

# AN UNSUPERVISED SEGMENTATION METHOD FOR REMOTE SENSING IMAGERY BASED ON CONDITIONAL RANDOM FIELDS

A. R. Soares\*, T. S. Körting, L. M. G. Fonseca, A. K. Neves

Coordination of Earth Observation (OBT) -  
Brazil's National Institute for Space Research (INPE)  
Av. dos Astronautas, 1758, São José dos Campos - SP, Brazil  
(anderson.soares,thales.korting,leila.fonseca,alana.neves)@inpe.br

ICWG

**KEY WORDS:** Image segmentation, Remote Sensing, Conditional Random Fields, GEOBIA

## ABSTRACT:

Segmentation is a fundamental problem in image processing and a common operation in Remote Sensing, which has been widely used especially in Geographic Object-Based Image Analysis (GEOBIA). In this paper, we propose a new unsupervised segmentation algorithm based on the Conditional Random Fields (CRF) theory. The method relies on two levels of information: (1) that comes from an unsupervised classification with Fuzzy *C*-Means algorithm; (2) the 8-connected neighbourhood of a pixel. The algorithm was tested on a WorldView-2 multispectral image, with 2m of spatial resolution. Results were evaluated using 6 quality measures, and their performance was compared with other image segmentation algorithms that are usually applied by the Remote Sensing community. Results indicate that the proposed algorithm achieved superior overall performance when compared others, despite some over-segmentation.

## 1. INTRODUCTION

Image segmentation is a fundamental problem in Remote Sensing and one of the most challenging tasks in digital image processing (Körting, 2012). Image segmentation is a process that splits an image, grouping its pixels by a similar feature - such as the grey level - so that the line which splits the segments is, ideally, an edge (Soille, 1999).

Segmentation algorithms should capture important groupings or regions, which often reflect global aspects of the image and should be highly efficient, running in nearly linear time based on the number of pixels (Felzenszwalb, Huttenlocher, 2004). Thus, the relationship among pixels and objects can be interpreted as spatial context, according to Gurney and Townshend (Gurney, Townshend, 1983).

Besides, the ability to detect how a pixel is related to its surroundings can be crucial in Remote Sensing. However, many of the traditional algorithms available on commercial software, and therefore frequently used for analysis, do not deal with such relationship directly. This negligence may cause, in some cases, problems in the definition of objects in the image and consequently in the result of classification. Usually, multiscale approaches are used to avoid different objects from incorrectly form a region. This approach performs several segmentations at different levels, evaluating different types of interaction between the pixels since the use of a single segmentation in many cases is not effective (Johnson et al., 2015).

An alternative to this approach, is based on the Bayesian framework. The algorithms based on this theory are commonly classified as energy-based segmentation algorithms. This approach regained attention over the last decades, in virtue of computational developments. It is known due to its solid mathematical and theoretical background, which favours tasks like image

segmentation where non-deterministic content, such as textures and statistical noise, is still a challenge to the traditional methods (Vantaram, Saber, 2012). The techniques based on this approach deal directly with the neighbourhood of a pixel, which in turn is the main reason why it makes them so interesting to model the spatial context in images.

There are different probabilistic models such as Markov Random Fields (MRF) that have been used for the segmentation of Remote Sensing images over the last years. MRF is a powerful stochastic tool that models joint probability distribution of the image pixels. This type of modelling, has been widely used in computer vision tasks due to its increased resistance to noise. In Remote Sensing, the first applications of MRF occurred in late 1980's (Geman, Geman, 1984, Kittler, Föglein, 1984, Mohn et al., 1987). Recently it is still used for classification (Zhang et al., 2018a, Zhang et al., 2018b, Fang et al., 2018, Borhani, Ghassemian, 2014) and image segmentation. The main disadvantage of MRF is the need to explicitly model the distribution of the likelihood and assume that the observed image data are conditionally independent, given the labels. A feasible solution for this problem relies on the Conditional Random Fields (CRF).

CRF is an undirected graphical models for structured prediction, which are trained to maximise the conditional probability of the outputs given the inputs (Lafferty et al., 2001). It was first proposed to label text sequences and has been successfully applied in computer vision tasks. CRF has characteristics that make it possible to model multivariate outputs, taking advantage of a vast number of input features for prediction. Thus, relaxing the assumption of conditional independence of the observed data, it becomes suitable for image segmentation task (Sutton et al., 2012, Zhang et al., 2015).

In this work, we propose an unsupervised pairwise energy-based segmentation approach for high spatial resolution Re-

\*Corresponding author

ote Sensing images. We combine well-known and simple techniques to create this algorithm, aim at label disjoint regions with similar characteristics.

## 2. IMAGE SEGMENTATION MODEL

Let  $I$  be a Remote Sensing image with  $N$  pixels and  $i$  each pixel of  $I$ . The region that  $i$  belongs to is specified by a label  $y_i$ , which is modelled as a discrete random variable in the interval  $y_i = \{1, 2, \dots, Y\}$ , where  $Y$  is the number of regions that compose the image  $I$ . The observed image features (i.e. colour and texture) are a function of the labels.

In this work we exploited the Markovianity property of random fields to condition the learning process of a pairwise energy model using a voting pool as in (Li et al., 2008, Yuan, Li, 2007). Models like this are a combination of two components; a unary potential, that represents the likelihood of a label being assigned to a pixel  $i$  and a pairwise potential that provides short-range contextual information, describes the compatibility between the pixel under analysis and its neighbours.

Let  $X = \{x_1, x_2, \dots, x_i\}$  denote a set of observed values of the pixels  $i$  and  $Y = \{y_1, y_2, \dots, y_i\}$  the set of labels. The objective of the learning phase is to assign an optimal label  $y$  to each pixel  $i$ , conditioned on observed data  $x_i$  and  $x_j$  and labels  $y_j$ , for all  $j$  in a small neighbourhood  $\eta_i$ , the voting pool. Given an image  $I$ , a pixel  $i$ ,  $\eta_i$ , that is the neighbourhood of the pixel  $i$ , and  $\mathcal{E}$  is the set of all pixel pairs:

$$E_i(y_i, y_{\eta_i}, x_i, x_{\eta_i}) = \psi_u(x_i, y_i) + \sum_i \sum_{j \in \eta_i} \psi_{ij}(x_i, x_j, y_i, y_j) \quad (1)$$

where the unary component  $\psi_u(x_i, y_i)$  is the cost of the pixel  $i$  to receive the label  $y_i$ , the pairwise components  $\psi_{ij}(x_i, x_j)$  computes the cost to the pixels  $i, j$  simultaneously and  $y$  is the image. Once the energy is calculated, the label with the lowest energy is applied to the pixel.

In our approach, when a pixel  $i$  is visited, a voting pool is formed encompassing all  $y$  labels that are defined by the morphological operation reconstruction (Section 3.1). The voting pool holds the energies of a pixel for all possible labels. This information is updated at each iteration using a deterministic selection approach, which chooses the label corresponding to the lowest energies value. This approach allows each pixel to interact with all possible labels in a way that it can find its own identity progressively.

### 2.1 Unary potential

The unary term in Eq. 2 corresponds to the negative log of the posterior probability of any label  $y_i$ , given observation  $x_i$ . This term only considers local features, therefore, the posterior probability of any local classifiers can be used. Despite the noise that might be produced by this approach, it provides a good initial understanding of the objects in the image. We chose the Fuzzy  $C$ -Means (FCM) clustering algorithm, which assesses the displacements between data and the centres of data groups with some degree of pertinence. We use the posterior probability of the FCM to compute the unary term as follows:

$$\psi_u(x_i, y_i) = -\log(P(x_i, y_i)_{FCM}) \quad (2)$$

### 2.2 Pairwise potential

The pairwise potential uses the label information to determine how a pixel relates to its neighbours. This term - on our approach - is designed by the contrast-sensitive Potts model. This potential considers the spectral features of pixels, therefore the probability of them taking the same label is decreased. Usually, the pairwise potential is defined as:

$$\psi(y_i, y_j, x_i, x_j, \theta_\psi) = \theta_{\phi_{ij}}^T \mu(x_i, x_j) \quad (3)$$

where  $\psi$  is the pairwise potential that imposes the contextual information provided by the 8-connected neighbourhood of a pixel  $i$ ,  $\theta_{\psi_{ij}}^T$  is a matrix of parameters obtained during the training,  $\mu(x_i, x_j)$  is a compatibility (Eq. 7) function that assesses the similarities between two pixels  $(x_i, x_j)$ .

To set  $\theta_{\psi_{pq}}$  is necessary a training phase that is prohibitive to an unsupervised approach. However, Zhang and Jia. (2012), proposed an adaptation to this function, in which the  $\theta_{\psi_{pq}}$  is defined as 1.0. This scalar inside the pairwise potential re-defines it, in a way that is solely dependent on the spectral differences between adjacent pixels. This adaptation allows the use of CRF without training, which enables the application of this type of model on unsupervised approach. This adaptation has been used in change detection and classification tasks (Zhang, Jia, 2012, Lv et al., 2018).

Therefore, the pairwise potential used on this work can be expressed as follows:

$$\psi(y_i, y_j, x_i, x_j, \theta_\psi) = \mu(x_i, x_j) \quad (4)$$

$$\mu(x_i, x_j) = \begin{cases} 0, & \text{if } y_i = y_j \\ 1 - g_{ij}, & \text{if } y_i \neq y_j \end{cases} \quad (5)$$

The function  $g_{ij}(y)$  models the interaction between the neighbouring pixels  $i$  and  $j$  measuring the difference in appearance between them. To measure this difference, we use the squared norm of the difference between the pixels.

$$g_{ij} = \frac{(\|x_i - x_j\|^2)}{2\sigma^2} \quad (6)$$

$x_i$  and  $x_j$  are the spectral vectors representing the appearance of the  $i$  and  $j$  pixels and  $\sigma$  is the similarity value defined by user. With this function interaction within a neighbourhood is related to the image data and encourages label coherence in adjacent pixels.

## 3. SEGMENTATION ALGORITHM

As shown in Figure 1, the proposed algorithm, from here called Unsupervised Contextual Segmentation (UCS), utilised a piecewise strategy to model the potential functions.

### 3.1 Clusters definition

In our approach, the number of clusters is defined based on a set of operations of mathematical morphology that is used as a pattern detector. The first stage of this process is high-pass

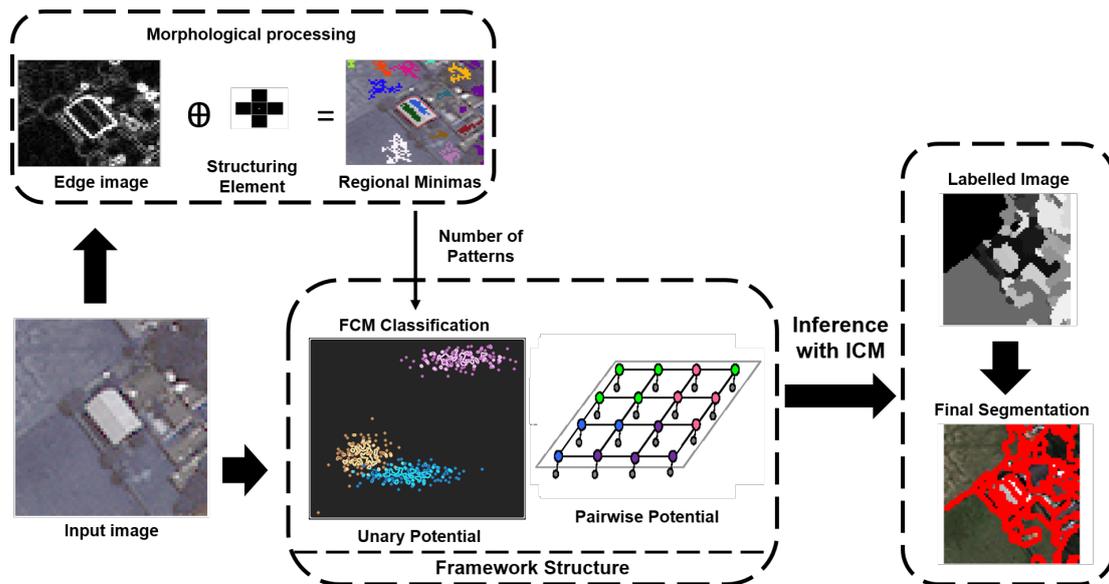


Figure 1. Image segmentation UCS segmentation method.

filtering, which obtains a normalised gradient magnitude image. With this approach, it is possible to manipulate the weights for each band, making it simple to highlight certain objects. Moreover, only similar pixels in all bands of the image will appear as regional minima.

A common problem is that this approach may lead to an over-segmentation since an object usually contain more than a single minimum. To overcome this, we perform a merge of the regional minima based on a similarity value. Thus, a group of regional minima can be merged into a single label based on the similarity value defined by the user, which is the only parameter requested by the proposed algorithm. It is important to state that the definition of seeds is completely unsupervised.

### 3.2 Clustering phase

Using the number of seeds defined by the regional minima, we perform the fuzzy clustering with the FCM algorithm. This classification returns two pieces of information; the probability of each pixel to each seed, used in (Eq.2) and a labelled image, which is used on the label compatibility function 5. We assume that this image is an appropriate starting point for the segmentation, despite that in certain cases it may not be, due to their spectral similarity.

### 3.3 Inference

This procedure maximises the local conditional probabilities iteratively, given an initial labelling, in our case provided by the FCM. Due to simplicity we use ICM algorithm. The energy of each pixel  $i$  (Eq. 1) is computed on each iteration and the label with the lowest energy cost is selected, following a winner-take-all approach. As an iterative procedure, it is performed until it reaches stability. In the end, each pixel in the image is assigned to the most likely object, resulting in a labelled image.

## 4. EXPERIMENTAL SETUP

Tests were conducted on a dataset of a WorldView-2 image of São José dos Campos city, in Brazil. This high spatial resolu-

tion image has 2 meters of spatial resolution and 8 spectral bands, each one with 11 bits of radiometric resolution.

The segmentation results were evaluated by comparing the geometries of resultant objects with 95 delineated objects, representing prominent features in the image, such as house rooftops, small buildings, public squares, industrial warehouses, water bodies and roads. For evaluation, the Precision, Recall and F-measure metrics, were used. All are available on the Python package *Scikit-learn*. Moreover, 3 metrics proposed by Delves et al. (1992), that evaluates the degree of displacement between segments (FITXY), size (FITN) and shape ( $G_{shape}$ ) similarity also were used to assess the regions created with the proposed algorithm.

## 5. RESULTS

To analyse the performance of the UCS method, comparisons were performed with the results obtained with commonly used segmentation algorithms. To that, we perform tests with the Multiresolution algorithm (MRS), (Baatz, Schäpe, 2000) implemented on Definiens eCognition software (Körting et al., 2013), the Meanshift algorithm, available on the Orfeo Toolbox, and the Region Growing (RG) algorithm proposed by (Bins et al., 1996), available on TerraView software.

The following tuning parameters for MRS algorithm were used: 45 scale, 0.3 shape and 0.7 compactness, and weight 2 for infrared bands and 1 for other bands. To perform the segmentation with Meanshift, the spatial radius was 30, 40 the range radius and minimum region size, respectively. For the RG algorithm it was used 30 as minimum size for the regions and 0.002 for similarity, which were the best in our tests with the proposed dataset. For the UCS algorithm, the solely parameter of the algorithm, the similarity, was set as 220.

The process of optimization was performed with the ICM algorithm to reduce computation complexity. As can be observed in Figure 2, the procedure despite been simple can achieve satisfactory results even with a small number of iteration. However, the performance is strongly dependent of the initial labelling.

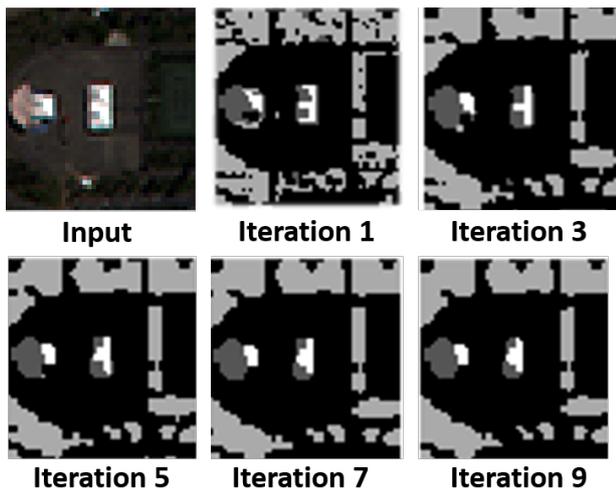


Figure 2. Evolution of labelling during optimization.

Table 1. Average performance of the algorithms.

Metric	MRS	Meanshift	RG	UCS
Precision	0.80	0.76	0.75	0.77
Recall	0.78	0.74	0.76	0.78
F1-score	0.75	0.69	0.70	0.74
FITXY	0.96	0.95	0.95	0.96
FITN	0.87	0.83	0.84	0.88
$G_{shape}$	0.72	0.67	0.67	0.71

Figure 3 shows the segmentation produced by the algorithms for some objects in our image. The segments (yellow line) provided by the MRS and UCS were the best when compared with the others and the reference (red line) at the first column. In the first row, MRS had slightly better results than others. The UCS produced a segment slightly bigger than the actual size, similar to the one produced by Meanshift. The segment produced by the RG algorithm had the worst result, due to over-segmentation.

In the second row, UCS produced a segment that was a little bit bigger than its actual size, but had good agreement with its shape. Similar to the MRS algorithm. The Meanshift produced a good segmentation, despite its shape did not fit as good as UCS and MRS. The RG algorithm produced a segment similar to Meanshift, with a less fitted shape, but with a good agreement.

In the third row, the proposed algorithm had the best performance than all others, with a strong agreement to the boundaries. Others produced segments that grew beyond the reference. In the fourth row, RG provided the best definition of the interest object. All other algorithms grew beyond the reference boundaries.

Table 1 shows the average obtained by each metric analysed for algorithms in all 95 reference regions. The UCS algorithm achieved similar results to MRS, specially on Recall, FITXY, FITN and  $G_{shape}$ . When compared to Meanshift and the RG algorithms it was better in all metrics.

## 6. CONCLUSION

In this work, we described and evaluated an unsupervised segmentation algorithm for high-resolution Remote Sensing images, based on Conditional Random Fields, the UCS algorithm.

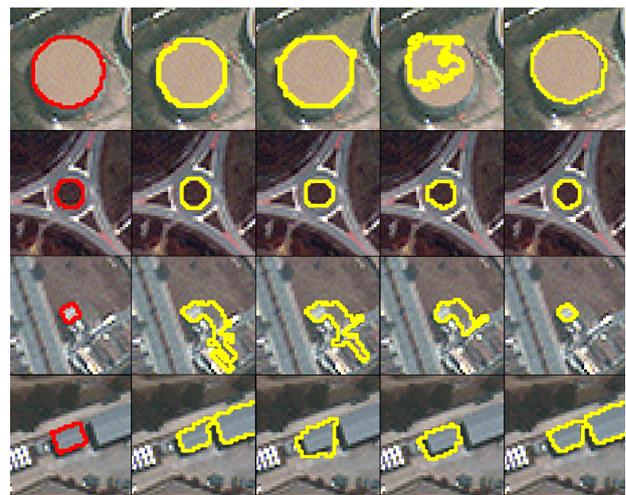


Figure 3. From left to right: Original image with reference polygon in red, MRS segmentation, Mean-shift segmentation, Region Growing segmentation and UCS segmentation algorithm.

We used 95 objects to evaluate its results, using visual assessment and 6 metrics. Moreover, we compared the results against the MRS, the Meanshift and the region growing algorithm proposed by (Bins et al., 1996). The UCS results were superior to Meanshift and the RG. When compared with MRS, it achieved inferior results in Precision and F1-score. It indicates the power of the proposed approach, opening possibilities for improvement, with the integration of long-range contextual information using a high-order model. Furthermore, it is important to highlight, from the analyst perspective, the simplicity of the algorithm, since our method requests only a similarity parameter.

## ACKNOWLEDGEMENTS

The authors thank the Brazilian National Research Council (CNPq) (grant N<sup>o</sup> 140681/2015-9) and São Paulo Research Foundation (FAPESP) (grant N<sup>o</sup>2017/24086-2,) for the financial support.

## REFERENCES

- Baatz, M., Schäpe, A., 2000. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. Wichmann-Verlag (ed.), *XII Angewandte Geographische Informationsverarbeitung*, pp, Herbert Wichmann Verlag, Heidelberg, 12–23.
- Bins, L. S., Fonseca, L. M. G., Erthal, G. J., Ii, F. M., 1996. Satellite imagery segmentation: a region growing approach. *Simpósio Brasileiro de Sensoriamento Remoto*, 8, 677–680.
- Borhani, M., Ghassemian, H., 2014. Hyperspectral image classification based on spectral-spatial features using probabilistic svm and locally weighted markov random fields. *Intelligent Systems (ICIS), 2014 Iranian Conference on*, IEEE, 1–6.
- Fang, Y., Xu, L., Peng, J., Yang, H., Wong, A., Clausi, D. A., 2018. Unsupervised Bayesian Classification of a Hyperspectral Image Based on the Spectral Mixture Model and Markov Random Field. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1-13.

- Felzenszwalb, P., Huttenlocher, D., 2004. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2), 167–181.
- Geman, S., Geman, D., 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6(6), 721–741. <https://doi.org/10.1109/TPAMI.1984.4767596>.
- Gurney, C., Townshend, J., 1983. The use of contextual information in the classification of remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 49, 55–64.
- Johnson, B. A., Bragais, M., Endo, I., Magcale-Macandog, D. B., Macandog, P. B. M., 2015. Image Segmentation Parameter Optimization Considering Within- and Between-Segment Heterogeneity at Multiple Scale Levels: Test Case for Mapping Residential Areas Using Landsat Imagery. *ISPRS International Journal of Geo-Information*, 4(4), 2292. <http://www.mdpi.com/2220-9964/4/4/2292>.
- Kittler, J., Föglein, J., 1984. Contextual classification of multispectral pixel data. *Image and Vision Computing*, 2(1), 13 - 29.
- Körting, T., Fonseca, L., Câmara, G., 2013. GeoDMA - Geographic Data Mining Analyst. *Computers & Geosciences*, 57, 133–145.
- Körting, T. S., 2012. GeoDMA: A toolbox for data mining, object-based and multi-temporal analysis of remote sensing imagery. PhD thesis, Brazil's National Institute for Space Research (INPE).
- Lafferty, J. D., McCallum, A., Pereira, F. C. N., 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 282–289.
- Li, C.-T., Yuan, Y., Wilson, R., 2008. An unsupervised conditional random fields approach for clustering gene expression time series. *Bioinformatics*, 24(21), 2467–2473.
- Lv, P., Zhong, Y., Zhao, J., Zhang, L., 2018. Unsupervised Change Detection Based on Hybrid Conditional Random Field Model for High Spatial Resolution Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 56(7), 4002–4015.
- Mohn, E., Hjort, N. L., Stovik, G. O., 1987. A Simulation Study of Some Contextual Classification Methods For Remotely Sensed Data. *IEEE Transactions on Geoscience and Remote Sensing*, GE-25(6), 796–804.
- Soille, P., 1999. Morphological image processing: principles and applications.
- Sutton, C., McCallum, A. et al., 2012. An introduction to conditional random fields. *Foundations and Trends® in Machine Learning*, 4(4), 267–373.
- Vantaram, S. R., Saber, E., 2012. Survey of contemporary trends in color image segmentation. *Journal of Electronic Imaging*, 21(4), 040901-1-040901-28. <http://dx.doi.org/10.1117/1.JEI.21.4.040901>.
- Yuan, Y., Li, C., 2007. Unsupervised clustering of gene expression time series with conditional random fields. *2007 Inaugural IEEE-IES Digital EcoSystems and Technologies Conference*, 571–576.
- Zhang, C., Sargent, I., Pan, X., Gardiner, A., Hare, J., Atkinson, P. M., 2018a. VPRS-based regional decision fusion of CNN and MRF classifications for very fine resolution remotely sensed images. *IEEE Transactions on Geoscience and Remote Sensing*.
- Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinson, P. M., 2018b. An object-based convolutional neural network (OCNN) for urban land use classification. *Remote Sensing of Environment*, 216, 57–70.
- Zhang, G., Jia, X., 2012. Simplified conditional random fields with class boundary constraint for spectral-spatial based remote sensing image classification. *IEEE Geoscience and Remote Sensing Letters*, 9(5), 856–860.
- Zhang, P., Li, M., Wu, Y., Li, H., 2015. Hierarchical Conditional Random Fields Model for Semisupervised SAR Image Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 53(9), 4933–4951.