



Looking Outside the Box: The Role of Context in Random Forest Based Semantic Segmentation of PolSAR Images

Ronny Hänsch^(✉)

SAR Technology, German Aerospace Center (DLR), Cologne, Germany
rww.haensch@gmail.com

Abstract. Context - i.e. information not contained in a particular measurement but in its spatial proximity - plays a vital role in the analysis of images in general and in the semantic segmentation of Polarimetric Synthetic Aperture Radar (PolSAR) images in particular. Nevertheless, a detailed study on whether context should be incorporated implicitly (e.g. by spatial features) or explicitly (by exploiting classifiers tailored towards image analysis) and to which degree contextual information has a positive influence on the final classification result is missing in the literature. In this paper we close this gap by using projection-based Random Forests that allow to use various degrees of local context without changing the overall properties of the classifier (i.e. its capacity). Results on two PolSAR data sets - one airborne over a rural area, one space-borne over a dense urban area - show that local context indeed has substantial influence on the achieved accuracy by reducing label noise and resolving ambiguities. However, increasing access to local context beyond a certain amount has a negative effect on the obtained semantic maps.

1 Introduction

Context refers to information not contained in an individual measurement but in its local proximity or at a larger (even global) range. For image analysis, this can refer to a spatial (i.e. pixels close to each other), temporal (measurements with a small time difference), or spectral (measurements taken at similar wavelengths) neighborhood. In this paper, context refers to the *spatial* neighborhood of a pixel.

In contrast to the (semantic) analysis of close-range photography, for a long time context had played only a minor role in remote sensing, in particular for data sources such as HyperSpectral Imagery (HSI) or Synthetic Aperture Radar (SAR). One reason is the historical approach and the scientific communities that pioneered in the analysis of images from both domains. The similarity of color photographs to the early stages of the human visual cortex (e.g. being based on angular measurements of the light intensity of primary colors), inspired to model also subsequent stages according to this biological role model for which it is well known that context (spatial as well as temporal) plays a vital role for the understanding of the image input [21]. HSI and SAR images, on the other hand, are too dissimilar to human perception to have inspired a similar

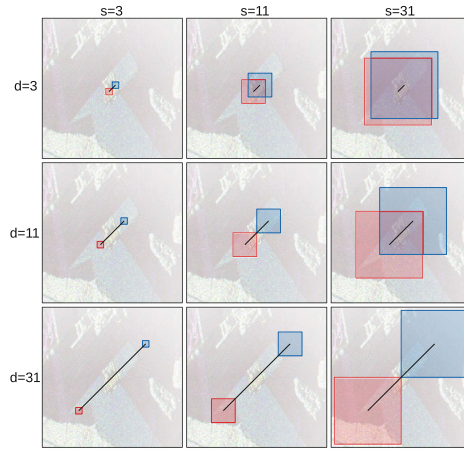


Fig. 1. We investigate the role of (visual spatial) context by varying the size of the spatial projections within the framework of projection-based Random Forests (pRFs), i.e. the size r_s and distance r_d of regions sampled relative to the patch center and used by the internal node tests of the decision trees to determine the semantic class.

approach during the early years of automated image analysis. On the contrary, early attempts to remote sensing image interpretation were often carried out by the same groups that built the corresponding sensors. Consequently, they took a rather physics-based approach and developed statistical models that aim to capture the complex relations between geo-physical and biochemical properties of the imaged object and the measured signal. Even today, approaches that aim to model the interaction of electro-magnetic waves with a scatterer with certain geometric and electro-physical properties are still in use for SAR image processing (see e.g. [7, 10]). Another reason is that the information contained in a single RGB pixel of a close-range photograph is rarely sufficient to make any reliable prediction of the semantic class this pixel might belong to. On the other hand, the information contained in a single HSI or PolSAR pixel does allow to make such predictions with a surprisingly high accuracy if processed and analysed correctly.

As a consequence, although there were early attempts to incorporate context (see e.g. [24, 28]) into the semantic analysis of remote sensing images, many classification methods ignored relations between spatially adjacent pixels and process each pixel independently (e.g. as in [6] for HSI and [16] for SAR data, respectively). This means in particular, that a random permutation of all pixels within the image would not effect classification performance during automatic image interpretation (quite in contrast to a visual interpretation by humans). However, neighboring pixels do contain a significant amount of information which should be exploited. On the one hand, adjacent pixels are usually correlated due to the image formation process. On the other hand, the depicted objects are usually large (with respect to the pixel size) and often rather homogeneous.

There are two distinct yet related concepts of context in images, i.e. visual context and semantic context. **Semantic context** refers to relationships on object level such as co-occurrence relations (e.g. a ship usually occurs together with water) for example modelled via Latent Dirichlet Allocations [23] or concept occurrence vectors [27] and topological relations (e.g. trees are more likely to be next to a road than on a road) capturing distances and directions (see e.g. [2]). This type of context is usually exploited during the formulation of the final decision rule, e.g. by applying a context-independent pixel-wise classification followed by a spatial regularization of the obtained semantic maps [5] or by applying Markov Random Fields (MRFs, see e.g. [11, 25] for usage of MRFs for the classification of SAR images). **Visual context** refers to relationships on the measurement level allowing for example to reduce the noise of an individual measurement (e.g. by local averaging) or to estimate textural properties. For example, visual context is implicitly considered during SAR speckle filtering. Another common example are approaches that combine spectral and spatial information in a pixel-wise feature vector and then apply pixel-based classification methods (e.g. [9, 22]). More recent approaches move away from the use of predefined hand-crafted features and use either variants of shallow learners that have been tailored towards the analysis of image data (such as projection-based Random Forests [12]) or deep neural networks. In particular the latter have gained on importance and are often the method of choice for the (semantic) analysis of remote sensing images in general (see e.g. [15] for an overview) and SAR data in particular [31].

In this paper we address the latter type, i.e. visual context, for the special case of semantic segmentation on polarimetric SAR images. In particular, we are interested whether different data representations that implicitly integrate context are helpful and in analysing how much local context is required or sufficient to achieve accurate and robust classification results. To the best of the authors knowledge, such an investigation is missing in the current literature of PolSAR processing. Corresponding works either stop at low-level pre-processing steps such as speckle reduction [4, 8] or simply assume that any amount of available contextual information leads to an improved performance.

Mostly to be able to efficiently vary available context information while keeping model capacity fixed, we use projection-based Random Forests (pRFs, [12]) which are applied to image patches and apply spatial projections (illustrated in Fig. 1) that sample regions of a certain size and distance to each other. Increasing the region size allows to integrate information over larger areas and thus adaptively reduce noise, while a larger region distance enables the RF to access information that is further away from the patch center without increasing the computational load (very similar to dilated convolutions in convolution networks [30]). Thus, the contribution of this paper is three-fold: First, we extend the general framework of [12] to incorporate node tests that can be directly applied to polarimetric scattering vectors; Second, we compare the benefits and limitations of using either scattering vectors or polarimetric sample covariance matrices for the semantic segmentation of PolSAR images; and third, we analyse how much context information is helpful to increase classification performance.

2 Projection-Based Random Forests

Traditional machine-learning approaches for semantic segmentation of PolSAR images either rely on probabilistic models aiming to capture the statistical characteristics of the scattering processes (e.g. [3, 29]) or apply a processing chain that consists of pre-processing, extracting hand-crafted features, and estimating a mapping from the feature space to the desired target space by a suitable classifier (e.g. [1, 26]). Modern Deep Learning approaches offer the possibility to avoid the computation of hand-crafted features by including feature extraction into the optimization of the classifier itself (see e.g. [17–20]). These networks are designed to take context into account by using units that integrate information over a local neighborhood (their receptive field). In principle, this would allow to study the role of context for the semantic segmentation of remotely sensed images with such networks. However, an increased receptive field usually corresponds to an increase of internal parameters (either due to larger kernels or deeper networks) and thus an increased capacity of the classifier.

This is why we apply projection-based Random Forests (pRFs [12]) which offer several advantages for the following experiments: Similar to deep learning approaches, pRFs learn features directly from the data and do not rely on hand-crafted features. Furthermore, they can be applied to various input data without any changes to the overall framework. This allows us to perform experiments on PolSAR data which are either represented through polarimetric scattering vectors $\mathbf{s} \in \mathbb{C}^k$ or polarimetric sample covariance matrices $\mathbf{C} \in \mathbb{C}^{k \times k}$

$$\mathbf{C} = \langle \mathbf{s}\mathbf{s}^\dagger \rangle_{w_C} \quad (1)$$

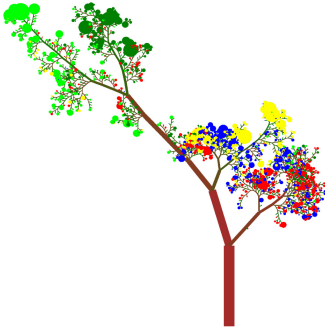
where $(\cdot)^\dagger$ denotes conjugate transpose and $\langle \cdot \rangle_{w_C}$ a spatial average over a $w_C \times w_C$ neighborhood.

Every internal node of a tree (an example of such a tree is shown in Fig. 2(a)) in a RF performs a binary test $t : D \rightarrow \{0, 1\}$ on a sample $\mathbf{x} \in D$ that has reached this particular node and propagates it either to the left ($t(\mathbf{x}) = 0$) or right child node ($t(\mathbf{x}) = 1$). The RF in [12] defines the test t as

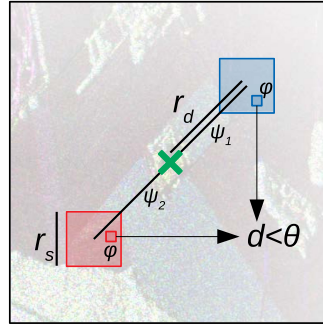
$$t(x) = \begin{cases} 0 & \text{if } d(\phi(\psi_1(\mathbf{x})), \phi(\psi_2(\mathbf{x}))) < \theta, \\ 1 & \text{otherwise.} \end{cases} \quad (2)$$

where $\psi(\cdot)$ samples a region from within a patch that has a certain size r_s and distance r_d to the patch center, $\phi(\cdot)$ selects a pixel within this region, $d(\cdot)$ is a distance function, and θ is the split threshold (see Fig. 2(b) for an illustration). Region size r_s and distance r_d to the patch center are randomly sampled from a user defined range. They define the maximal possible patch size $w = 2r_d + r_s$ and thus the amount of local context that can be exploited by the test. To test whether a multi-scale approach is beneficial for classification performance, we allow the region distance to be scaled by a factor α which is randomly drawn by a user defined set of possible scales.

The pixel selection function ϕ as well as the distance function are data type dependent. The RF in [12] proposes test functions that apply to $w \times w$ patches



(a) A typical decision tree within a pRF (visualized via [14]) that was trained on the *OPH* data set (see Section 3.1). Leaf colors represent the dominant class in a leaf (see Figure 3(c)); leaf size represents the number of samples in this leaf.



(b) A node test t uses spatial projections of a patch centered at \mathbf{x} that sample regions of size r_s and distance r_d to the patch center via $\psi(\cdot)$ from which $\phi(\cdot)$ selects a single pixel. The distance d between these two pixel values is then compared to the split threshold θ .

Fig. 2. Visualisation of a single decision tree of a trained pRF (left) as well as the applied spatial node projections (right).

of polarimetric covariance matrices, (i.e. $D = \mathbb{C}^{w \times w \times k \times k}$). In this case, ϕ either computes the average over the region or selects the covariance matrix within a given region with minimal, maximal, or medium span r_s , polarimetric entropy H , or anisotropy A , i.e.

$$S = \sum_{i=1}^k \lambda_i, \quad H = \sum_{i=1}^k \frac{\lambda_i}{S} \log \left(\frac{\lambda_i}{S} \right), \quad A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3} \quad (3)$$

where $\lambda_1 > \lambda_2 > \lambda_3$ are the Eigenvalues of the covariance matrix. Note, that for $k = 2$, i.e. dual-polarimetric data, the covariance matrix has only two Eigenvalues which means that the polarimetric anisotropy cannot be computed.

Any measure of similarity between two Hermitian matrices P, Q (see [13] for an overview) can serve as distance function d , e.g. the Bartlett distance

$$d(P, Q) = \ln \left(\frac{|P + Q|^2}{|P||Q|} \right). \quad (4)$$

We extend this concept to polarimetric scattering vectors $\mathbf{s} \in \mathbb{C}^k$ by adjusting ϕ to select pixels with minimal, maximal, or medium total target power ($\sum_i |s_i|$). Note that polarimetric scattering vectors are usually assumed to follow a complex Gaussian distribution with zero mean which means that the local sample average tends to approach zero and thus does not provide a reasonable projection. While it would be possible to use polarimetric amplitudes only, we

want to work as closely to the data as possible. Extracting predefined features and using corresponding projections is possible within the pRF framework but beyond the scope of the paper. As distance $d(p, q)$ we use one of the following distance measures between polarimetric scattering vectors $p, q \in \mathbb{C}^k$:

$$\text{Span distance: } d(p, q) = \sum_{i=1}^k |p_i| - \sum_{i=1}^k |q_i| \quad (5)$$

$$\text{Channel intensity distance: } d(p, q) = |p_i| - |q_i| \quad (6)$$

$$\text{Phase difference: } d(p, q) = \arg(p_i) - \arg(q_i) \quad (7)$$

$$\text{Ratio distance: } d(p, q) = \left| \log \left(\frac{|p_i|}{|p_j|} \right) \right| - \left| \log \left(\frac{|q_i|}{|q_j|} \right) \right| \quad (8)$$

$$\text{Euclidean distance: } d(p, q) = \sqrt{\sum_{i=1}^k |p_i - q_i|^2}, \quad (9)$$

where $\arg(z)$ denotes the phase of z .

An internal node creates multiple such test functions by randomly sampling their parameters (i.e. which ψ defined by region size and position, which ϕ , and which distance function d including which channel for channel-wise distances) and selects the test that maximises the information gain (i.e. maximal drop of class impurity in the child nodes).

3 Experiments

3.1 Data

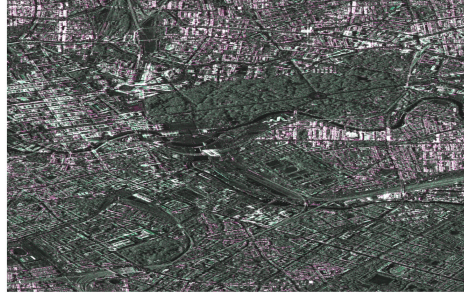
We use two very different data sets to evaluate the role of context on the semantic segmentation of PolSAR images. The first data set (shown in Fig. 3(a), 3(c)) is a fully polarimetric SAR image acquired over Oberpfaffenhofen, Germany, by the E-SAR sensor (DLR, L-band). It has 1390×6640 pixels with a resolution of approximately 1.5 m. The scene contains rather large homogeneous object regions. Five different classes have been manually marked, namely City (red), Road (blue), Forest (dark green), Shrubland (light green), and Field (yellow).

The second data set (shown in Fig. 3(b)) is a dual-polarimetric image of size 6240×3953 acquired over central Berlin, Germany, by TerraSAR-X (DLR, X-band, spotlight mode). It has a resolution of approximately 1 m. The scene contains a dense urban area and was manually labelled into six different categories, namely Building (red), Road (cyan), Railway (yellow), Forest (dark green), Lawn (light green), and Water (blue) (see Fig. 3(d)).

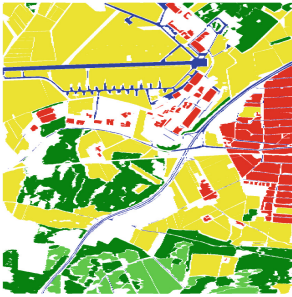
The results shown in the following sections are obtained by dividing the individual image into five vertical stripes. Training data (i.e. 50,000 pixels) are drawn by stratified random sampling from four stripes, while the remaining stripe is used for testing only. We use Cohen's κ coefficient estimated from the test data and averaged over all five folds as performance measure.



(a) *OPH* image data
(E-SAR, DLR, L-Band)



(b) *BLN* image data
(TerraSAR-X, DLR, X-Band)



(c) *OPH* reference data:
City (red), Road (blue),
Forest (dark green), Shrub-
land (light green), Field
(yellow), unlabelled pixels
in white



(d) *BLN* reference data:
Building (red), Road (cyan), Railway (yel-
low), Forest (dark green), Lawn (light
green), Water (blue), unlabelled pixels in
white.

Fig. 3. False color composite of the used PolSAR data (top) as well as color-coded reference maps (bottom) of the Oberpfaffenhofen (*OPH*, left) and Berlin (*BLN*, right) data sets. Note: Images have been scaled for better visibility. (Color figure online)

3.2 Polarimetric Scattering Vectors

As a first step we work directly on the polarimetric scattering vectors by using the projections described in Sect. 2 with $r_d, r_s \in \{3, 11, 31, 101\}$. Figure 4 shows the results when using the polarimetric scattering vectors directly without any pre-processing (i.e. no presumming, no speckle reduction, etc.). The absolute accuracy (in terms of the kappa coefficient) differs between the air- ($\kappa \in [0.64, 0.80]$) and space-borne ($\kappa \in [0.29, 0.44]$) PolSAR data. There are several reasons for this difference. One the one hand, the *OPH* data was acquired by an fully-polarimetric airborne sensor while the *BLN* data was acquired by a dual-polarimetric space-borne sensor. As a consequence, the *OPH* data contains more information (one more polarimetric channel) and has in general a better signal to noise ratio. On the other hand, the scene is simpler in terms of semantic classes, i.e. the

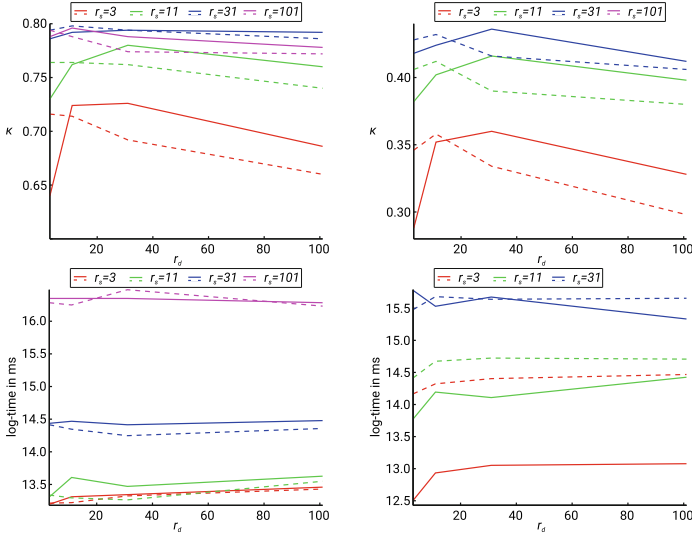


Fig. 4. Achieved κ (top) and prediction time (bottom) for *OPH* (left) and *BLN* (right) using polarimetric scattering vectors. The solid lines denote the single scale ($\alpha = 1$), the dashed lines the multi-scale ($\alpha \in \{1, 2, 5, 10\}$) case.

reference data contains less classes and object instances are rather large, homogeneous segments. In contrast, the *BLN* data contains fine grained object classes such as buildings and roads in a dense urban area.

Despite the difference in the absolute values for both data sets, the relative performance between the different parameter settings is very similar. In general, larger region sizes lead to a better performance. While the difference between 3×3 and 11×11 regions are considerable, differences between 11×11 and 31×31 regions are significantly smaller. Large regions of 101×101 pixels lead to worse results than moderate regions of 31×31 . Larger regions allow to locally suppress speckle and noise and are better able to integrate local context. However, beyond a certain region size, the patches start to span over multiple object instances which makes it impossible to distinguish between the different classes.

A similar although less pronounced effect can be seen for increasing region distances. At first, performance does increase with larger distance. However, the improvement soon saturates and for very large distances even deteriorates. This effect is strongest in combination with small region sizes as the distance relative to the region size is much smaller for tests with large regions, i.e. for a test with a region distance of $r_d = 11$, regions of $r_s = 31$ still overlap.

The optimal parameter combination in terms of accuracy is $r_s = r_d = 31$, i.e. patches with $w = 93$ (note, that this only determines the maximal patch size while the actually used size depends on the specific tests selected during node optimisation). Interestingly, this seems to be independent of the data set.

A large region size has the disadvantage of an increased run time during training and prediction (the latter is shown in Fig. 4). The run time per node

test increases quadratically with the region size r_s but is independent of r_d . The overall run time also depends on the average path length within the trees which might in- or decrease depending on the test quality (i.e. whether a test is able to produce a balanced split of the data with a high information gain). In general, an increased region size leads to a much longer prediction time, while an increased region distance has only a minor effect. As a consequence, if computation speed is of importance in a particular application, it is recommendable to increase sensitivity to context by setting a larger region distance than increasing the region size (at the cost of a usually minor loss in accuracy).

The dashed lines in Fig. 4 show the results when access to context is increased beyond the current local region by scaling the region distance by a factor α which is randomly selected from the set $R = \{1, 2, 5, 10\}$ (e.g. if r_d is originally selected as $r_d = 5$ and α is selected as $\alpha = 10$, the actually used region distance is 50). If the original region distance is set to a small value (i.e. $r_d = 3$) using the multi-scale approach leads to an increased performance for all region sizes. For a large region size of $r_s = 101$ this increase is marginal, but for $r_s = 3$ the increase is substantial (e.g. from $\kappa = 0.64$ to 0.72 for *OPH*). However, even for medium region distances ($r_d = 11$) the effect is already marginal and for large distances the performance actually decreases drastically. The prediction time is barely affected by re-scaling the region distance. In general, this reconfirms the results of the earlier experiments (a too large region distance leads inferior results) and shows that (at least for the used data sets) local context is useful to solve ambiguities in the classification decision, but global context does rarely bring further benefits. On the one hand, this is because local homogeneity is a very dominant factor within remote sensing images, i.e. if the majority of pixels in a local neighborhood around a pixel belong to a certain class, the probability is high that this pixel belongs to the same class. On the other hand, typical objects in remote sensing images (i.e. such as the here investigated land cover/use classes) are less constrained in their spatial co-occurrence than close range objects (e.g. a road can go through an urban area, through agricultural fields as well as through forest or shrubland and can even run next to a river).

3.3 Estimation of Polarimetric Sample Covariance Matrices

In a second experiment, we use the projections described in Sect. 2, i.e. the RF is applied to polarimetric sample covariance matrices instead of scattering vectors. While in contrast to scattering vectors, covariance matrices can be locally averaged, we exclude node tests that perform local averaging in order to be better comparable to the experiments on scattering vectors.

As covariance matrices are computed by locally averaging the outer product of scattering vectors, they implicitly exploit context. In particular distributed targets can be statistically described only by their second moments. Another effect is that large local windows increases the quality of the estimate considerably. However, too large local windows will soon go beyond object borders and include pixels that belong to a different physical process, i.e. in the worst case to a different semantic class, reducing the inter-class variance of the samples.

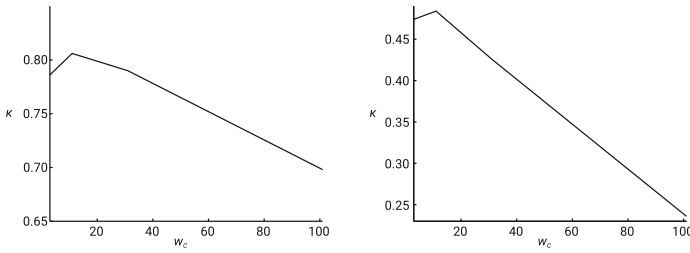


Fig. 5. Achieved κ (top) and prediction time (bottom) for *OPH* (left) and *BLN* (right) using covariance matrices computed over local windows of size w_C .

Figure 5 shows that performance barely changes for medium window sizes but degrades drastically for larger windows. A reasonable choice is $w_C = 11$, which is used in the following experiments. Note that covariance matrices are precomputed and thus do not influence computation times of the classifier.

3.4 Polarimetric Sample Covariance Matrices

In the last set of experiments, we fix the local window for computing the local polarimetric covariance matrix to $w_C = 11$ and vary region distance r_d and size r_s in the same range as for the experiments based on the scattering vector, i.e. $r_d, r_s \in \{3, 11, 31, 101\}$. The results are shown in Fig. 6. Compared to using scattering vectors directly, the achieved performance increased from $\kappa \in [0.64, 0.798]$ to $\kappa \in [0.786, 0.85]$ for *OPH* and from $\kappa \in [0.288, 0.436]$ to $\kappa \in [0.448, 0.508]$ for *BLN* which demonstrates the benefits of speckle reduction and the importance to use second-order moments. The relative performance among different settings for region size and distance, however, stays similar. Large regions perform in general better than small regions. An interesting exception can be observed for $r_s = 3$ and $r_s = 11$: While for small distances ($d \leq 11$) the larger $r_s = 11$ leads to better results, the accuracy for $r_s = 3$ surpasses the one for $r_s = 11$ if $r_d = 31$. In general the results follow the trend of the experiments based on scattering vectors: First, the performance increases with increasing distance, but then declines if the region distance is too large. This is confirmed as well by the experiments with upscaled distances: While for $r_d = 3$ the results of the scaled distance is often superior to the results achieved using the original distance, the performance quickly decreases for $d > 11$.

3.5 Summary

Figure 7 shows qualitative results by using projections that allow 1) a minimal amount of context (being based on scattering vectors with $r_d = r_s = 3$ and no scaling), 2) the optimal (i.e. best κ in the experiments) amount of context (being based on covariance matrices with $r_d = r_s = 31$ and no scaling); and 3) a large amount of context (being based on covariance matrices with $r_d = 101$, $r_s = 31$ and scaling with $\alpha \in \{1, 2, 5, 10\}$). There is a significant amount of label noise if only a small amount of local context is included but even larger structures tend to be misclassified if they are locally similar to other classes. By increasing the amount of context, the obtained semantic maps become considerably smoother. Note, that these results are obtained without any post-processing. Too much context, however, degrades the results as the inter-class differences decrease leading to misclassifications in particular for smaller structures.

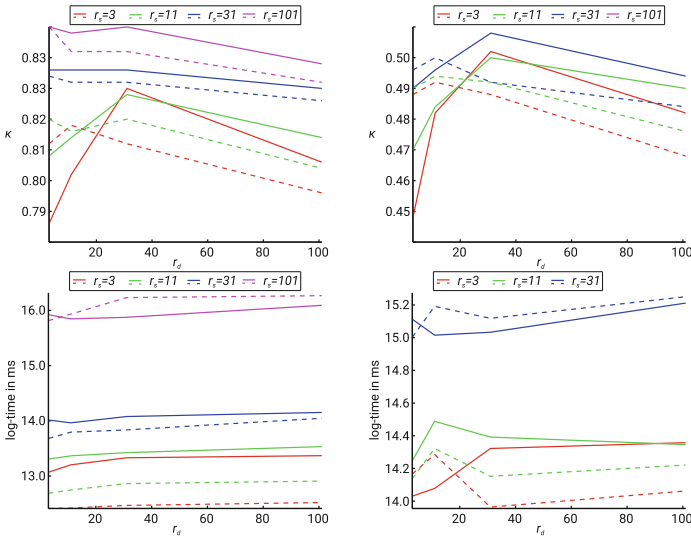
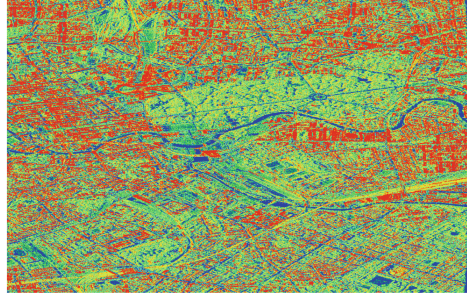
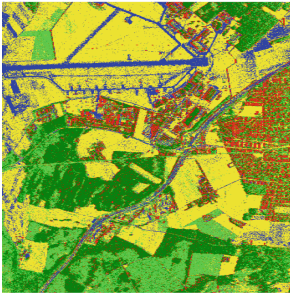
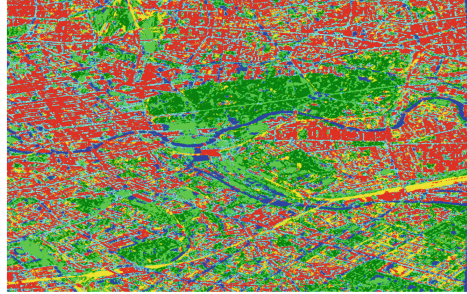
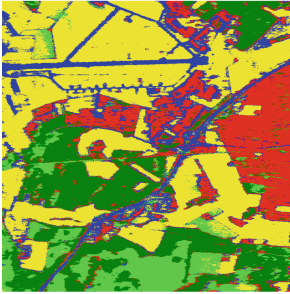


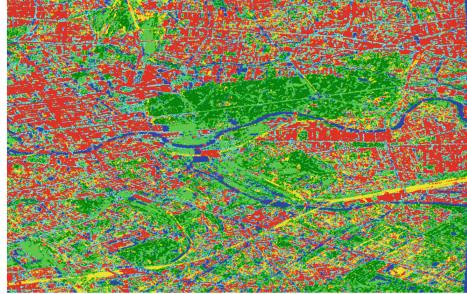
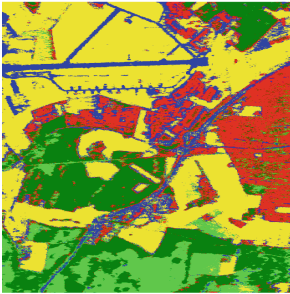
Fig. 6. Achieved κ (top) and prediction time (bottom) for *OPH* (left) and *BLN* (right) using polarimetric sample covariance matrices. The solid lines denote the single scale ($\alpha = 1$), the dashed lines the multi-scale ($\alpha \in \{1, 2, 5, 10\}$) case.



(a) Minimal context: Scattering vector, $r_d = r_s = 3$, $\alpha = 1$



(b) Optimal context: Covariance matrix ($w_C = 11$), $r_d = r_s = 31$, $\alpha = 1$



(c) Maximal context: Covariance matrix ($w_C = 11$), $r_d = 31$, $r_s = 31$, $\alpha \in \{1, 2, 5, 10\}$

Fig. 7. Obtained semantic maps (stitching of corresponding test sets) by exploiting different amounts of spatial context. Note: Images have been scaled for better visibility.

4 Conclusion and Future Work

This paper extended the set of possible spatial projections of pRFs by exploiting distance functions defined over polarimetric scattering vectors. This allows a time- and memory efficient application of pRFs directly to PolSAR images without any kind of preprocessing. However, the experimental results have shown that usually a better performance (in terms of accuracy) can be obtained by

using polarimetric sample covariance matrices. We investigated the influence of the size of the spatial neighborhood over which these matrices are computed and showed that medium sized neighborhoods lead to best results where the relative performance changes were surprisingly consistent between two very different data sets. Last but not least we investigated the role context plays by varying the region size and distance of the internal node projections of pRFs. Results show that the usage of context is indeed essential to improve classification results but only to a certain extent after which performance actually drastically decreases.

Future work will confirm these findings for different sensors, i.e. HSI and optical images, as well as for different classification tasks. Furthermore, while this paper focused on visual context (i.e. on the measurement level), semantic context (i.e. on the level of the target variable) is of interest as well. On the one hand, the test selection of the internal nodes of pRFs allows in principle to take semantic context into account during the optimisation process. On the other hand, post processing steps such as MRFs, label relaxation, or stacked Random Forests should have a positive influence on the quality of the final semantic maps.

References

1. Aghababae, H., Sahebi, M.R.: Game theoretic classification of polarimetric SAR images. *Eur. J. Remote Sens.* **48**(1), 33–48 (2015)
2. Aksoy, S., Koperski, K., Tusk, C., Marchisio, G., Tilton, J.C.: Learning Bayesian classifiers for scene classification with a visual grammar. *IEEE Trans. Geosci. Remote Sens.* **43**(3), 581–589 (2005)
3. Anfinson, S.N., Eltoft, T.: Application of the matrix-variate Mellin transform to analysis of polarimetric radar images. *IEEE Trans. Geosci. Remote Sens.* **49**(6), 2281–2295 (2011)
4. Bouchemakh, L., Smara, Y., Boutarfa, S., Hamadache, Z.: A comparative study of speckle filtering in polarimetric radar SAR images. In: 2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications, pp. 1–6 (2008)
5. Bovolo, F., Bruzzone, L.: A context-sensitive technique based on support vector machines for image classification. In: Pal, S.K., Bandyopadhyay, S., Biswas, S. (eds.) *Pattern Recogn. Mach. Intell.*, pp. 260–265. Springer, Heidelberg (2005). https://doi.org/10.1007/11590316_36
6. Camps-Valls, G., Bruzzone, L.: Kernel-based methods for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **43**(6), 1351–1362 (2005)
7. Deng, X., López-Martínez, C., Chen, J., Han, P.: Statistical modeling of polarimetric SAR data: a survey and challenges. *Remote Sens.* **9**(4) (2017). <https://doi.org/10.3390/rs9040348>
8. Farhadiani, R., Homayouni, S., Safari, A.: Impact of polarimetric SAR speckle reduction on classification of agriculture lands. *ISPRS - Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci. XLII-4/W18*, 379–385 (2019). <https://doi.org/10.5194/isprs-archives-XLII-4-W18-379-2019>
9. Fauvel, M., Chanussot, J., Benediktsson, J.A., Sveinsson, J.R.: Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. In: 2007 IEEE International Geoscience and Remote Sensing Symposium, pp. 4834–4837 (2007)

10. Fischer, G., Papathanassiou, K.P., Hajnsek, I.: Modeling and compensation of the penetration bias in InSAR DEMs of Ice sheets at different frequencies. *IEEE J. Selected Topics Appl. Earth Observ. Remote Sens.* **13**, 2698–2707 (2020)
11. Fjortoft, R., Delignon, Y., Pieczynski, W., Sigelle, M., Tupin, F.: Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **41**(3), 675–686 (2003)
12. Hänsch, R., Hellwich, O.: Skipping the real world: classification of PolSAR images without explicit feature extraction. *ISPRS J. Photogrammetry Remote Sens.* **140**, 122–132 (2017). <https://doi.org/10.1016/j.isprsjprs.2017.11.022>
13. Hänsch, R., Hellwich, O.: A comparative evaluation of polarimetric distance measures within the random forest framework for the classification of PolSAR images. In: *IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 8440–8443. IEEE, July 2018. <https://doi.org/10.1109/IGARSS.2018.8518834>
14. Hänsch, R., Wiesner, P., Wendler, S., Hellwich, O.: Colorful trees: visualizing random forests for analysis and interpretation. In: *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 294–302. IEEE, January 2019. <https://doi.org/10.1109/WACV.2019.00037>
15. Hoerer, T., Kuenzer, C.: Object detection and image segmentation with deep learning on earth observation data: a review-part I: evolution and recent trends. *Remote Sens.* **12**, 1667 (2020)
16. Jong-Sen Lee, Grunes, M.R., Pottier, E., Ferro-Famil, L.: Unsupervised terrain classification preserving polarimetric scattering characteristics. *IEEE Trans. Geosci. Remote Sens.* **42**(4), 722–731 (2004)
17. Ley, A., D’Hondt, O., Valade, S., Hänsch, R., Hellwich, O.: Exploiting GAN-based SAR to optical image transcoding for improved classification via deep learning. In: *EUSAR 2018; 12th European Conference on Synthetic Aperture Radar*, pp. 396–401. VDE, June 2018
18. Liu, X., Jiao, L., Tang, X., Sun, Q., Zhang, D.: Polarimetric convolutional network for polSAR image classification. *IEEE Trans. Geosci. Remote Sens.* **57**(5), 3040–3054 (2019)
19. Mohammadimaneh, F., Salehi, B., Mahdianpari, M., Gill, E., Molinier, M.: A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem. *ISPRS J. Photogramm. Remote. Sens.* **151**, 223–236 (2019)
20. Mullissa, A., Persello, C., Stein, A.: Polsarnet: a deep fully convolutional network for polarimetric sar image classification. *IEEE J. Selected Topics Appl. Earth Observ. Remote Sens.* **12**(12), 5300–5309 (2019)
21. Paradiso, M.A., Blau, S., Huang, X., MacEvoy, S.P., Rossi, A.F., Shalev, G.: Lightness, filling-in, and the fundamental role of context in visual perception. In: *Visual Perception, Progress in Brain Research*, vol. 155, pp. 109–123. Elsevier (2006)
22. Pesaresi, M., Benediktsson, J.A.: A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Trans. Geosci. Remote Sens.* **39**(2), 309–320 (2001)
23. Tang, H.H., et al.: A multiscale latent Dirichlet allocation model for object-oriented clustering of VHR panchromatic satellite images. *IEEE Trans. Geosci. Remote Sens.* **51**(3), 1680–1692 (2013)
24. Tilton, J.C., Swain, P.H.: Incorporating spatial context into statistical classification of multidimensional image data. *LARS (Purdue University. Laboratory for Applications of Remote Sensing)*, vol. 072981 (1981)

25. Tison, C., Nicolas, J., Tupin, F., Maitre, H.: A new statistical model for Markovian classification of urban areas in high-resolution SAR images. *IEEE Trans. Geosci. Remote Sens.* **42**(10), 2046–2057 (2004)
26. Uhlmann, S., Kiranyaz, S.: Integrating color features in polarimetric SAR image classification. *IEEE Trans. Geosci. Remote Sens.* **52**(4), 2197–2216 (2014)
27. Vogel, J., Schiele, B.: Semantic modeling of natural scenes for content-based image retrieval. *Int. J. Comput. Vis.* **72**, 133–157 (2007). <https://doi.org/10.1007/s11263-006-8614-1>
28. Watanabe, T., Suzuki, H.: An experimental evaluation of classifiers using spatial context for multispectral images. *Syst. Comput. Japan* **19**(4), 33–47 (1988)
29. Wu, Y., Ji, K., Yu, W., Su, Y.: Region-based classification of polarimetric SAR images using wishart MRF. *IEEE Geosci. Remote Sens. Lett.* **5**(4), 668–672 (2008)
30. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions (2015)
31. Zhu, X.X., et al.: Deep learning meets SAR (2020)