COMPARISON OF MACHINE LEARNING METHODS FOR PREDICTING QUAD-POLARIMETRIC PARAMETERS FROM DUAL-POLARIMETRIC SAR DATA

Katalin Blix, Martine M. Espeseth, Torbjørn Eltoft

Department of Physics and Technology
UiT The Arctic University of Norway, Tromsø, Norway

ABSTRACT

This work evaluates three machine learning methods with respect to their ability of learning the functional relationship between dual-polarimetric (dual-pol) input data and quad-polarimetric (quad-pol) output parameters. We chose to study and compare the learning strength of a Neural Network (NN) approach, two kernel-methods, the Support Vector Machine (SVM) and the Gaussian Process Regression (GPR). Overlapping quad-pol Radarsat-2 (RS2) and dual-pol ScanSAR Sentinel-1 (S1) sea ice Synthetic Aperture Radar (SAR) scenes, with 20 minutes time difference, were used for establishing the relationship between the dual-pol input data and corresponding quad-pol output parameters. We then used the learned relationship to predict quad-pol parameters for the overlapping S1 dual-pol scene, and show the results of the three machine learning methods, visually, by showing images of the predicted polarimetric features, and quantitively, by computing statistical performance measures. The results indicate that all three methods have strong learning capacity, however, the computed statistical measures and the visual comparisons suggest the best performance for the GPR model.

Index Terms— Neural Network, Support Vector Machine, Gaussian Process Regression, Polarimetry, Sea Ice

1. INTRODUCTION

The on-going climate change has dramatic impact on the Arctic sea ice cover, and has enabled more human activities in this sensitive environment. This puts forward new demands for reliable, high-resolution, large scale sea ice monitoring. Currently, this type of monitoring can only be performed from space using single or dual-pol Synthetic Aperture Radar (SAR) systems, such as the Sentinel 1A and 1B satellites. On the other hand, previous studies have found (see e.g. [1]) that some quad-pol parameters are sensitive to ice type characterics, like roughness and dielectric properties. Unfortunately, the narrow swath-width of a quad-pol SAR (usually around 25-50 km) makes it challenging to use it for operational monitoring purposes. In our previous work [2], we studied the possibility of using the machine learning Gaussian Process Regression (GPR) model for learning the relationship between dual-pol observations (input) and quad-pol parameters (output). Although the GPR model showed promising performance, there are several other relevant machine learning approaches frequently used for remote sensing applications; for instance, Neural Networks (NN) and the Support Vector Machine (SVM). The SVM, similarly to the GPR, is a non-linear kernel method, however, the underlying mathematical principles differs from those of the GPR approach.

Our objective in this work is to examine and compare the performances of the three sophisticated machine learning methods, the NN, the GPR and the SVM for polarimetric information up-scaling from dual-pol to quad-pol systems. Dual-pol ScanSAR Sentinel-1 (S1) scene overlapping with a quad-pol Radarsat-2 (RS2) scene are used to evaluate the up-scaling performances of these machine learning methods. In the training phase, the *true* output parameters, which describe sea ice surface characteristics such as surface roughness and dielectric properties, were generated from the RS2 image, whereas the input data was obtained from the S1 data. Then we trained and tested all the three methods on the same training and test data. The remaining of this paper gives an overview of the materials and machine learning methods used (Sec. 2). Sec. 3 shows the results of the comparison study, and Sec. 4 concludes the findings.

2. MATERIALS AND METHODS

2.1. Data

2.1.1. Data for learning

We aim to estimate two quad-pol output parameters from a dual-pol input data. The output is denoted by y^i, for $i = 1, 2$, and the input by X. In this study, we will estimate the circular $RR - LL$ *coherence*, defined in Eq.(1), and the *co-pol ratio* defined in Eq. (2), below. In the X-Bragg surface scattering model, the circular $RR - LL$ coherence is directly related to surface roughness, whereas, the co-pol ratio in e.g. the Bragg

Radarsat-2 data were provided by KSAT under the Norwegian-Canadian Radarsat agreement 2016.

scattering model is related to the dielectric constant. The two outputs can be expressed by

$$y^1 = \langle |S_{RR}S_{LL}^{\star}| \rangle / \sqrt{\langle |S_{RR}|^2 \rangle \langle |S_{LL}|^2 \rangle} \quad (1)$$

$$y^2 = \langle |S_{HH}|^2 \rangle / \langle |S_{VV}|^2 \rangle \quad (2)$$

In order to avoid speckle noise, spatial averaging is used. This is indicated by $\langle \cdot \rangle$ in Eq. (1) and (2). The outputs y^1 and y^2 have been used to describe the geo-physical properties of sea ice surface [3], [4], which are surface roughness and dielectric parameters, respectively. The co-pol ratio is also sensitive to thin sea ice thickness [5].

The dual-pol input data is represented by the covariance matrix $\mathbf{C}$, and can be written by

$$\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} \\ C_{12}{}^{\star} & C_{22} \end{bmatrix}. \quad (3)$$

The dual-pol system C depends on the polarization of the transmitted and received signals. We use a system that transmits on the horizontal H channel and receives on H and vertical channel V, i.e.:

$$\mathbf{C} = \left\langle \begin{bmatrix} |S_{HH}|^2 & S_{HH}S_{HV}^{\star} \\ S_{HV}S_{HH}^{\star} & |S_{HV}|^2 \end{bmatrix} \right\rangle. \quad (4)$$

Using this input system C we generated five features for the input matrix $\mathbf{X}$. The five input features are denoted by $f_1, ..., f_5$, and can be written as

$$\begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \end{bmatrix} = \begin{bmatrix} 10\log_{10}(C_{11}) \\ 10\log_{10}(C_{22}) \\ C_{11}/C_{22} \\ C_{11} - C_{22} \\ (C_{11} - C_{22})/(C_{11} + C_{22}) \end{bmatrix}. \quad (5)$$

The input matrix $\mathbf{X}$ has a dimension of $5 \times N$, corresponding to the five input features, where N is the total number of pixels. Finally, the dataset for the learning can be expressed by $\mathcal{D} = \{\mathbf{y}^i, \mathbf{X}\}$, for $i = 1, 2$. The rationality for our approach is based on the anticipation that there is some functional relationships between the input features and output parameters, and that the ML models are able learn these in the training phase.

2.1.2. *Data for prediction*

We tested the prediction power of the machine learning approaches on an dual-pol ScanSAR S1. This can be seen in Fig. 1. The three machine learning models learn the relationship between the input systems obtained from the ScanSAR S1 data and the corresponding output features generated from the overlapping RS2 scene. The overlapping scenes are resampled, geocoded, and co-registered on a common grid with equal spatial resolution. This is used to predict the quad-pol features for the overlapping S1 scene. This way of predicting features that can only be estimated from a quad-pol system by using data from a dual-pol system, is what we refer to as *information up-scaling*.

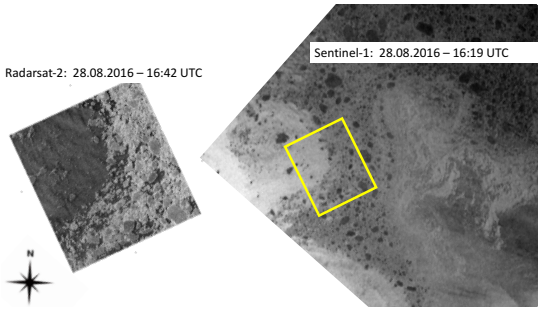

Fig. 1. Left is the HH intensity of RS2, and right is the HH intensity image of the S1 scene. The yellow box indicates the area of overlap.

2.2. Machine Learning Methods

2.2.1. *Neural Network*

The chosen NN was a Multi Layer Perceptron (MLP). The input to the MLP is the dual-pol input system, and the output is the corresponding quad-pol output features. The MLP has one hidden layer and 25 neurons. (It was found that this set up yielded robust and high accuracy results.) We used the Levenberg - Marquardt (L-M) training algorithm. (Note, we have also tested Bayesian regularization, however this has not impacted the results.) We used 70 % of the learning data for training, 15 % for testing and 15 % for validating, which is default in case of the NN.

The MLP operates by assigning weights (w) to the five input features, and then summing the products of the neurons in the hidden layer by a bias term (b). This sum goes through an activation function g to result in the output of the given node (n_i). This can be mathematically expressed be $n_i = g\big(\sum_{d=1}^{D} w_{di}\mathbf{x}_d + b_i\big)$, where d is the dimension of the input data and $i = 1, ..., 25$ is the number of neurons in the hidden layer. The final outputs of the hidden layer of the MLP NN was trained by using a back-propagation algorithm (L - M). In this work, the hyperbolic tangent sigmoid activation function was used.

2.2.2. *Support Vector Machine*

The SVM for regression estimates the output y_n from the input $\mathbf{x}_n$ by $y_n = \mathbf{w}^{\mathrm{T}}\mathbf{x}_n + b$, where the term $\mathbf{w}^{\mathrm{T}}$ is the transposed weight vector and b is the bias [6, 7, 8, 9]. An ϵ-intensitive loss function is used to obtain estimates, which are penalized by errors exceeding ϵ, while keeping the regression function flat. In order to obtain the estimates, the weights need to be computed. This is done by introducing an objective function $J = \frac{1}{\beta}\sum_{n=1}^{N}(\zeta_n^+ + \zeta_n^-) + \frac{1}{2}||\mathbf{w}||^2$, and minimizing it with respect to the weights and constrained to

$$y_n - \mathbf{w}^{\mathrm{T}}\mathbf{x}_n - b \leq \epsilon + \zeta_n^+ \quad \text{for} \quad n = 1, ..., N \quad (6)$$

$$\mathbf{w}^{\mathrm{T}}\mathbf{x}_n + b - y_n \leq \epsilon + \zeta_n^- \quad \text{for} \quad n = 1, ..., N \quad (7)$$

$$\zeta_n^+, \zeta_n^- \geq 0 \quad \text{for} \quad n = 1, ..., N. \quad (8)$$

The terms ζ_n^+ and ζ_n^- are the so-called slack variables. They ensure flexibility of the approach by allowing measurements to take larger values than ϵ, and $\beta > 0$. The weights are estimated by introducing support vectors, which are Lagrangian multipliers, and are obtained from the Lagrange function. This can be expressed by

$$\hat{\mathbf{w}} = \sum_{n=1}^{N} \left(\alpha_n^+ - \alpha_n^-\right) \mathbf{x}_n, \tag{9}$$

where the terms α_n^+ and α_n^- are the support vectors. Finally, introducing $a_n = \alpha_n^+ - \alpha_n^-$, an estimate for the output vector ($\hat{\mathbf{y}}$) can be computed by

$$\hat{\mathbf{y}} = \hat{\mathbf{w}}^{\mathrm{T}}\mathbf{x} + \hat{\mathbf{b}} = \sum_{n=1}^{N} a_n \mathbf{x}_n^{\mathrm{T}}\mathbf{x} + \hat{\mathbf{b}}. \tag{10}$$

We used the Squared Exponential (SE) kernel function to apply it to the term $\mathbf{x}_n^{\mathrm{T}}\mathbf{x}$, and hence the output parameters are estimated by

$$\hat{\mathbf{y}} = \sum_{n=1}^{N} a_n k(\mathbf{x}_n, \mathbf{x}) + \hat{\mathbf{b}}. \tag{11}$$

The SE kernel function can be written by

$$k(X_m, X_n) = \nu^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{D}\left(\frac{X_m^d - X_n^d}{\lambda_d}\right)^2\right), \tag{12}$$

where the hyper-parameters are λ_d and ν, and they are the length-scale and scaling-factor for the five dimension d.

2.2.3. Gaussian Process Regression

The GPR assumes that the output $\mathbf{y}^i$ is a function of the input $\mathbf{X}$, which can be expressed by $\mathbf{y} = f(\mathbf{X}) + \boldsymbol{\varepsilon}$. The term $\boldsymbol{\varepsilon}$ is the noise, which is independently, identically Gaussian distributed with zero mean and constant variance.

Fitting a multivariate joint Gaussian distribution over the function values allows to derive the posterior distribution of the output values [10]. For a new output $\mathbf{y}_*$, this can be written by

$$\begin{aligned} p(\mathbf{y}_*|\mathbf{X}_*, \mathcal{D}) &= \mathcal{N}(\mathbf{y}_*|\mu_{\mathrm{GP}*}, \sigma^2_{\mathrm{GP}*}) \\ \boldsymbol{\mu}_{\mathrm{GP}*} &= \mathbf{k}_{\mathbf{f}*}^{\top}(\mathbf{K}_{\mathbf{ff}} + \boldsymbol{\sigma}^2\mathbf{I}_n)^{-1}\mathbf{y} = \mathbf{k}_{\mathbf{f}*}^{\top}\boldsymbol{\alpha} \\ \boldsymbol{\sigma}_{\mathrm{GP}*} &= \sigma^2 + k_{**} - \mathbf{k}_{\mathbf{f}*}^{\top}(\mathbf{K}_{\mathbf{ff}} + \sigma^2\mathbf{I}_n)^{-1}\mathbf{k}_{\mathbf{f}*}, \end{aligned}$$

where $\boldsymbol{\mu}_{\mathrm{GP}*}$ is the estimated output, $\boldsymbol{\sigma}_{\mathrm{GP}*}$ is the certainty level of the estimates, $\mathbf{k}_{\mathbf{f}*}$ is the covariance between the training vector and the test point, $\boldsymbol{\alpha} = (\mathbf{K}_{\mathbf{ff}} + \sigma^2\mathbf{I}_n)^{-1}\mathbf{y}$ is the weight vector and k_{**} is the covariance between the test point with itself. The elements of the covariance matrices are computed by the SE kernel function, defined in Eq. (12).

Fig. 2 shows the workflow of the different steps of the three models.

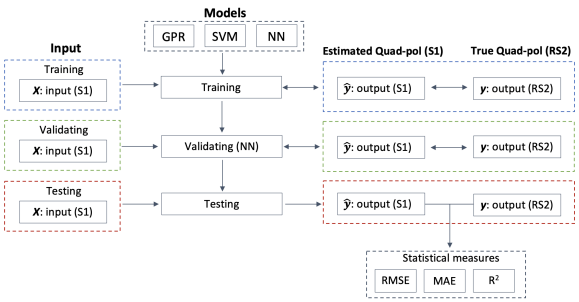

Fig. 2. Flowchart of the different steps involving the data input, the three models tested, training, testing, and validation.

Table 1. Computed statistical measures for the machine learning models for the overlapping S1 scene.

	y^1			y^2		
Model	NRMSE	MAE	R^2	NRMSE	MAE	R^2
SVM	0.0326	0.0201	0.9936	0.1061	0.2009	0.8450
GPR	0.0300	0.0185	0.9946	0.0919	0.1579	0.8827
NN	0.0307	0.0188	0.9943	0.0924	0.1654	0.8816

3. RESULTS

Fig. 3 shows the results of the estimated outputs, y^1 (left-column) and y^2 (right-column). The top-row is the ground truth, which is the generated output data form the RS2 quad-pol system, the second-, third- and bottom rows show the estimated outputs from the S1 input system for the SVM, GPR and NN models, respectively. We used 494 randomly sampled observations for training the machine learning methods, and for testing and prediction the whole RS2 and S1 scenes were used, respectively. This corresponds to $4.9{\cdot}10^5$ pixels (observations). Fig. 3 shows that although all the thee machine learning methods seem to follow the pattern of the ground truth, there are differences in the assigned values between the methods. The SVM seems to show underestimates in both outputs. Both the NN and GPR show correct estimates for y^1 and y^2. In case of the output y^1, the GPR model might show some smoothing in comparison to the NN. Note, this can be adjusted by setting the length-scale parameters in the kernel function. The computed statistical measures (Table 1) also confirms these results. We used the Normalized Root Mean Squared Errors (NRMSE), the Mean Absolute Error (MAE) and the squared Pearson correlation coefficient for comparing the performance of the three machine learning methods. Although both the NN and GPR performed well for both outputs, the GPR showed slightly lower NRMSE and MAE, and higher R^2 than the NN.

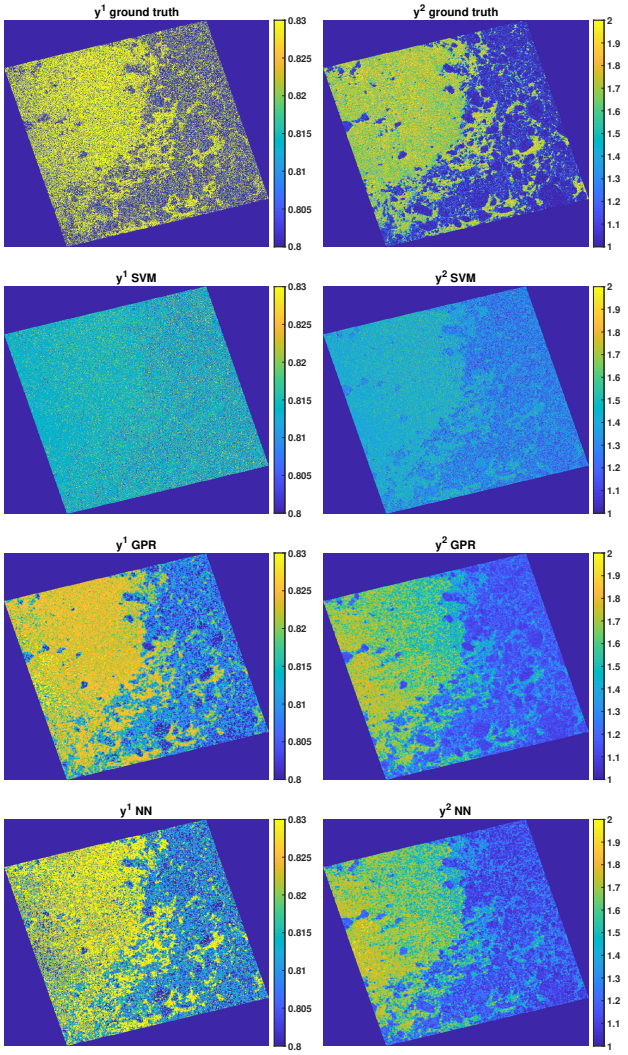

Fig. 3. The output features y^1 (left column) and y^2 (right-column) for the ground truth, the SVM, GPR and NN models. The number of samples used for training is 494 and for testing is $4.9{\cdot}10^5$.

4. CONCLUSION

In this study, we compared three well-recognized machine learning approaches for spatial up-scaling of quad-pol information using dual-pol observations. Our results indicate that the NN and GPR models are excellent candidates to establish a functional relationship between the quad-pol outputs and the dual-pol input system. We evaluated the three methods for two outputs describing sea ice surface characteristics. The NN and the GPR could capture the smoother (yellow in Fig. 3) and rougher surfaces (blue in Fig. 3) according to the generated ground truth RS2 data. This was in contrast with the SVM model, which showed in general underestimates for both y^1 and y^2. This was also supported by the computed statistical measures (Table 1), which revealed that both the NN and GPR models performed well for this task. This comparison study showed, that based on the computed measures, the strongest performance was obtained by using the GPR model. Both the NN and GPR models have different advantageous properties. Both models can be applied to large datasets, hence they can be used for operational purposes. Therefore, for future work, we will continue to further exploit the possibilities of using NN, in addition to the GPR for for up-scaling.

5. ACKNOWLEDGMENTS

This research is financed by CIRFA (RCN Grant no. 237906).

6. REFERENCES

[1] S. V. Nghiem, R. Kwok, S. H. Yueh, and M. R. Drinkwater, "Polarimetric signatures of sea ice: 1. theoretical model," *J. Geophys. Res.,*, vol. 100(C7), pp. 13 681–13 698, 1995.

[2] K. Blix, M. M. Espeseth, and T. Eltoft, "Up-scaling from quad-polarimetric to dual-polarimetric sar data using machine learning gaussian process regression," in *IEEE Int. Geosci. and Remote Sens. Symp., IGARSS, Valencia, Spain*, 2018.

[3] F. Mattia, T. L. Toan, J.-C. Souyris, G. D. Carolis, N. Floury, F. Posa, and G. Pasquariello, "The effect of surface roughness on multifrequency polarimetric sar data," *IEEE Trans. Geosci. Remote Sens*, vol. 35(4), pp. 954–996, 1997.

[4] R. Ressel and S. Singha, "Comparing near coincident space borne c and x band fully polarimetric sar data for arctic sea ice classification," *Remote Sensing*, vol. 8(3), p. 198, 2016.

[5] R. G. Onstott, *SAR and Scatterometer Signatures of Sea Ice*. American Geophysical Union (AGU), 2013, ch. 5, pp. 73–104.

[6] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, 2004.

[7] B. Schölkopf and A. Smola, "Learning with kernels-support vector machines, regularization, optimization and beyond," *MIT Press*, 2002.

[8] K. P. Murphy, *Machine Learning A probabilistic Perspective*. The MIT Press, 2012.

[9] S. Y. Kung, *Kernel Methods and Machine Learning*. Cambridge University Press, 2014.

[10] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2005.