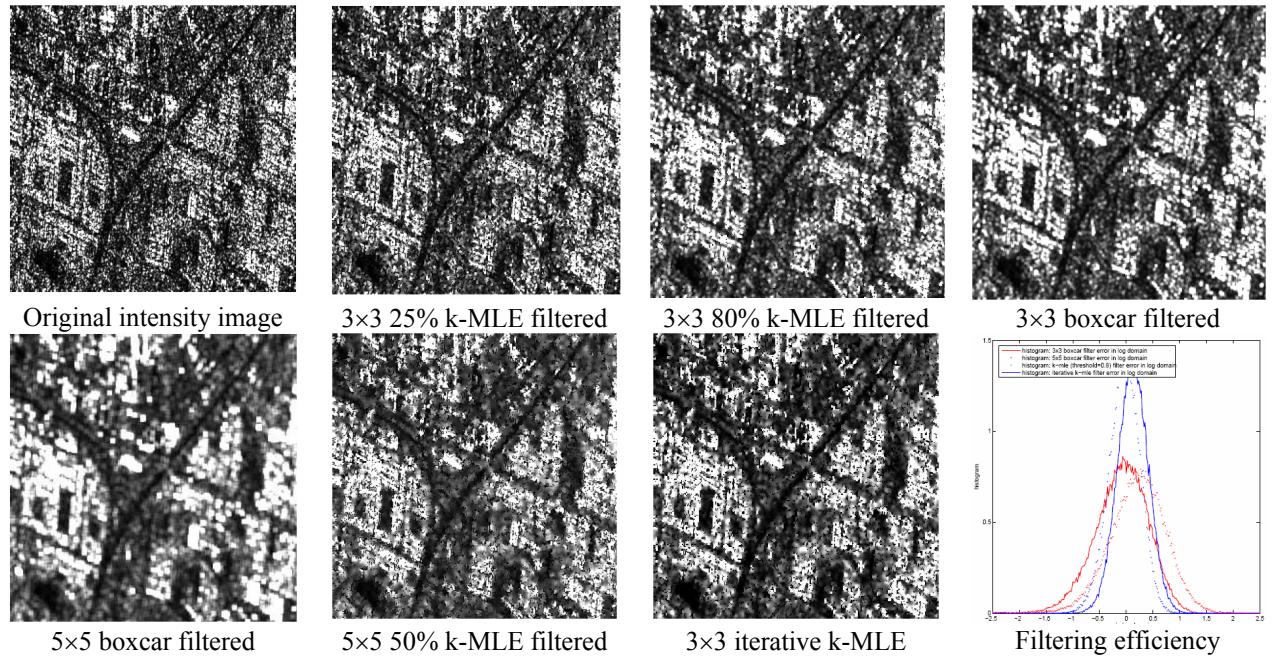


Figure 2: Effects of k-MLE threshold-values

4.2. Improving and Quantitatively Evaluating Speckle Filtering Efficiency on Simulated Homogenous Areas

As can be seen in Figure 2, a normal 3×3 (80%) k-MLE filter does not filter as much noise as a normal 3×3 boxcar filter does. One way to improve the speckle filtering effect is by working on a larger window of analysis. In this case, when compared to the 5×5 boxcar filter, the k-MLE filter with a threshold-value below 100% would result in adaptive partitioning and is characterized by its better preservation of edges. Figure 3 shows the effectiveness of k-MLE filtering using 5×5 windows. Obviously, a larger window choice will have better speckle filtering power at the cost of increased blurring. One way to improve the filtering efficiency of the k-MLE filter, while avoiding the blurring artifacts, is to use the k-MLE filter iteratively.

To evaluate the degree of speckle removed, a 256×256 homogenous area was generated and SAR speckle was simulated with the given PDF (3). Filters were applied to the patch and error is considered as the difference between filtered output and the “ground-truth” input values in the log-domain. Figure 2 presents a plot of this error; not only as averaged values, but the whole error distribution of different filters is reported. From the plot, the 3×3 iterative k-MLE filter appears to have similar quantitative performance with the 5×5 boxcar filter, while spatial feature preservation appears to be better.

4.3. Qualitative Evaluation on Real Images

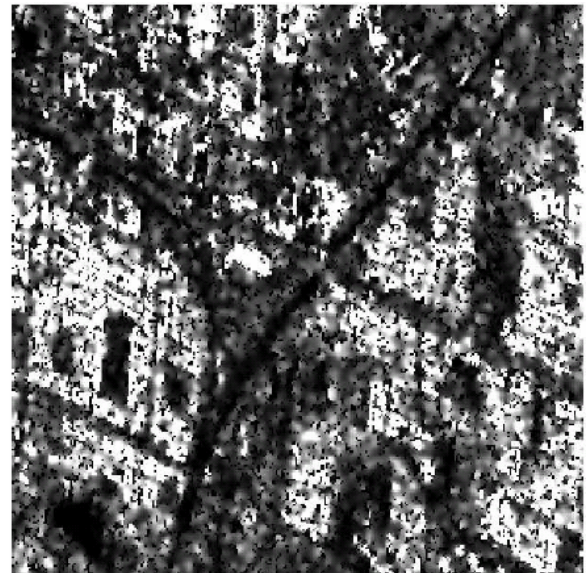
The filters are applied to an SLC RadarSat-2 image covering the Muada Merbok area of Malaysia. Figure 3 shows patches of original and filtered images for areas of both urban and natural landscapes. The filter’s effect on speckle suppression is visually observable.

5. Conclusion

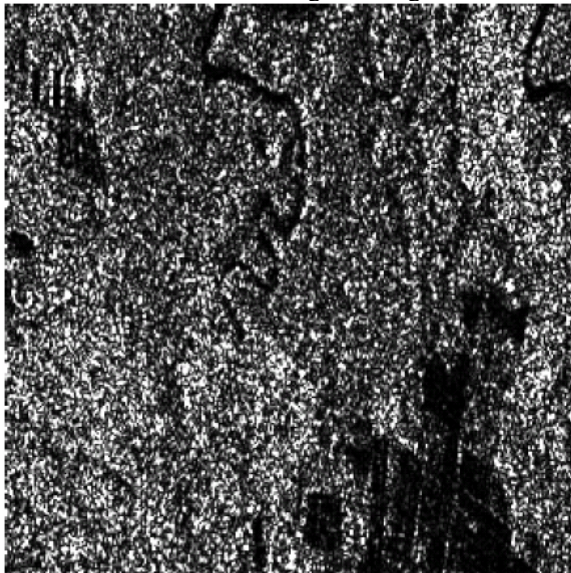
The nature of SAR speckle is stochastic and SAR data suffer from heteroskedastic phenomena. In this paper, logarithmic transformation has been shown to provide homoskedastic properties. Thus, a novel speckle filter was introduced exploit the consistent sense of variance and distance within the log-transformed domain. Image and statistical evidence has been presented which demonstrate the effectiveness of the method.



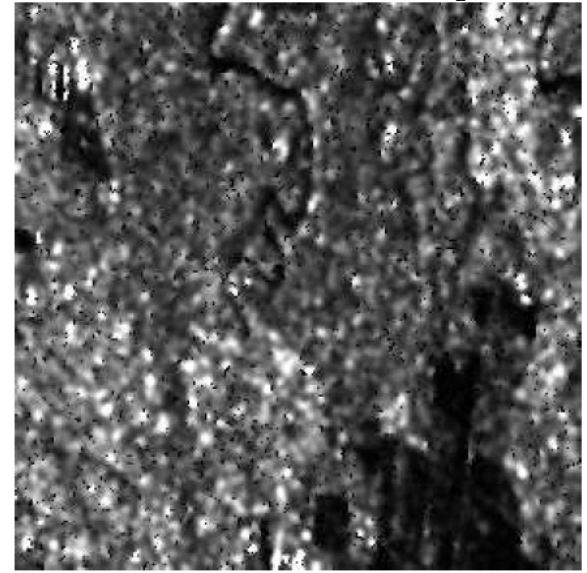
Urban area: original image



Urban area: k-MLE filtered image



Natural surfaces: original image



Natural Surfaces: k-MLE filtered image

Figure 2: Qualitative Evaluation of k-MLE filter

References

- Arsenault, H. H. & April, G., 1976. *Properties of speckle integrated with a finite aperture and logarithmically transformed*. J. Opt. Soc. Am., 66, pp.1160-1163
- Goodman, J. W., 1976. *Some fundamental properties of speckle*. J. Opt. Soc. Am., 66, pp.1145-1150.
- Le, T. H.; McLoughlin, I. V.; Lee, K. Y. & Bretschneider, T., 2010. *Neural Network-Assisted Reconstruction of Full Polarimetric SAR Information*. Proceedings of the International Symposium on Communications, Control and Signal Processing, Cyprus, Mar. 2010.
- Lee, J.S., 2009. *Polarimetric Radar Imaging: from Basics to Applications*. CRC Press.
- Leith, E., 1971. *Quasi-holographic techniques in the microwave region*. Proceedings of the IEEE, 59, pp. 1305–1318.
- Touzi, R., 2002. *A review of speckle filtering in the context of estimation theory*. IEEE Transactions on Geoscience and Remote Sensing, 40, pp.2392 - 2404.