# Estimation of Soil Moisture Content from L- and P-band AirSAR Data: A Case Study in Jeju, Korea

E. Y. Kwon[1], S. E. Park[1], K. K. Lee[1], and W. M. Moon[1,2]

1. School of Earth and Environmental Sciences, Seoul National University, Seoul 151-742 KOREA
2. Geophysics, The University of Manitoba, Winnipeg, Manitoba R3T 2N2 CANADA

(vanares@kebi.com, wmoon@eos1.snu.ac.kr)

***Abstract*- During the PacRimII AirSAR campaign in Korea, the ground truth data about soil moisture content and surface roughness characteristics were collected. We intend to retrieve the surface parameters over the bare soil from multi-polarization and multi-frequency AirSAR data. In this study, the theoretical scattering model, the IEM model is inverted by two existing algorithms – the multi-dimensional regression technique by Dawson et al. [1] and the inversion using 3-layer artificial neural networks (ANNs) [4]. As the first step, backscatter coefficients are calculated based on the ground truth information, and then training patterns are generated from within the valid ranges of surface parameters using the IEM model. The trained inversion models are tested to a set of AirSAR data as well as synthetic data. Root mean square (RMS) errors of estimated soil moisture from the AirSAR data are average 3.1% in the regression and 4.2% in the inversion using the ANNs. The methods to improve the inversion accuracy are investigated. First, the normalization of signal parameters reduced the number of pixels that fail to reasonable results in the regression model. Second, the use of co-polarization ratio as input units in the ANNs inversion scheme improve the soil moisture estimation, which results in an average RMS error of 2.9%.**

## I. INTRODUCTION AND TEORETICAL BACKGROUND

Soil moisture content is an essential parameter in agriculture and geo-hydrology and it has been a target in many academic research projects. Applications of active synthetic aperture radar (SAR) data have been investigated by many scientists for its potential to monitor soil moisture over large area regardless of the weather and the presence of the sun. For example, Oh et al. [7] have developed empirical relations between backscatter coefficients and surface soil moisture content. And Dawson et al. [1] have examined a multidimensional statistical estimation method based on theoretical scattering model and applied to experimental data. Furthermore, Hoeben et al. [5] used active microwave observations of the surface soil moisture content to estimate the root zone soil moisture profile by data assimilation.

During the PacRim-II AirSAR campaign in Korea, the polarimetric SAR experiment was accompanied by detailed ground truth measurements, which include soil moisture contents, surface roughness characteristics, and surface cover mapping. The surface roughness characteristics investigated include root mean square (RMS) height and correlation length in bare surface test sites. The goal of our experiment is to correlate the analysed surface information with retrieved information from fully polarimetric AirSAR data in L- and P-band.

In this paper, soil moisture content and surface roughness parameters are estimated from the AirSAR data observed at HH and VV polarization at L- and P-band. In this section, the theoretical background is illustrated. And this is applied to the AirSAR data over test sites in section II and III.

### *A. Scattering Model as a Forward Model*

The backscatter coefficient is defined as a ratio of received signal intensity to transmitted signal intensity. The returned signal after being scattered against land surfaces depends on the wave parameters and surface parameters. The wave parameters consist of frequency, polarization state, and an incidence angle. Although this information is included in polarimetric SAR data set, the surface parameters are unknown variables, which are implicit in the backscatter coefficients. In case of bare surface, dielectric properties and surface roughness characteristics such as RMS height and correlation length belong to the surface parameters. The relationships between backscatter coefficients and surface parameters have been developed. For example, theoretical models such as small perturbation model, physical optics model, and geometrical optics model and empirical models by Oh et al. [7] and Dubois et al. [3] have been developed and applied. Among them, the physically driven scattering model, the integral equation method (IEM) model is known to have a relatively large domain of validity, whereas the other theoretical models are limited by the severe assumptions and empirical models have over-restrictions [10]. Therefore, the IEM model is selected as a forward model. The IEM model used in this study is a simplified version through approximations so as to simulate single scattering from randomly rough surfaces. A criteria is

$$(ks)\cdot(kl) < 1.2\sqrt{\varepsilon_r} \qquad (1)$$

where $k$ is the wave number, $s$ is the surface RMS height, $l$ is the surface correlation length, and $\varepsilon_r$ is a relative dielectric constant of the soil. For further knowledge about the IEM model and the model equations, refer to the literature by Fung [4].

*B. Conversion of Soil Moisture Content to Dielectric Constant*

The dielectric constant of a surface is one of variables in the IEM model. So, the values of volumetric moisture content need to be converted to dielectric constant. According to the semi-empirical model of Peplinski et al. [8], which is valid in the frequency range between 0.3 and 1.3 GHz, the dielectric constant is a function of soil texture, wave frequency, soil temperature, and soil water content. The principle and the procedure for conversion of soil moisture content to dielectric constant are explained in Dobson et al. [2] and Peplinski et al. [8]. Figure 1 illustrates the changes of dielectric constant depending on soil water content variations at C-, L-, and P-band in the case of the soil texture and temperature over our test sites.

*C. Methodologies for Inversion*

In the applications of remotely sensed data, it is important to solve inverse problem as well as to construct the forward relationships. Once a rigorous forward model, IEM model is established, inversion method determines the accuracy of retrieved surface parameters. The goal of the inversion is to extract surface RMS height $s$, surface correlation length $l$, and volumetric soil moisture content $m_v$ from four backscatter coefficients and an incidence angle, denoted by $s^{L}_{HH}$, $s^{L}_{VV}$, $s^{P}_{HH}$, $s^{P}_{VV}$, and $j$. Two methodologies are reviewed in this subsection. One is the multi-dimensional regression technique developed by Dawson et al. [1] and the other is the inversion using artificial neural networks (ANNs) [4].

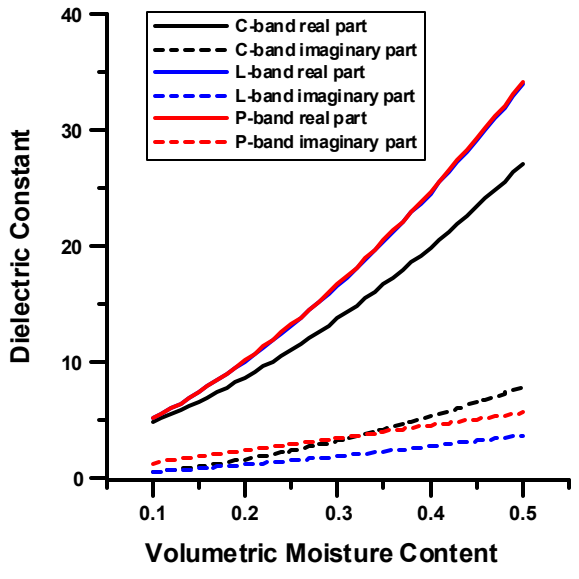


Fig. 1. The curves of dielectric constant depending on volumetric soil moisture content variations at C-, L-, and P-band according to the semi-empirical model of Dobson et al. and Peplinski et al.

*(a) Review of the Multi-Dimensional Regression Technique*

For the regression model, a $q$ th input vector is composed of four backscatter coefficients and an incidence angle as $\vec{Z}_q = \{z_q(1), z_q(2), z_q(3), z_q(4), z_q(5)\} = \{\sigma^{L}_{HH}, \sigma^{L}_{VV}, \sigma^{P}_{HH}, \sigma^{P}_{VV}, \varphi\}_q$. And basis functions are organized as second order polynomial functions. Given the dimension 5 of the input vector $\vec{Z}_q$ and the highest polynomial degree 2, total 21 basis functions compose a set of basis function as $\vec{\Phi}_q = \{\phi_q(1), \cdots, \phi_q(21)\} = \{1, z_q(1), \cdots, z_q(1)^2, z_q(1)\cdot z_q(2), \cdots, z_q(5)^2\}$. $j$ th output $\tilde{q}(j, \vec{Z}_q)$ is expressed as the summation for all 21 members of the basis function which is evaluated for the $q$ th input vector $\vec{Z}_q$ as

$$\tilde{q}(j, \vec{Z}_q) = \sum_{i=1}^{21} w(j,i) \cdot f_q(i) \tag{2}$$

where $w(j, i)$ is the weight connecting the $i$ th basis function $\phi_q(i)$ to $j$ th output $\tilde{q}(j, \vec{Z}_q)$. The desired output vector of $q$ th pattern is $\vec{\theta}_q = \{\theta_q(1), \theta_q(2), \theta_q(3)\} = \{s, l, m_v\}_q$. Assuming that total $N_v$ training patterns are given, the error for $j$ th output is

$$E_j = \sum_{q=1}^{N_v} [\theta_q(j) - \tilde{\theta}_q(j)]^2 = \sum_{q=1}^{N_v} [\theta_q(j) - \sum_{i=1}^{21} w(j,i) \cdot \phi_q(i)]^2 . \tag{3}$$

The total mean square error for the system is

$$MSE = \frac{1}{N_v} \sum_{j=1}^{3} E_j = \frac{1}{N_v} \sum_{j=1}^{3} \sum_{q=1}^{N_v} [\theta_q(j) - \sum_{i=1}^{21} w(j,i) \cdot \phi_q(i)]^2 . \tag{4}$$

The training patterns of $N_v$ = 4,000 are used to compute the weights in this study. The best set of coefficient $w(j, i)$ can be determined so that the mapping error of (3) is minimized. This is done by taking the partial derivative of the error with respect to each unknown coefficient and by equating the result to zero as follows.

$$\frac{\partial E_j}{\partial w(j,i)} = \sum_{q=1}^{N_v} 2\{\theta_q(j) - \sum_{i=1}^{21} w(j,i) \cdot \phi_q(i)\} \cdot \{-\phi_q(i)\}$$
$$= -2 \sum_{q=1}^{N_v} \{\theta_q(j) - \sum_{i=1}^{21} w(j,i) \cdot \phi_q(i)\} \cdot \phi_q(i) = 0 \tag{5}$$

from the above relations, the next results are

$$\therefore \sum_{q=1}^{N_v} \theta_q(j) \cdot \phi_q(i) = \sum_{q=1}^{N_v} \phi_q(i) \cdot \sum_{i=1}^{21} w(j,i) \cdot \phi_q(i) \tag{6}$$

Matrix form of equation (6) is as follows. 3 linear matrixes of the size $21 \times 21$ can be derived as

$$\begin{pmatrix} \sum_{q=1}^{N_v}\theta_q(j)\phi_q(1) \\ \sum_{q=1}^{N_v}\theta_q(j)\phi_q(2) \\ \sum_{q=1}^{N_v}\theta_q(j)\phi_q(3) \\ \vdots \\ \sum_{q=1}^{N_v}\theta_q(j)\phi_q(21) \end{pmatrix} = \begin{pmatrix} \sum_{q=1}^{N_v}\phi_q(1)\phi_q(1) & \sum_{q=1}^{N_v}\phi_q(1)\phi_q(2) & \cdots & \sum_{q=1}^{N_v}\phi_q(1)\phi_q(21) \\ \sum_{q=1}^{N_v}\phi_q(2)\phi_q(1) & \sum_{p=1}^{N_v}\phi_q(2)\phi_q(2) & \cdots & \sum_{q=1}^{N_v}\phi_q(2)\phi_q(21) \\ \sum_{q=1}^{N_v}\phi_q(3)\phi_q(1) & \sum_{q=1}^{N_v}\phi_q(3)\phi_q(2) & \cdots & \sum_{q=1}^{N_v}\phi_q(3)\phi_q(21) \\ \vdots & \vdots & \cdots & \vdots \\ \sum_{q=1}^{N_v}\phi_q(21)\phi_q(1) & \sum_{q=1}^{N_v}\phi_q(21)\phi_q(2) & \cdots & \sum_{q=1}^{N_v}\phi_q(21)\phi_q(21) \end{pmatrix} \begin{pmatrix} w(j,1) \\ w(j,2) \\ w(j,3) \\ \vdots \\ w(j,21) \end{pmatrix} \quad \textbf{(7)}$$

where $N_v$ = 4,000 and $j$=1, 2, or 3 in this study. The results of the inversion are described in section **III**.

*(b) Review of the Inversion Using ANNs*

Inversion is also attempted using ANNs, which is suitable for multidimensional retrieval from nonlinear scattering model. A 3-layer perceptron networks with one hidden layer is chosen. Each unit has a nonlinear sigmoid activation function. An input vector into the first layer contains five input signals, which are four backscatter coefficients and an incidence angle as $\{u^1(1), u^1(2), u^1(3), u^1(4), u^1(5)\} = \{\sigma^L_{HH}, \sigma^L_{VV}, \sigma^P_{HH}, \sigma^P_{VV}, \varphi\}$. In the second or third layer, the input signal $u^2(i)$ or $u^3(i)$ is the weighted linear summation of the outputs from the previous layer as

$$u^2(i) = \sum_{j=1}^{5} w^2(i,j) \cdot x^1(j) \tag{8}$$

in the second layer, and

$$u^3(i) = \sum_{j=1}^{45} w^3(i,j) \cdot x^2(j) \tag{9}$$

in the third layer because there are 45 units in a hidden layer. These are activated via the nonlinear sigmoide activation function, in the case of the second layer,

$$x^2(i) = f(u^2(i)) = \frac{1}{1+\exp(-u^2(i))}. \tag{10}$$

3 output signals $s$, $l$, and $m_v$ come out from the networks after being scaled linearly. The fully connected networks is trained by backpropagation learning algorithm [9], which learning in the network is the process to minimize the sum of the squared errors between the desired outputs called teacher signal and the computed network outputs as

$$E = \frac{1}{2}\sum_{i=1}^{3}\left(x^3(i) - d(i)\right)^2. \tag{11}$$

While a training pattern is fed into the networks sequentially, the weights are updated as much as the difference of

$$\Delta w^m = -\mu \frac{\partial E}{\partial w^m(i,j)} \tag{12}$$

where $\mu$ is the learning rate. Specifically,

$$\begin{aligned} \Delta w^3(i,j) &= -\mu \frac{\partial E}{\partial w^3(i,j)} = -\mu \frac{\partial E}{\partial x^3(i)} \cdot \frac{\partial x^3(i)}{\partial u^3(i)} \cdot \frac{\partial u^3(i)}{\partial w^3(i,j)} \\ &= -\mu \cdot \left(x^3(i) - d(i)\right) \cdot x^3(i) \cdot \left(1 - x^3(i)\right) \cdot x^2(j) \end{aligned} \tag{13}$$

and

$$\begin{aligned} \Delta w^2(i,j) &= -\mu \frac{\partial E}{\partial w^2(i,j)} = -\mu \frac{\partial E}{\partial x^2(i)} \cdot \frac{\partial x^2(i)}{\partial u^2(i)} \cdot \frac{\partial u^2(i)}{\partial w^2(i,j)} \\ &= -\mu \cdot \frac{\partial E}{\partial x^2(i)} \cdot x^2(i) \cdot \left(1 - x^2(i)\right) \cdot x^1(j) \end{aligned} \tag{14}$$

where

$$\begin{aligned} \frac{\partial E}{\partial x^2(i)} &= \sum_{k=1}^{3} \frac{\partial E}{\partial x^3(k)} \cdot \frac{\partial x^3(k)}{\partial u^3(k)} \cdot \frac{\partial u^3(k)}{\partial x^2(k)} \\ &= \sum_{k=1}^{3} \left(x^3(k) - d(k)\right) \cdot x^3(k) \cdot \left(1 - x^3(k)\right) \cdot w^3(k,i). \end{aligned} \tag{15}$$

In this study, total 4,000 training patterns are generated from the IEM model and the ANNs are trained by total 40,000 iterations, ten times of the training patterns. One example of the training errors during iterations is shown in Fig. 2 and the results of the ANNs inversion are presented and discussed in section **III**.

*D. Spatial Analysis of the Distributed Data*

The statistics considering not only the quantities but also the sampling locations can be achieved by variogram modeling [12]. The variogram measures the degree of similarity of two samples taken the lag distance apart. When

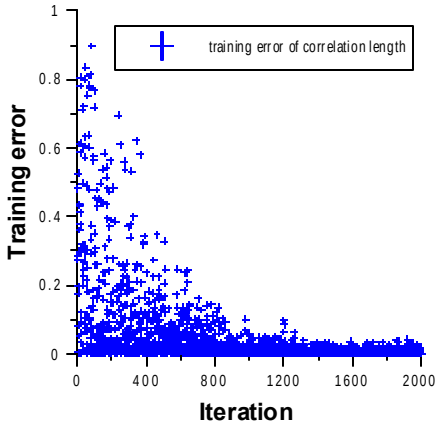


Fig. 2. Scatter plot of training errors during the backpropagation learning (One scatter point represents an error per one training).

$v(x_i)$ is an attribute of a pixel at a point $x_i$ and $N_s$ is the number of pairs of data points located by the lag distance $h \pm \Delta h$, the definition of semi-variogram is.

$$\gamma(h \pm \Delta h) = \frac{1}{2N_s}\sum_{i=1}^{N_s}[v(x_i) - v(x_i + h \pm \Delta h)]^2 . \quad (16)$$

The semi-variograms are estimated for the AirSAR signal parameters and soil moisture content measured on ground in section **III**.

## II. DESCRIPTIONS OF DATA

### *A. Test Sites*

Experimental sites are located at Jeju-si on Jeju island in Korea. Since the Jeju island originated from the eruption of basaltic volcano, the soil covering the island is characterized by high porosity which ranges from 48.7% to 60.2% and high permeability. The Soil is composed of sand of 15.9% and clay of 28.2%. And the volumetric density and the density of soil particle measure average 1.09 $g/cm^3$ and 2.47 $g/cm^3$, respectively. On September 30, the average temperature was 23℃. And it was cloudy and rained intermittently. However the surface soil was not saturated and runoff didn't occur by rapid infiltration. Two flat bare surfaces named as site L1 (Fig. 3) and site L2 are selected for the estimation of soil moisture content from AirSAR data.

Data observed by P-band (0.45GHz frequency, 67cm wavelength) and L-band (1.26GHz, 23cm) at HH (Horizontally received and Horizontally transmitted), HV (Horizontally received and Vertically transmitted), and VV (Vertically received and Vertically transmitted) polarizations were collected. Azimuth pixel spacing and range pixel spacing of the AirSAR data are 4.63m and 3.3m, respectively. Fig. 4 shows a set of AirSAR images around the L1 and L2 sites.

Fig. 3. PacRim-II AirSAR experimental site L1 on Jeju island in Korea.

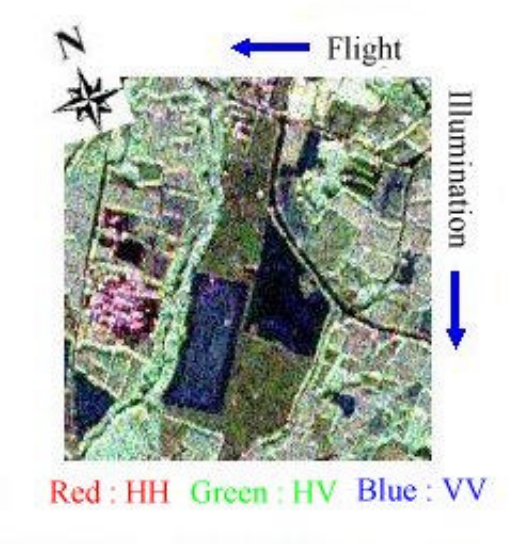


Fig. 4. A set of AirSAR images around the test sites (the upper right trapezoid of blue color depicts site L1 and the lower left rectangle of blue color depicts site L2).

### *B. Roughness Characteristics Measured on Ground*

The profiles of surface roughness were obtained each location by spraying over a scale paper attached to a tin plate fixed vertically on ground. The profiles were obtained at right angles when there is a linear directional trend. Two profiles at each site are measured; the east-west direction (EW) and the north-south direction (NS) in both L1 and L2 sites. And then, two roughness statistics were estimated from the digitisation of surface heights. One is the standard deviation of the surface height and the other is the surface correlation length, which is estimated from the Gaussian autocorrelation function fitted in Fig. 5. The Gaussian autocorrelation function is defined as

$$\rho(x) = \exp\left(-\frac{|x|^2}{l^2}\right) \quad (17)$$

where $x$ is the displacement and $l$ is the correlation length such that $\rho(l) = \frac{1}{e}$. Equation (17) can be converted to the roughness spectrum via Fourier transformation as

$$W^{(1)}(2k\sin\theta) \approx \frac{l^2}{2}\cdot\exp\left(-\frac{(kl\sin\theta)^2}{}\right) \quad (18)$$

where $k$ is the wave number and $\theta$ is an incidence angle. Equation (18) is inputted to the IEM model directly as a roughness parameter. Table 1 shows roughness statistics and normalized parameters with respect to the AirSAR wave numbers information.

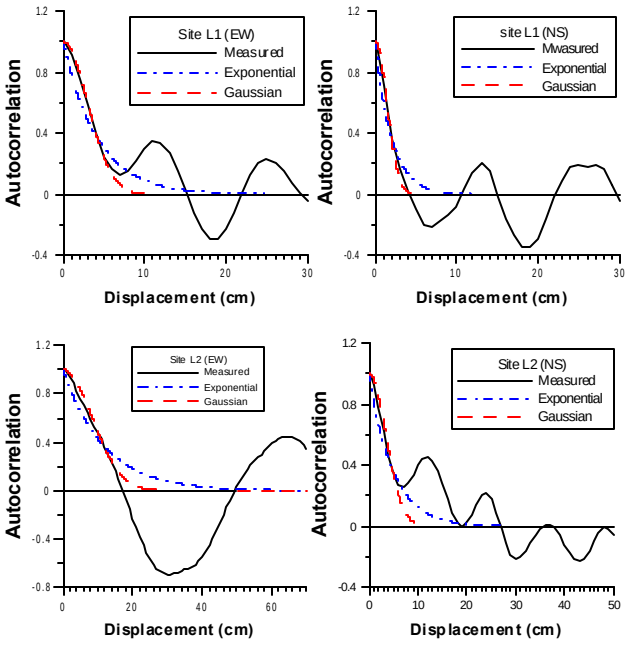


Fig. 5. Measured and fitted surface autocorrelation functions on site L1 and L2.

TABLE 1
MEASURED ROUGHNESS PARAMETERS IN EXPERIMENTAL SITES

| Name of Sites | Direction | RMS Height (cm) | Correlation Length (cm) | *ks* (L) | *Ks* (P) | *kl* (L) | *kl* (P) |
|---|---|---|---|---|---|---|---|
| L1 | EW | 1.0 | 4.2 | 0.27 | 0.094 | 1.1 | 0.39 |
| | NS | 0.49 | 2.0 | 0.13 | 0.046 | 0.55 | 0.19 |
| L2 | EW | 1.3 | 11.8 | 0.35 | 0.12 | 3.22 | 1.11 |
| | NS | 0.94 | 4.8 | 0.26 | 0.088 | 1.3 | 0.45 |

*s* is a surface root mean square height in cm, *l* is a surface autocorrelation length in cm, and *k* is the wave number, which result in 0.27 for L-band and 0.094 for P-band.

### *B. Soil Moisture Content Measured on Ground*

Volumetric soil moisture content was measured at intervals of 30m with the use of Time Domain Reflectometry (TDR) Probe. The probe measures the water content of 0~12cm depth from land surfaces. The averages and standard deviations of soil moisture content are 37.3% and 3.7% in L1 site, respectively and 33.3% and 4.0% in L2 site, respectively. The frequency distributions of measured soil moisture content are shown in Fig. 6. The relatively wide ranges of the soil water content seem to be owing to raindrop around the test date.

Particularly inside a selected area, total 60 soil samples were collected with the regular interval of 4m in EW direction and the interval of 7m in NS direction. The gravimetric moisture contents were derived by comparing the weights of soil contained in a controlled volume after oven drying at 105° C with those before oven drying. This detail measurements are used for the spatial analysis of the soil moisture distribution. This is compared with the spatial

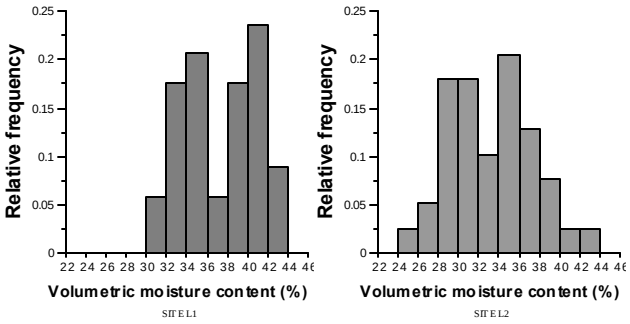


Fig. 6. Frequency distributions of the soil moisture content measured at site L1 and L2.

distribution of the AirSAR signal parameters in section **III**.

## III. RESULTS AND DISCUSSION

### *A. Estimation of Backscatter Coefficients*

The backscatter coefficients by single scattering from randomly rough surfaces are calculated using the IEM model. The backscatters at L- and P-band at HH and VV polarizations are simulated with *s*, *l*, and $m_v$ within the limits measured on ground. And then its mean and standard deviation are compared with those of the backscatter coefficients in the AirSAR data, of which the pixels except the bare soil area are masked. This statistics of calculated and observed backscatter coefficients are tabulated in Table 2. The comparison shows the largest deviation at VV polarization of L-band, whereas the smallest deviation at HH of L-band. The biases influence the performance of inversions, because the inversion models trained with the synthetic data set are applied to the observed data set. And so, both synthetic and observed backscatter coefficients are normalized while training the inversion models and estimating the surface parameters with the use of the inversion models. The normalization of signal parameters reduced the number of pixels that fail to reasonable results in the regression model.

As a preliminary step for inversion, sensitivities are examined. The scopes of surface parameters and the corresponding changes of backscatter coefficients are based on the ground truth information (Fig. 7). And the scope of an incidence angle in Fig. 7 is based on the AirSAR data taken

TABLE 2
STATISTICS OF BACKSCATTER COEFFICIENTS IN dB CALCULATED FROM THE IEM MODEL AND OBSERVED BY AIRSAR

| | | $\sigma^L_{HH}$ | $\sigma^L_{VV}$ | $\sigma^P_{HH}$ | $\sigma^P_{VV}$ |
|---|---|---|---|---|---|
| Calculation by IEM | Mean | -21.93 | -13.06 | -28.76 | -20.90 |
| | SD | 6.036 | 5.587 | 3.947 | 4.069 |
| AIRSAR data (L1 and L2) | Mean | -22.09 | -19.02 | -25.80 | -24.79 |
| | SD | 3.846 | 3.607 | 4.349 | 4.136 |

SD means the standard deviation. The region except for L1 and L2 sites is masked in the AirSAR data.

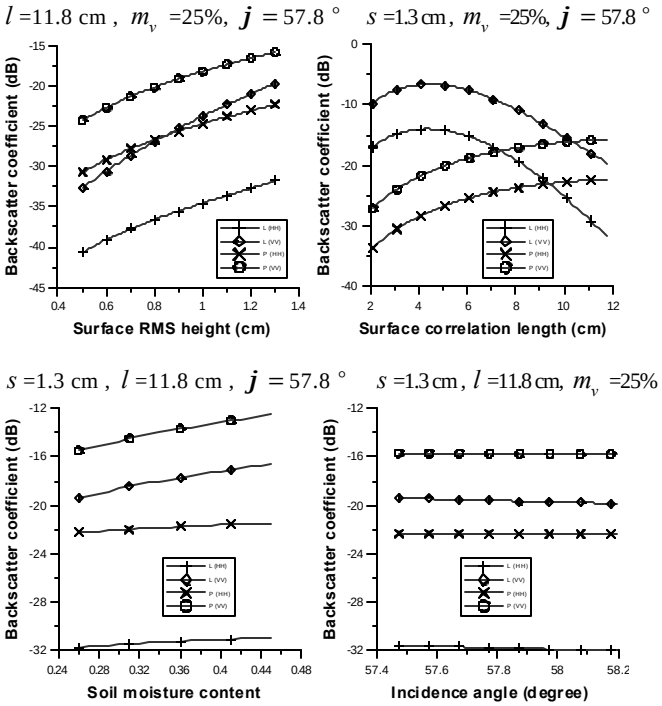


Fig. 7. Change of the backscatter coefficients responding surface parameters and an incidence angle.

in test sites. The backscatter coefficients are less sensitive to soil moisture content than to surface roughness parameters. This makes the estimation of soil water content difficult in two inversion models, which is discussed in the next subsection.

### *B. Estimation of Surface Parameters*

4000 data sets are generated synthetically from the IEM model for the purpose of training two inversion models. Each set contains the three surface parameters $s$, $l$, and $m_v$, four backscatter coefficients $s^{L}_{HH}$, $s^{L}_{VV}$, $s^{P}_{HH}$, and $s^{P}_{VV}$, and an incidence angle $\boldsymbol{j}$. Table 3 shows the lower and upper limits of the measured and synthetic surface parameters. The limits of synthetic data follow the criteria of equation (1), the condition for the application to the IEM model, in addition to cover the limits of measured surface parameters. The synthetic data sets are used to train the multi-dimensional regression model and the ANNs inversion model as mentioned in section **I**. The process for a training error of the surface correlation length to converge has been shown already in Fig. 2. That shows the best convergence. Namely, average RMS error during last 100 iterations is 0.080cm for $l$, whereas total error of equation (11) is 0.11. This observation is consistent with the best sensitivity of $l$ in the previous subsection. After the training, the synthetic backscatter coefficients and incidence angle are re-inputted to the trained inversion models. The estimated outputs versus desired surface parameters are shown in Fig. 8. This cross-validation confirms the propriety of the inversion models. The RMS errors of the estimated outputs are summarized in Table 4. Poorer results in the ANNs inversion model than in the regression model requires further investigation about the method.

Two inversion models are also applied to the AirSAR data. The estimated surface parameters are shown in Table 6. The average RMS errors of soil moisture content are 3.1% in the regression model and 4.2% in the ANNs inversion model.

TABLE 3
LOWER AND UPPER LIMITS OF SURFACE PARAMETERS

| | $s$ (cm) | $l$ (cm) | $m_v$ (%) |
|---|---|---|---|
| Ground truth | [0.49, 1.30] | [2.0, 11.8] | [25.0, 45.0] |
| Synthetic data | [0.1, 3.9] | [0.2, 15.0] | [21, 56] |

TABLE 4
THE ROOT MEAN SQUARE ERRORS OF SURFACE PARAMETERS ESTIMATED FROM SYNTHETIC DATA

| | $s$ (cm) | $l$ (cm) | $m_v$ (%) |
|---|---|---|---|
| Regression | 0.28 | 0.52 | 3.9 |
| ANNs inversion | 0.45 | 0.90 | 5.4 |

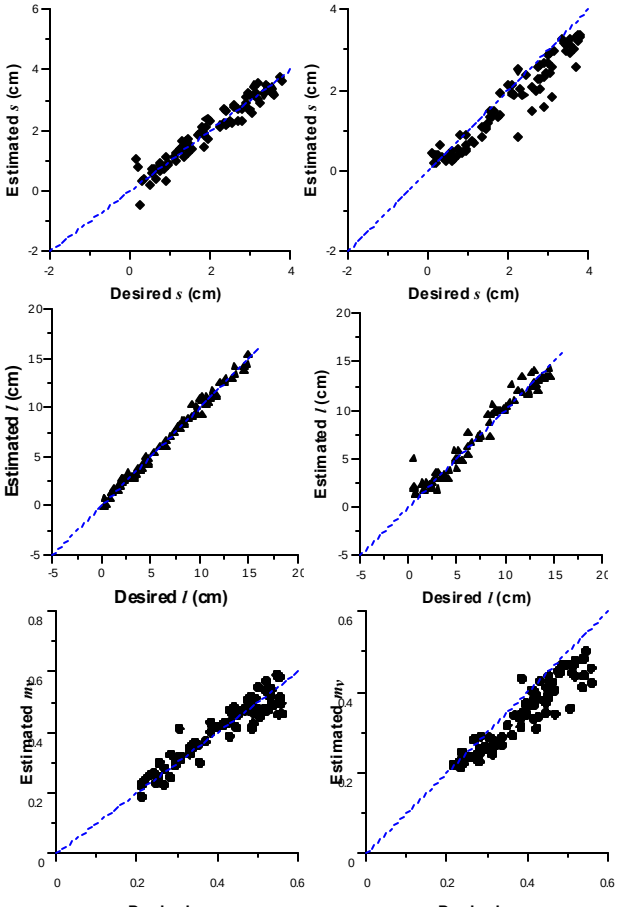


Fig. 8. Cross validation of two inversion models (Left: Multi-dimensional regression technique, Right: Inversion using artificial neural network).

A modified input vector is investigated to improve the soil moisture estimation by the ANNs inversion. For the purpose of searching signal parameters that have similar spatial structure with measured soil moisture, semi-variograms are estimated for soil moisture and signal parameters. The results are shown in Fig. 9 and 10. The scatter plots in Fig. 9 and 10 represent the experimental semi-variogram defined in equation (16). A solid line and a dotted line in Fig. 9 are a fitted exponential and Gaussian semi-variogram model, respectively. The exponential model of

$$\gamma(h) = 12.02\left(1-\exp\left(-\frac{3h}{36.35}\right)\right)+1.219 \qquad (19)$$

fits better than Gaussian. According to (19), the range value of measured soil moisture content is 36.35m. This means that the two variables at farther distance than 36.35m are independent. The signal parameters investigated include $\sigma^L_{HH}$, $\sigma^L_{VV}$, $\sigma^L_{HV}$, $\sigma^P_{HH}$, $\sigma^P_{VV}$, $\sigma^P_{HV}$, $\frac{\sigma^L_{VV}}{\sigma^L_{HH}}$, and $\frac{\sigma^P_{VV}}{\sigma^P_{HH}}$. Fig. 10 shows the experimental semi-variogram of co-polarization backscatter ratio $\frac{\sigma_{vv}}{\sigma_{hh}}$ at L- and P-band, which is sensitive to soil moisture according to Kim and van Zyl [6]. The range values of each signal parameter in Table 5 are obtained by fitting with the exponential semi-variogram model. It is found that the range value of co-polarization at site L1 approaches to that of soil moisture. Therefore the co-polarization ratios at L- and P- band are used for the elements of input vector, that is, $\left\{\frac{\boldsymbol{s}^L_{VV}}{\boldsymbol{s}^L_{HH}}, \boldsymbol{s}^L_{VV}, \frac{\boldsymbol{s}^P_{VV}}{\boldsymbol{s}^P_{HH}}, \boldsymbol{s}^P_{VV}, \boldsymbol{j}\right\}$ instead of $\{\boldsymbol{s}^L_{HH}, \boldsymbol{s}^L_{VV}, \boldsymbol{s}^P_{HH}, \boldsymbol{s}^P_{VV}, \boldsymbol{j}\}$. This attempt that denoted by the ANNs2 in Table 6, is proved to improve soil moisture content estimation; Average RMS error is reduced to 2.9% and it is consistent with the ground truth information that soil water content is higher in L1 site than in L2. The mean values of extracted and measured data are summarized in Table 6.

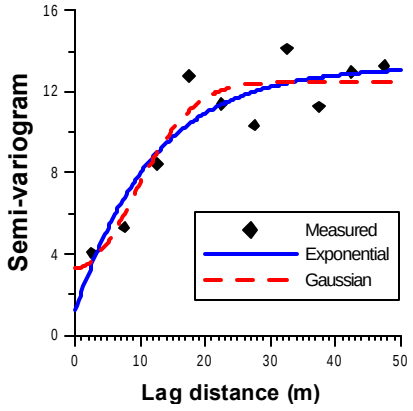


Fig. 9. Experimental semi-variogram of the soil moisture distribution measured on ground.

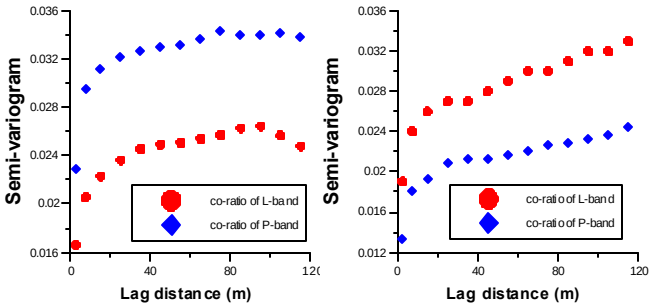


Fig. 10. Experimental semi-variograms of the co-polarization ratio distributions at L-and P- band (left: L1, right: L2).

TABLE 5
THE RANGE VALUES OF SIGNAL PARAMETERS IN m AT SITE L1 AND L2

| | $\sigma^L_{HH}$ | $\sigma^L_{VV}$ | $\sigma^L_{HV}$ | $\sigma^P_{HH}$ | $\sigma^P_{VV}$ | $\sigma^P_{HV}$ | $\frac{\sigma^L_{VV}}{\sigma^L_{HH}}$ | $\frac{\sigma^P_{VV}}{\sigma^P_{HH}}$ |
|---|---|---|---|---|---|---|---|---|
| L1 | 61.0 | 39.0 | 73.4 | 77.0 | 52.3 | 107 | 38.9 | 21.0 |
| L2 | 24.0 | 40.0 | 67.6 | 181 | 39.0 | 694 | 116 | 47.3 |

TABLE 6
SURFACE PARAMETERS ESTIMATED FROM THE AIRSAR DATA AND MEASURED ON GROUND

| | Site L1 | | | Site L2 | | |
|---|---|---|---|---|---|---|
| | $s$ (cm) | $l$ (cm) | $m_v$ (%) | $s$ (cm) | $l$ (cm) | $m_v$ (%) |
| Regression | 1.45 | 6.86 | 38.4 | 1.40 | 8.21 | 38.4 |
| ANNs | 1.73 | 7.97 | 35.4 | 1.77 | 7.00 | 39.8 |
| ANNs2 | 1.39 | 8.43 | 40.3 | 2.31 | 6.70 | 36.0 |
| Ground truth | 0.75 | 3.10 | 37.3 | 1.12 | 8.30 | 33.3 |

The ANNs means the ANNs inversion model using original input vectors, whereas the ANNs2 means the ANNs inversion model using the modified input vectors as described in section **III**.

## IV. SUMMARY

Estimation of bare surface soil moisture content is investigated using multi-frequency and multi-polarization AirSAR data set. During the PacRim-II AirSAR experiment in Korea, the AirSAR data and ground truth data were collected. Two existing inversion algorithms are applied to the AirSAR data set taken at the bare soil area in Jeju. The algorithms were developed to retrieve surface dielectric and roughness parameters simultaneously. One is the multi-dimensional regression model [1] and the other is the inversion using artificial neural networks [4]. The IEM model is used to calculate the backscatter coefficients with the surface parameters measured on ground and to generate synthetic data set in order to train the two inversion models. The regression model is proved to perform better than the ANNs inversion model when applied to the AirSAR data as well as the synthetic data. However, the ANNs inversion model can be modified to improve the soil moisture inversion. For example, the use of co-polarization ratio as elements of input vector improves the soil moisture estimation qualitatively and quantitatively. Still further

investigation should be continued for the ANNs inversion model.

Both *in situ* and AirSAR data are taken on the day, when it was cloudy and rained intermittently. Therefore this soil moisture estimation example shows the case of almost saturated surface. In this paper, only the average of surface parameters has been validated. But, the distributed estimation would be valuable both for the utilization in agriculture and assimilation to hydrological model [11].

## ACKNOWLEDGEMENT

We thank Prof. H. N. Hyun at Jeju National University, Mr. Y. C. Kim, and Mr. J. Y. Lee for helping ground truth during PacRim-II AirSAR experiment. And the advice of Prof. Y. Oh at Hong-Ik University was very helpful to prepare the ground truth experiment.

## REFERENCES

[1] M. S. Dawson, A. K. Fung, and M. T. Manry, "A robust statistical-based estimator for soil moisture retrieval from radar measurements," *IEEE Trans. Geosci. Remote Sensing*, vol. 35(1), pp. 57-67, 1997.

[2] M. C. Dobson, F. T. Ulaby, M. T. Hallikainen, and M. A. El-Rayes, "Microwave dielectric behavior of wet soil, part II: Dielectric mixing models," *IEEE Trans. Geosci. Remote Sensing*, vol. 32(2), pp. 438-448, 1994.

[3] P. C. Dubois, J. van Zyl, and T. Engman, "Measuring soil moisture with imaging radars," *IEEE Trans. Geosci. Remote Sensing*, vol. 33(4), pp. 915-926, 1995.

[4] A. K. Fung, *Microwave Scattering and Emission models and Their Applications Norwood*, MA: Artech House, 1994.

[5] R. Hoeben and P. A. Troch, "Assimilation of active microwave observation data for soil moisture profile estimation," *Water Resour. Res*., vol. 36(10), pp. 2805-2819, 2000.

[6] Y. Kim and J. van Zyl, "On the relationship between polarimetric parameters, " *Proceedings, Annual IGARSS*, pp. 1298-1300, 2000.

[7] Y. Oh, K. Sarabandi, and F. T. Ulaby, "An empirical model and an inversion technique for radar scattering from bare soil surfaces," *IEEE Trans. Geosci. Remote Sensing*, vol. 30(2), pp. 370-381, 1992.

[8] N. R. Peplinski, F. T. Ulaby, and M. C. Dobson, "Dielectric properties of soils in the 0.3-1.3GHz range," *IEEE Trans. Geosci. Remote Sensing*, vol. 33(3), pp. 803-807, 1995.

[9] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing*, Eds. Cambridge, MA: MIT Press, Vol. 1, 1986.

[10] P. J. van Oevelon and D. H. Hoekman, "Radar backscatter inversion techniques for estimation of surface soil moisture: EFEDA-Spain and HAPEX-Sahel case study," *IEEE Trans. Geosci. Remote Sensing*, vol. 37 (1), pp. 113-123, 1999.

[11] Valentijn R. N. Pauwls, Rudi Hoeben, Niko E. C. Verhost, and Francois P. De Troch, "The importance of the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation," *Jour. of Hydrology*, 251, pp. 88-102, 2001.

[12] P. K. kitanidis, *Introduction to Geostatistics: Application to Hydrogeology*, Cambridge University Press, 1997.