CHAPTER 7

COMPREHENSIVE EVALUATION OF SOIL MOISTURE RETRIEVAL MODELS UNDER DIFFERENT CROP COVER TYPES USING SENTINEL-1A DATA

7.1 INTRODUCTION

Soil moisture plays an important role for the proper crop growth in the field of agriculture (Gupta et al., 2017b). Information about soil moisture conditions is crucial in the crop yield prediction, hydrological and meteorological applications (Srivastava et al., 2006). The regular approximations of soil moisture from cropped surfaces are essential for the effective irrigation management and scheduling (Glenn et al., 2011). Some researcher's computed bare soil moisture by ground based scatterometer using microwave remote sensing as a powerful tool (Gupta et al., 2014; 2017a). However, retrieving soil moisture under vegetation using microwave remote sensing is a challenging task because of vegetation covered volume scattering and underlying soil surface scattering (Prakash et al., 2012). Vegetation canopies contain their own moisture contents which make difficulties in the underlying soil moisture retrieval (Bindlish and Barros, 2001; Jain and Singh, 2016). Due to multiple scattering effects, the relations between the observed backscattering and combined contributions of the vegetation and soil surface water content is found highly nonlinear (Bindlish and Barros, 2001; Notarnicola et al., 2006).

However, in the last decade few modelling approaches were applied for the retrieval of soil moisture under vegetation (Singh et al., 1996; Srivastava et al., 2002; 2006). At the higher angle of incidence, SAR signal experiences an increased path length through the vegetation volume and consequently interacted more with the crop canopy. The retrieved soil moisture by the model is improved considerably after including the influence of crop cover (Patel et al., 2001; Srivastava et al., 2002). Some semi-empirical and empirical models are used to retrieve soil moisture with vegetation and compared results at different bands (Saleh et al., 2006; Gupta et al., 2017a). Theoretical approach based on physical models was used to incorporate the effect of surface roughness and crop cover (Fung, 1994).

RF is the robust model to the outliers and can run efficiently on large data-sets for the regression problems (Breiman, 2001). Robustness of RFR model over SVR and ANNR has been shown for the biomass estimation of wheat crop (Wang et al., 2016). Soil mapping is done using random forest model in Africa and achieved improved accuracy results (Hengl et al., 2015). External parameter orthogonalization coupled with RF, SVM, partial least squares regression and ANN models applied on a wider set of soil properties have provided satisfactory results (Wijewardane et al., 2016).

SVR is becoming popular in the field of geo/ biophysical parameters retrieval in the last few years. It was anticipated firstly for the regression problems by Vapnik et al. (1997). Good results are found for the soil moisture retrieval using ground based scatterometer data at X-band by SVR model (Gupta et al., 2015). A comparison is done between SVR and ANNR for the soil moisture retrieval in the base agricultural areas using C-band scatterometer data (Pasolli et al., 2011). SVR model using RADARSAT-2 data at HH and HV polarizations enhanced the estimated soil moisture content in the mountain area because the vegetation effect is found separated from the radar signal at HV polarization (Pasolli et al., 2012). The estimated values of soil moisture by SVR indicates its better performance than BPANN and MLR models using tropical rainfall measuring mission and the advanced very high resolution radiometer data (Ahmad et al., 2010).

BPANN, radial basis function neural network (RBFANN), generalized regression artificial neural network (GRANN) and linear regression models (LRM) are compared for the estimation of soil moisture. Marginally better results were found using BPANN at HH polarization while RBFANN has shown good results using VV polarization followed by GRANN and LRM models (Gupta et al., 2017a). The performance of SVM models are found better than ANN models for the soil moisture forecasting during comparison between the results obtained by SVM and ANN (Gill et al., 2006). However, ANN model yields superior results over the linear statistical technique for the mapping of soil moisture patterns (Narayanan and Hegde, 2008). A review is given by Ali et al. (2015) for better understanding of the machine learning models for the soil moisture retrieval.

After a comprehensive survey of the literature, very few studies are carried out for the soil moisture retrieval using Sentinel-1 satellite data (Hornáček et al., 2012; Paloscia et al., 2013; Navarro et al., 2016). This study is done for the first time to evaluate the Sentinel-1A SAR data for the soil moisture retrieval in tropical condition of India. On purview of the above, the main objectives of the study are (i) the evaluation of RFR, SVR and ANNR models for the retrieval of soil moisture covered by winter wheat, barley and corn crops, (ii) performance assessment of the retrieved soil moisture content using the VV and VH polarizations.

7.2 MATERIALS AND METHODOLOGY

7.2.1 Ground soil samples collection

The soil samples covered by the winter wheat, barley and corn crops were collected from the 5 different locations in the each field for the gravimetric soil moisture content measurement on 28 February 2015, 05 March 2015 and 29 April 2015.The ground data collection was made on the same date of Sentinel-1A satellite passing over the study area in the Varanasi district, India. The study covered a total area of 192 km^2 and had centre latitude 25° 17' 51''N and longitude 82° 56' 36''E. A part of study area is shown using Sentinel-1A satellite image in Figure 7.1.

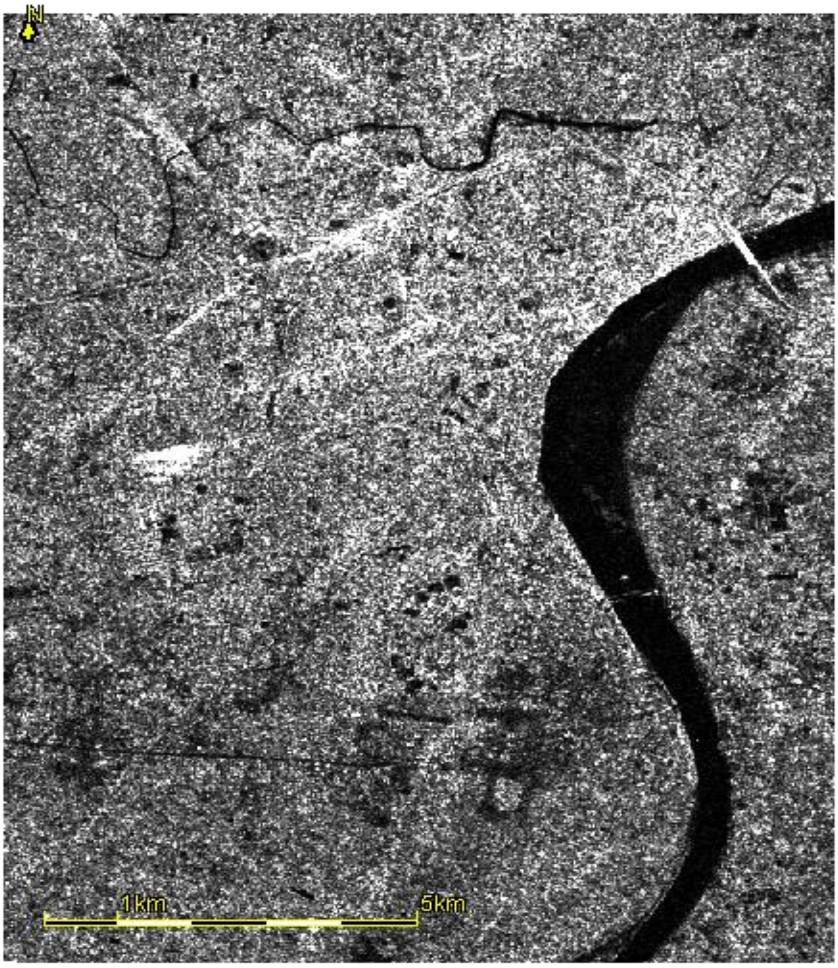

Figure 7.1 Study area using Sentinel-1A satellite image at VH polarization

After ground sampling, the samples were dried in an oven at 110°C for 18h. The samples were weighted before and after drying for the computation of gravimetric soil moisture. The gravimetric soil moisture was defined as the ratio of weight of water present in soil to the weight of dry soil (Gupta et al., 2014). The average values of the of the gravimetric soil moisture content were taken to compute the percent of soil moisture. The gravimetric soil moisture can be computed using the equation 7.1.

$$SM\ (\%) = \frac{W_{wet} - W_{dry}}{W_{dry}} \times 100 \quad (7.1)$$

where W_{wet} and W_{dry} are the weights of the soil samples before and after dryness respectively

7.2.2 SAR data collection and processing

Sentinel-1A satellite was launched in 3 April 2014 by the ESA. Satellite data of Varanasi district from 28 February 2015 to 29 April 2015 at VV and VH polarizations in the descending pass direction at C-band (5.405 GHz) was freely downloaded (https://scihub.copernicus.eu/dhus/#/home). The specification of the Sentinel-1A satellite data is described in the Table 7.1.

Table 7.1 Specification of Sentinel-1A SAR satellite data

Satellite and band	Date of acquisition	Mode	Pass direction	Polarization	Product type	Resolution (m x m)	Product level
Sentinel-1A and C-band	28/02/2015	IW	Descending	VV & VH	GRD	5x20	L1
	05/03/2015	IW	Descending	VV & VH	GRD	5x20	L1
	29/04/2015	IW	Descending	VV & VH	GRD	5x20	L1

For the processing of satellite data-sets, the SNAP software version 1.1.1 with Sentinel-1 toolbox (S-1 TBX) was freely downloaded (http://step.esa.int/main/download/). The downloaded Interferometric Wide-Ground Range Detected (IW-GRD) product file data in the ZIP format was extracted for the further processing. The raster data was imported and subsetting was performed for the required area and radiometric calibration was done. Speckle filtering using refined lee filter and geometric correction were done by applying ellipsoid correction. For the proper computation of the backscattering values, the linear to/from dB raster conversion was done using SNAP software to convert intensity (m^2m^{-2}) values into decibel (dB) values.

7.2.3 Random forest regression model

RFR model uses the samples of the training data to generate various regression trees. These regression trees are the set of conditions applied from the root to the leaves of the tree. The RFR model combined the results from individual trees and generally produces better results (Breiman, 1996). Regression trees are grown by selecting the best split at each node using predictor variables. The combination of output from trees inclines to smooth the variance between trees and gives the more generalization capacity of the model. The RFR does not overfit, always converges, and has bounded generalization error on adding more trees. The out-of-bag error estimates are created from the data that are not included in the bootstrapped samples (Breiman, 2001). The two-thirds of the selected training data (bootstrap samples) were used to train the model, whereas one third (out-of-bag samples) data were used for the model validation (Siegmann and Jarmer, 2015). The total number of regression trees (ntree; 112) and the number of input variables per node (mtry; 2) were optimized using training data that provided the lowest RMSE.

7.2.4 Support vector regression model

SVM has shown the robustness for the classification and regression problems (Walton, 2008; Mishra et al., 2014). SVM maps the independent variables where complex nonlinear decision boundaries between the classes become linear in the higher dimensional space. An optimal linear separator is established that maximizes the margin between the classes in the high-dimensional space (Vapnik, 2000; Russell and Norvig, 2003). The solution is generalized by maximizing the margin and overfitting is reduced. Large input data spaces and training data-sets can be handled using the model (Burges, 1998). SVR is used to describe a real-valued output function for the given independent input variables. The model relates the concept of ε-insensitive loss function that ignores point errors within a distance of ε from the true value by weighting them with zero (Smola and Schölkopf, 2004).The parameters such as loss function ($\varepsilon = 1$) and penalty parameter ($C = 10$) were used for the optimization.

7.2.5 Artificial neural network regression model

The back propagation ANN is consists of an input layer, a hidden layer and an output layer. These layers are fully interconnected having simple processing units (neurons) at every layer. The input data information is given to the input layer and then forwarded after multiplying by a weight factor and adding a bias to the hidden layer. The output values of the hidden layer neurons are considered as input for the output layer (Rumelhart et al., 1986).The σ^0 were taken as input data at input layer neurons and vegetated soil moisture as output data at output layer neurons. A transfer function was defined to know the networks nonlinear relationship between input and output data in the hidden and output layers. Weights and bias values were updated according to back-propagation of the error to obtain the best fitness function (Haykin, 1994; Munakata, 1998). The ANN was optimized using tan-sigmoid transfer function at hidden neurons and log-sigmoid transfer function at the output neurons. Eight neurons at the hidden layer were taken to optimize the ANN using trial and error method.

7.2.6 Performance indicators

The performance of the developed regression models was evaluated by analysing the different indices such as %bias, RMSE and Nash Sutcliffe efficiency (NSE).

7.2.6.1 % bias calculation

The %bias makes available the average tendency of the retrieved values to be larger or smaller than their observed values. The optimum value of %bias is zero whereas; positive and negative values indicate the over and under retrieved values by the model, respectively. The %bias calculation can be made by the mathematical relation given by the equation 7.2.

$$\%bias = 100 * \left[\frac{\sum(y_i - x_i)}{\sum x_i}\right] \tag{7.2}$$

7.2.6.2 Nash Sutcliffe efficiency

The NSE (Nash and Sutcliffe, 1970) is an important parameter for the performance evaluation. It is based on the sum of the square of difference between the retrieved and observed values normalized by the variance of observed values. The NSE can be computed using equation 7.3.

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(y_i - \bar{x}_i)^2} \tag{7.3}$$

7.2.6.3 Root mean square error calculation

The RMSE was used to measure the differences between observed and retrieved values of vegetated soil moisture. The RMSE was calculated using the relation given by the equation 7.4.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \tag{7.4}$$

where x_i, y_i and n are the observed, retrieved values and number of observations, respectively.

7.3 RESULTS AND DISCUSSION

The soil samples collected in the field and σ^0 were interpolated into 61 data-sets at the interval of one day for measurements made during 28 February, 2015 to 29 April, 2015. The 41 sample data-sets were used for the training and 20 samples for testing of the RFR, SVR and ANNR models. The results were found better at VV polarization in compare to VH polarization for the all the crop types covered soil moisture. The microwave response at different polarizations depends on several factors such as shape, size and orientation of the target elements. The horizontal polarization provides the measure of the horizontal dimension, whereas the vertical polarization provides the measure of the vertical dimension of the backscattering elements (Prasad, 2009). The RFR model performance depends on determining the number of tress and predictors in each node (Shataee et al., 2012). The total 112 number of trees were used to make a graph between average squared error and number of trees for the training and testing samples. The produced graphs indicate an optimum number of trees 40 which showed stable response with lowest error.

7.3.1 Retrieval of wheat crop covered soil moisture using RFR, SVR and ANNR models

The graphs between average squared error and number of trees for the wheat crop covered soil moisture at VV and VH polarizations using RFR model are shown in Figures 7.2 and 7.3, respectively. The results were evaluated and compared using adj. R^2, %bias, NSE and RMSE values. The scatter plots between observed and retrieved soil moisture covered by wheat crop using RFR, SVR and ANNR models at VV and VH polarizations are shown in Figures 7.4 and 7.5, respectively.

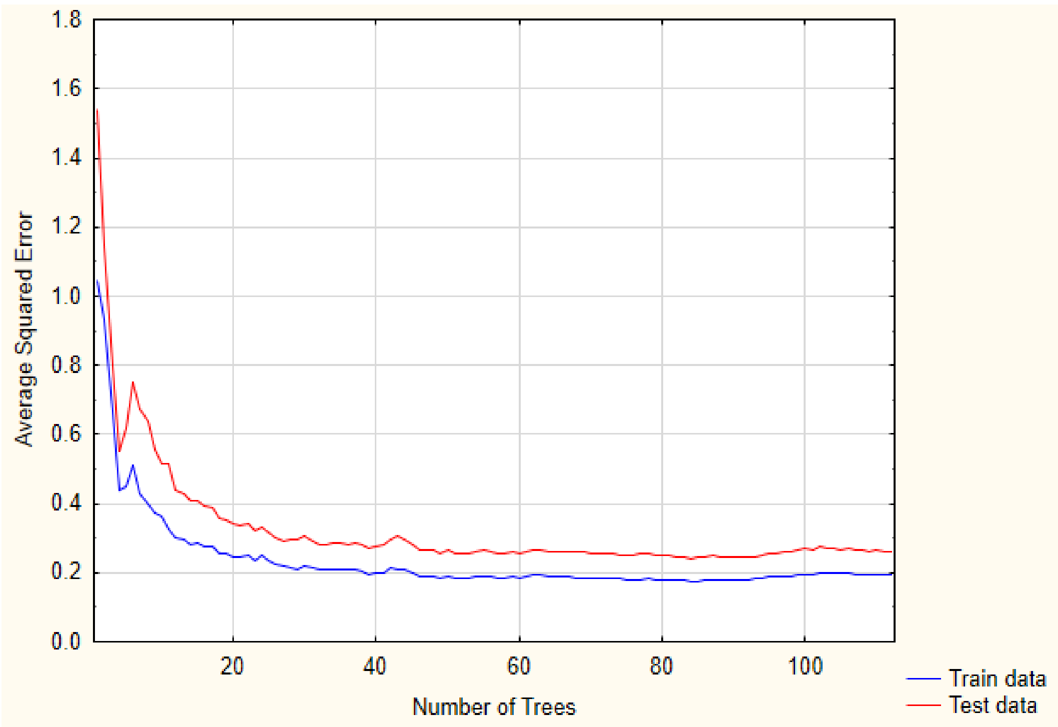

Figure 7.2 Average square error against number of trees in training and testing data for the soil moisture retrieval covered by wheat crop at VV polarization

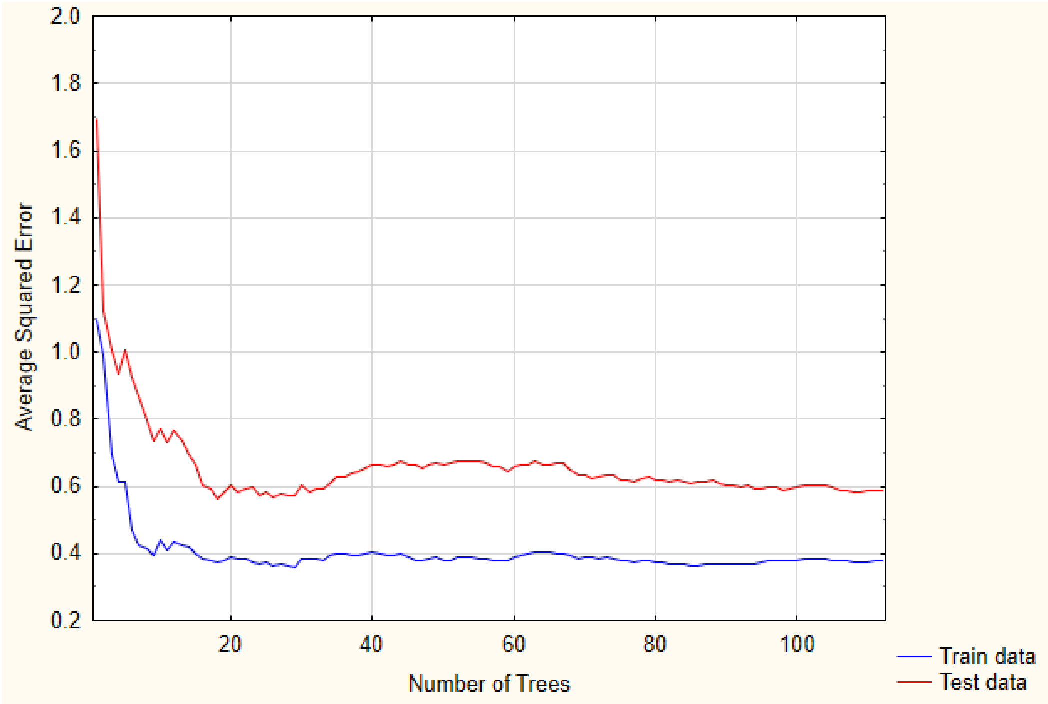

Figure 7.3 Average square error against number of trees in training and testing data for the soil moisture retrieval covered by wheat crop at VH polarization

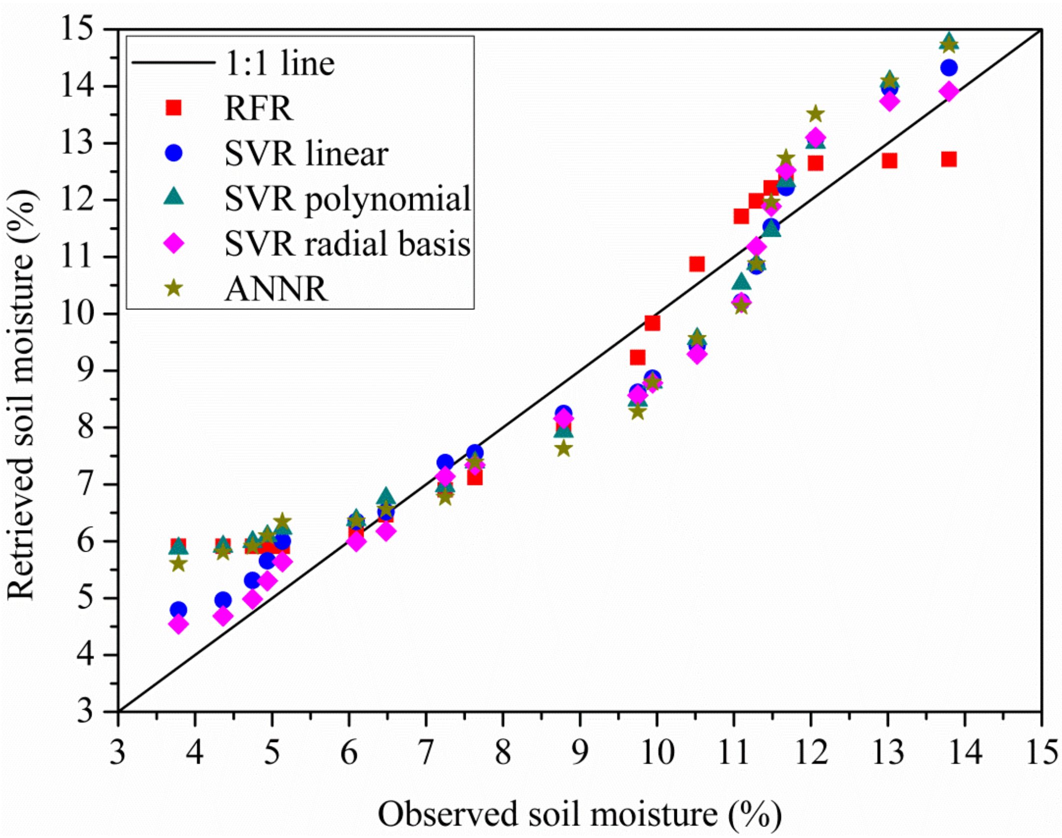

Figure 7.4 Scatter plots of observed and retrieved soil moisture covered by wheat crop using RFR, SVR and ANNR models at VV polarization

The overall better results were found using SVR with radial basis kernel (adj. R^2 = 0.95 and RMSE = 0.68) in comparison to RFR (adj. R^2 = 0.94 and RMSE = 0.86), SVR with linear (adj. R^2 = 0.94 and RMSE = 0.72), SVR with polynomial (adj. R^2 = 0.90 and RMSE = 0.99) and ANNR (adj. R^2 = 0.89 and RMSE = 1.05) models at VV polarization. However, the results obtained by SVR with linear kernel (adj. R^2 = 0.94 and RMSE = 0.72) provided similar results as RFR (adj. R^2 = 0.94 and RMSE = 0.86) at VV polarization. It indicates the high sensitivity of σ° with the wheat crop covered soil moisture at VV polarization. The retrieval of soil moisture covered by wheat crop using RFR, SVR and ANNR models at VV and VH polarizations are described in Table 7.2.

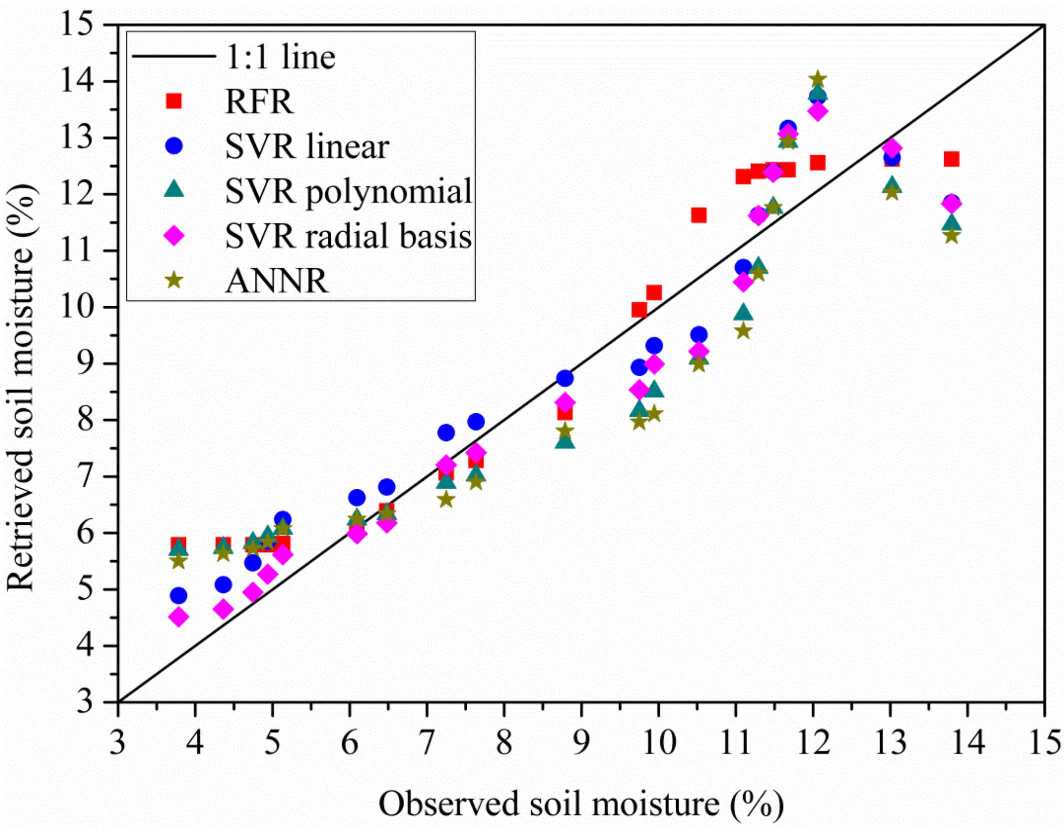

Figure 7.5 Scatter plots of observed and retrieved soil moisture covered by wheat crop using RFR, SVR and ANNR models at VH polarization

Table 7.2 Retrieval of soil moisture covered by wheat crop using RFR, SVR and ANNR models at VV and VH polarizations

Crop covered soil moisture	Polarizations	Models	Regression equations	adj- R square	%bias	NSE	RMSE
Wheat	VV	RFR	y = 1.47+0.87x	0.94	3.84	0.92	0.86
		SVR linear	y = 0.66+0.94x	0.94	1.15	0.95	0.72
		SVR polynomial	y = 1.33+0.88x	0.90	3.19	0.90	0.99
		SVR radial basis	y = 0.17+0.98x	0.95	-0.41	0.95	0.68
		ANNR	y = 1.08+0.91x	0.89	3.03	0.88	1.05
	VH	RFR	y = 1.26+0.91x	0.94	5.33	0.92	0.90
		SVR linear	y = 1.39+0.87x	0.92	3.15	0.91	0.93
		SVR polynomial	y = 1.79+0.78x	0.84	-1.20	0.84	1.22
		SVR radial basis	y = 0.50+0.94x	0.92	-0.79	0.92	0.86
		ANNR	y = 1.68+0.78x	0.82	-2.28	0.82	1.30

In case of VH polarization, the performance of RFR (adj. R^2 = 0.94 and RMSE = 0.90) was found better than SVR with linear (adj. R^2 = 0.92 and RMSE = 0.93), polynomial (adj. R^2 = 0.84 and RMSE = 1.22), radial basis function (adj. R^2 = 0.92 and RMSE = 0.86) kernels and ANNR (adj. R^2 = 0.82 and RMSE = 1.30) models for the retrieval of soil moisture covered under wheat crop. However, the performance at VV polarization was found better than the VH polarization because of relatively higher attenuation of the signals at VH polarization.

7.3.2 Retrieval of barley crop covered soil moisture using RFR, SVR and ANNR models

The graphs between average squared error and number of trees for the barley crop covered soil moisture at VV and VH polarizations using RFR model are shown in Figures 7.6 and 7.7, respectively. The results obtained for the barley crop covered soil moisture were found somewhat lower than the results obtained for the wheat crop covered soil moisture at both the polarizations. However, the performance of SVR with polynomial kernel (adj. R^2 = 0.92 and RMSE = 0.85) was found better for the retrieval of soil moisture covered with barley than the soil moisture covered with wheat crop using SVR with polynomial kernel (adj. R^2 = 0.90 and RMSE = 0.99) at VV polarization.

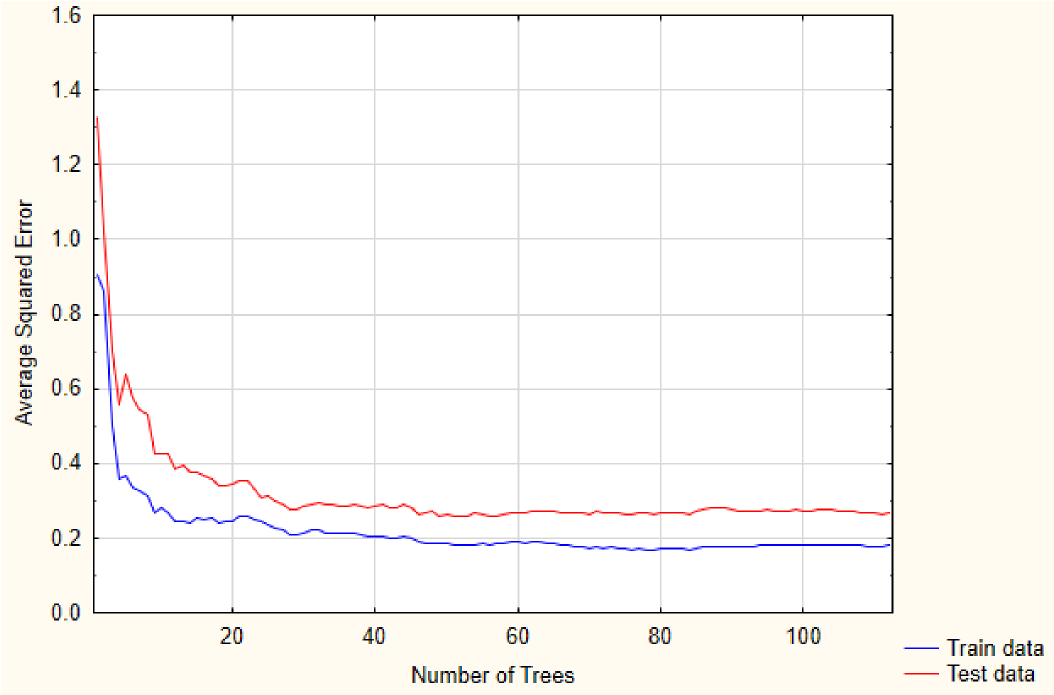

Figure 7.6 Average square error against number of trees in training and testing data for the soil moisture retrieval covered by barley crop at VV polarization

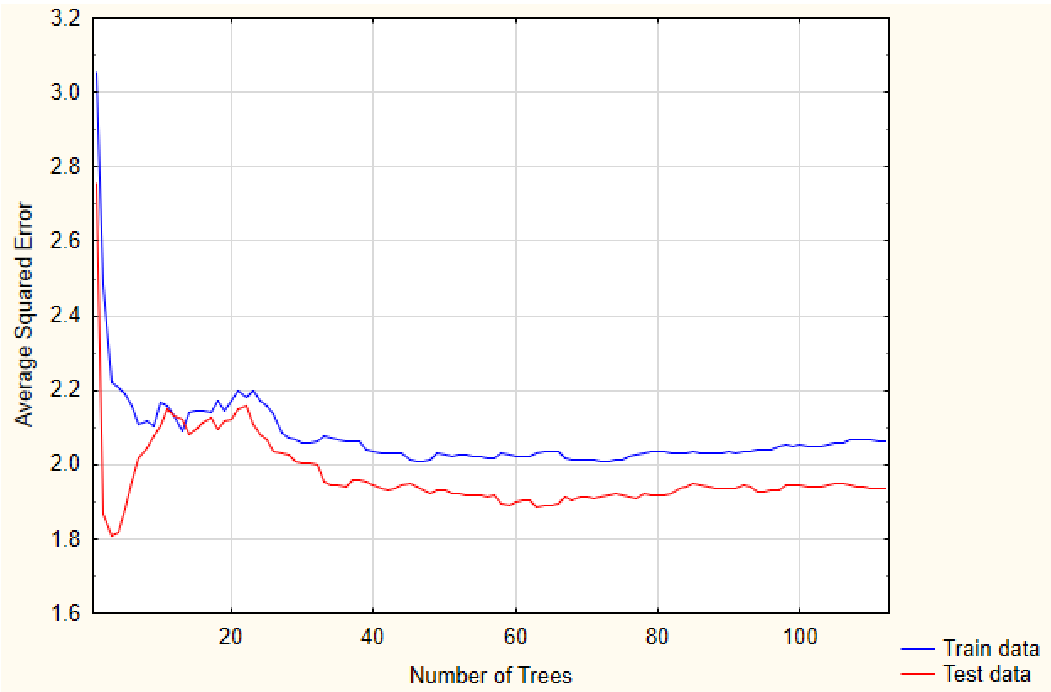

Figure 7.7 Average square error against number of trees in training and testing data for the soil moisture retrieval covered by barley crop at VH polarization

The scatter plots between observed and retrieved soil moisture covered by barley crop using RFR, SVR and ANNR models at VV and VH polarizations are shown in Figures 7.8 and 7.9, respectively.

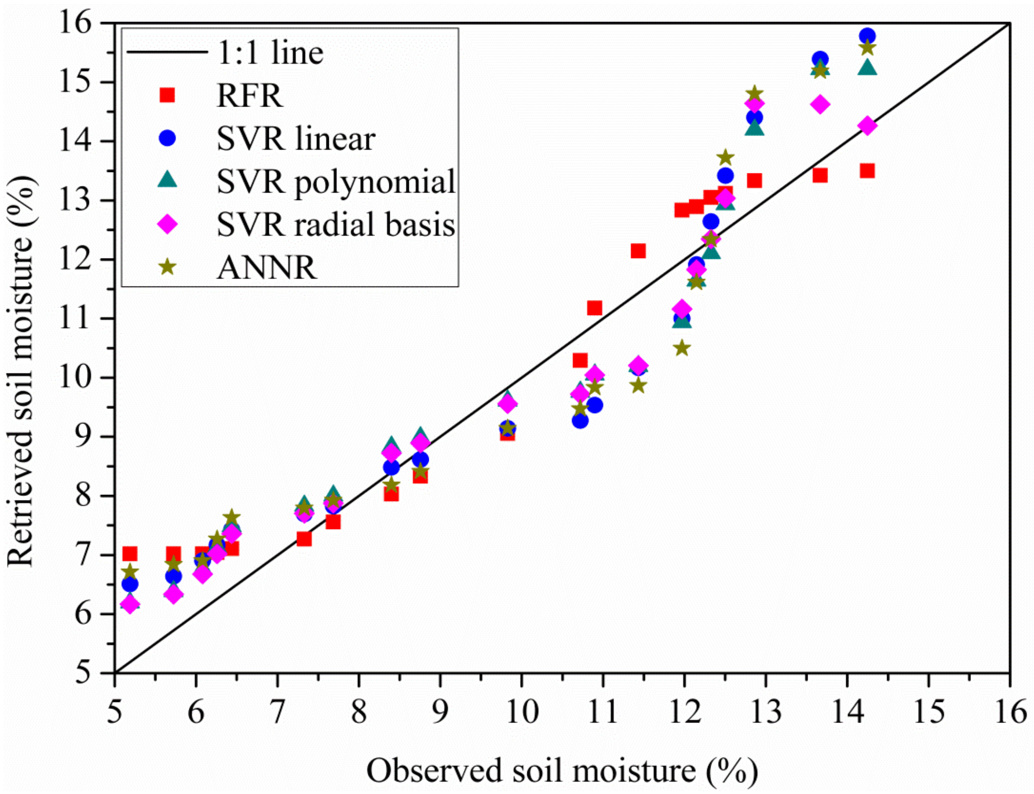

Figure 7.8 Scatter plots of observed and retrieved soil moisture covered by barley crop using RFR, SVR and ANNR models at VV polarization

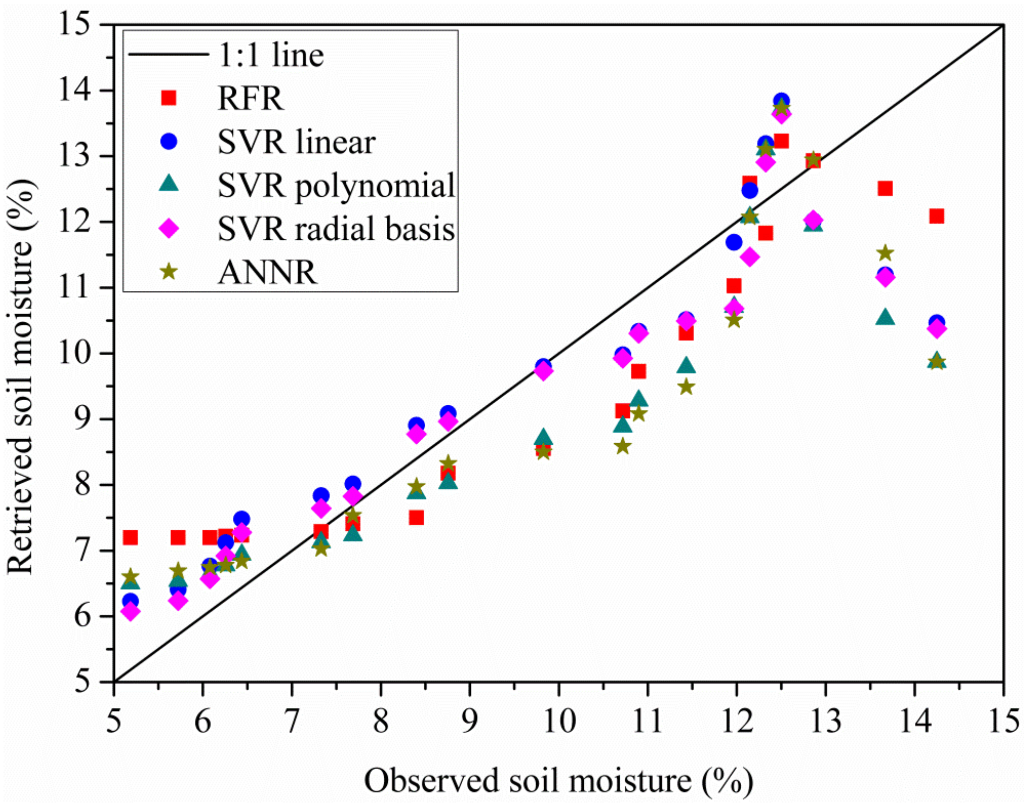

Figure 7.9 Scatter plots of observed and retrieved soil moisture covered by barley crop using RFR, SVR and ANNR models at VH polarization

The sensitivity of the σ° with the barely crop covered soil moisture was found better at the VV polarization than the VH polarization. The poor and comparable results were found using SVR polynomial (adj. R^2 = 0.74 and RMSE = 1.55) and ANNR (adj. R^2 = 0.74 and RMSE = 1.51) models in compare to other models at VH polarization for the barley crop covered soil moisture. The poor performance at VH polarization was found because of highly attenuated signals received at the VH polarization. The random orientation of the leaves was one of the major issues in case of barley crop covered soil moisture for the highly attenuated signals. The retrieval results for the soil moisture covered by barley crop using RFR, SVR and ANNR models at VV and VH polarizations are given in Table 7.3.

Table 7.3 Retrieval of soil moisture covered by barley crop using RFR, SVR and ANNR models at VV and VH polarizations

Crop covered soil moisture	Polarizations	Models	Regression equations	adj- R square	%bias	NSE	RMSE
Barley	VV	RFR	y = 1.25+0.91x	0.94	3.49	0.94	0.77
		SVR linear	y = 0.66+0.96x	0.88	2.83	0.89	1.02
		SVR polynomial	y = 0.92+0.93x	0.92	2.60	0.92	0.85
		SVR radial basis	y = 0.93+0.92x	0.93	1.92	0.94	0.77
		ANNR	y = 0.93+0.93x	0.85	2.72	0.87	1.11
	VH	RFR	y = 2.33+0.74x	0.86	-2.11	0.86	1.12
		SVR linear	y = 2.69+0.72x	0.82	-0.57	0.84	1.23
		SVR polynomial	y = 2.47+0.68x	0.74	-6.27	0.74	1.55
		SVR radial basis	y = 2.59+0.71x	0.84	-2.80	0.83	1.23
		ANNR	y = 2.33+0.71x	0.74	-5.42	0.75	1.51

7.3.3 Retrieval of corn crop covered soil moisture using RFR, SVR and ANNR models

The graphs between average squared error and number of trees for the corn crop covered soil moisture at VV and VH polarizations using RFR model are shown in Figures 7.10 and 7.11, respectively.

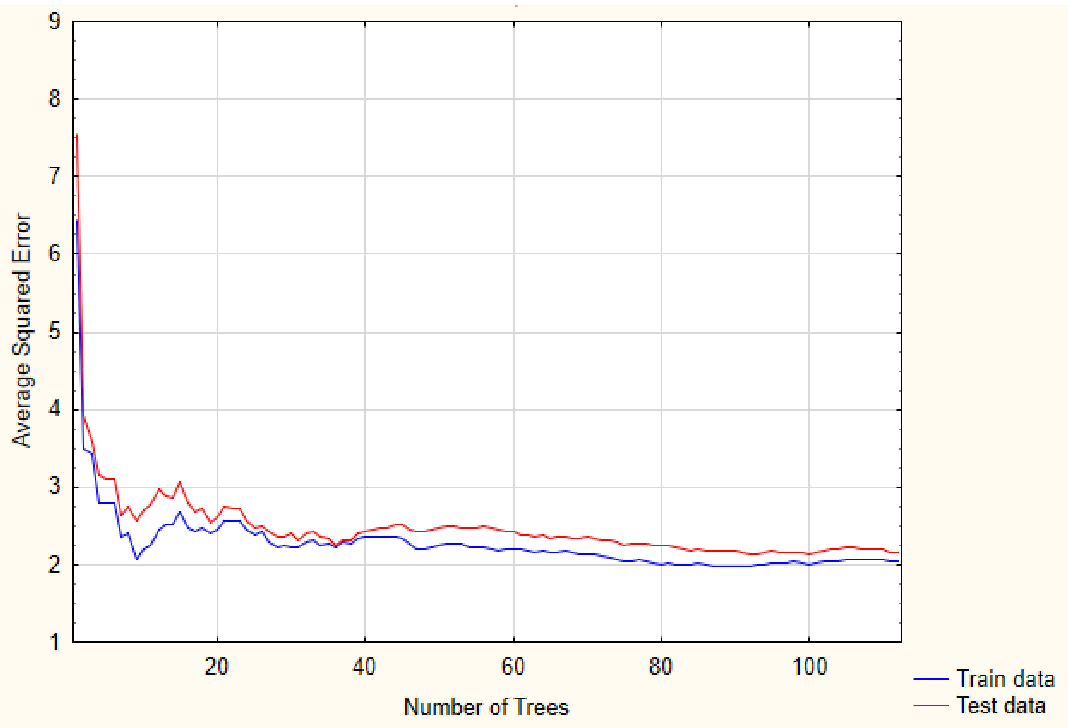

Figure 7.10 Average square error against number of trees in training and testing data for the soil moisture retrieval covered by corn crop at VV polarization

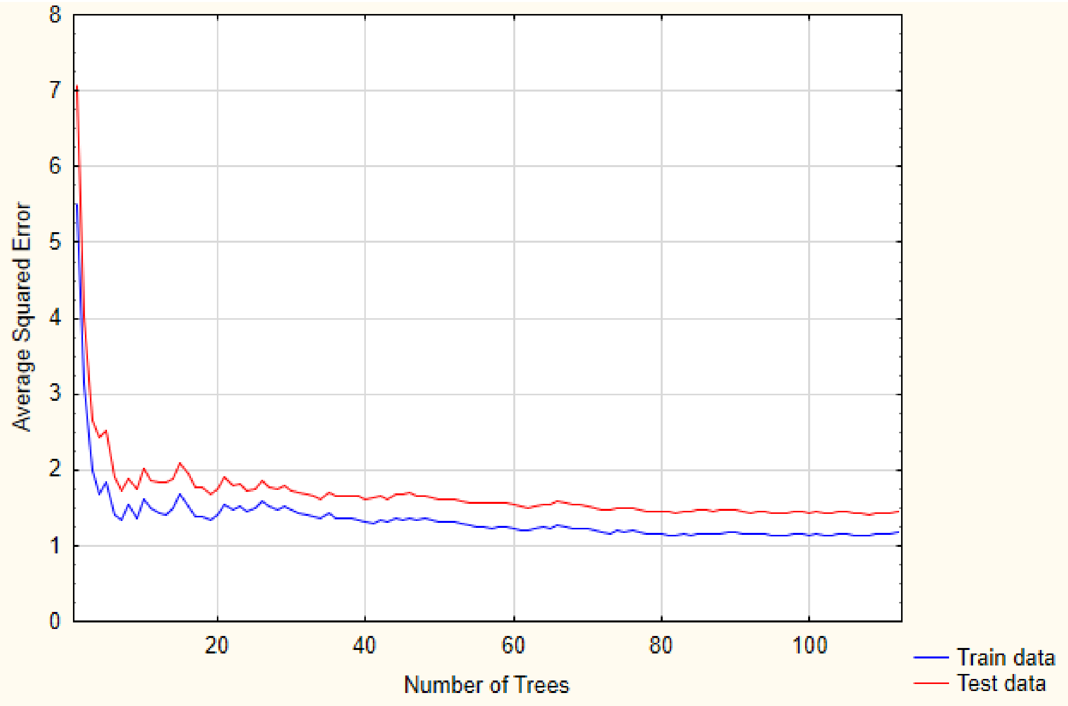

Figure 7.11 Average square error against number of trees in training and testing data for the soil moisture retrieval covered by corn crop at VH polarization

Overall good and similar results were found using SVR with linear (adj. R^2 = 0.94 and RMSE = 0.79), SVR with radial basis (adj. R^2 = 0.94 and RMSE = 0.87) kernels and RFR (adj. $R^2 = 0.94$ and RMSE = 0.87) models for the corn covered soil moisture retrieval at VV polarization. Results were also found good using RFR (adj. $R^2 = 0.93$ and RMSE = 0.94) and SVR with radial basis kernel (adj. $R^2 = 0.91$ and RMSE = 1.12) at VH polarization. A large difference was found between SVR with linear (adj. R^2 = 0.94 and RMSE = 0.79) at VV polarization and SVR with linear (adj. $R^2 = 0.76$ and RMSE = 1.46) at VH polarization for the corn crop covered soil moisture. The scatter plots between observed and retrieved soil moisture covered by corn crop using RFR, SVR and ANNR models at VV and VH polarizations are shown in Figures 7.12 and 7.13, respectively.

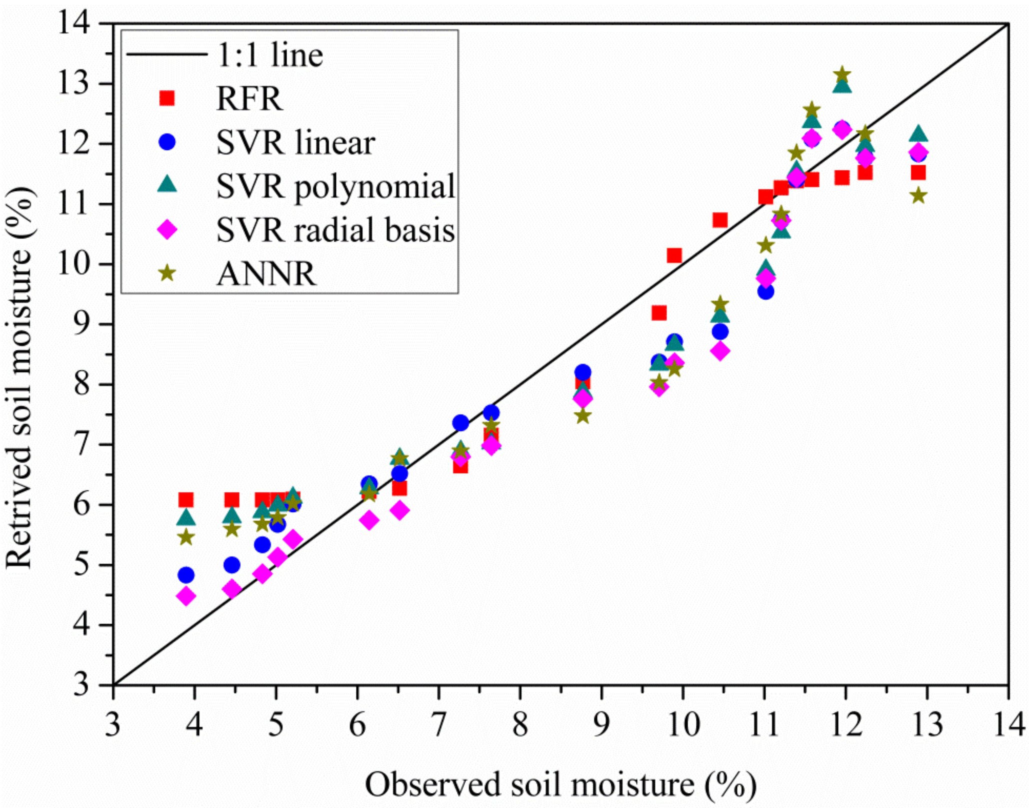

Figure 7.12 Scatter plots of observed and retrieved soil moisture covered by corn crop using RFR, SVR and ANNR models at VV polarization

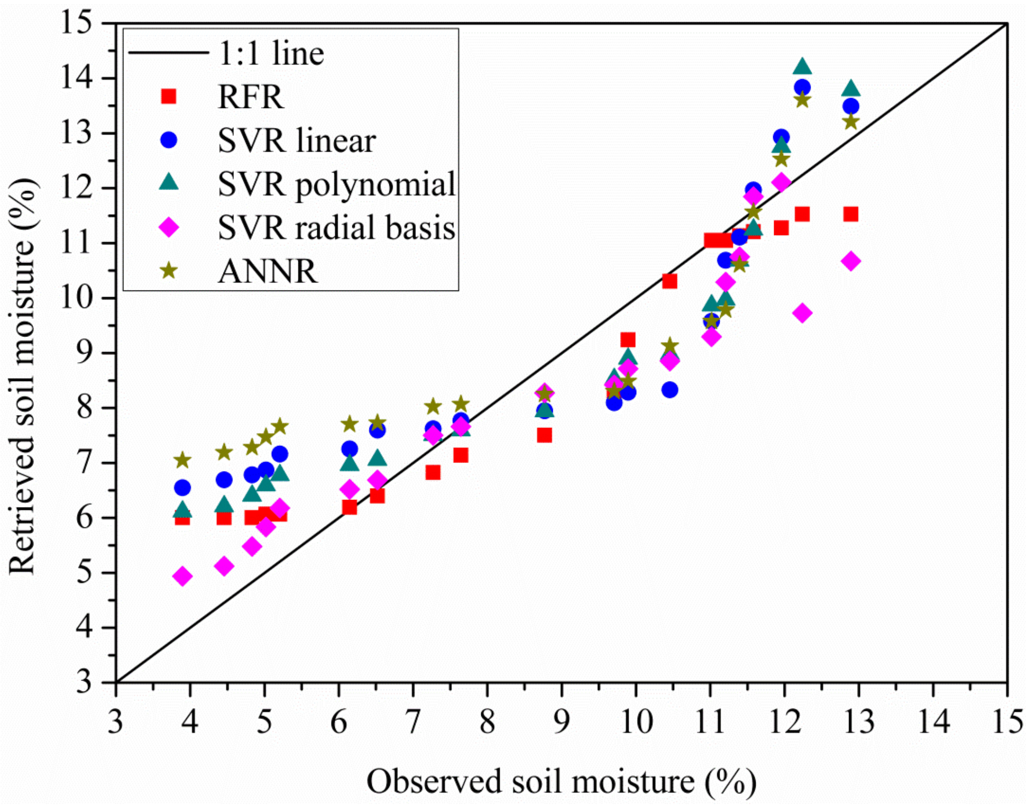

Figure 7.13 Scatter plots of observed and retrieved soil moisture covered by corn crop using RFR, SVR and ANNR models at VH polarization

Poor performances were shown by SVR with polynomial (adj. R^2 = 0.74 and RMSE = 1.55), ANNR (adj. R^2 = 0.74 and RMSE = 1.51) models for the soil moisture covered with barley crop and ANNR model with adj. R^2 = 0.74 and RMSE = 1.63 for the soil moisture covered by corn crop at VH polarization. The high sensitivity of backscattering values with the corn crop covered soil moisture was found at VV polarisation in comparison to the VH polarization. The soil moisture retrieval results covered by corn crop using RFR, SVR and ANNR models at VV and VH polarizations are described in Table 7.4.

Table 7.4 Retrieval of soil moisture covered by corn crop using RFR, SVR and ANNR models at VV and VH polarizations

Crop covered soil moisture	Polarizations	Models	Regression equations	adj- R square	%bias	NSE	RMSE
Corn	VV	RFR	y = 2.00+0.78x	0.94	1.39	0.91	0.87
		SVR linear	y = 1.27+0.83x	0.94	-2.16	0.93	0.79
		SVR polynomial	y = 1.66+0.81x	0.89	-0.13	0.89	0.97
		SVR radial basis	y = 0.36+0.90x	0.94	-5.61	0.91	0.87
		ANNR	y = 1.41+0.83x	0.87	-0.77	0.88	1.02
	VH	RFR	y = 2.07+0.75x	0.93	-0.74	0.90	0.94
		SVR linear	y = 2.90+0.71x	0.76	4.89	0.75	1.46
		SVR polynomial	y = 2.27+0.77x	0.83	3.45	0.82	1.23
		SVR radial basis	y = 2.16+0.71x	0.91	-4.20	0.85	1.12
		ANNR	y = 3.98+0.60x	0.74	6.45	0.69	1.63

where x and y are the observed and retrieved vegetation covered soil moisture, respectively

7.4 CONCLUSION

In the present study, three different types of regression models such as RFR, SVR and ANNR were evaluated for the retrieval of soil moisture covered with the winter wheat, barley and corn crops. The performing efficiencies of SVR and RFR models were found better in comparison to ANNR model at VV polarization for the soil moisture retrieval under different crops. The performance at VH polarization was found lower than the VV polarization for soil moisture retrieval using the Sentinel-1A satellite data. The results provided could be beneficial for the accurate retrieval of soil moisture under different crop type.