Parameter tuning in the support vector machine and random forest and their performances in cross- and same-year crop classification using TerraSAR-X

Rei Sonobe*[a], Hiroshi Tani[b], Xiufeng Wang[b], Nobuyuki Kobayashi[c] and Hideki Shimamura[d]

[a] *JSPS Research Fellow / Graduate School of Agriculture. Hokkaido University, Sapporo, Hokkaido 060-8589, Japan*

[b] *Research Faculty of Agriculture. Hokkaido University, Sapporo, Hokkaido 060-8589, Japan*

[c] *Smart Link Hokkaido, Iwamizawa, Hokkaido 068-0034, Japan*

[d] *PASCO Corporation, Meguro-Ku, Tokyo 153-0043, Japan*

Corresponding author

Kita 9, Nishi 9 Kita-ku, Sapporo, 060-8589, Japan

reysnb@gmail.com

Parameter tuning in the support vector machine and random forest and their performances in cross- and same-year crop classification using TerraSAR-X

This paper describes the comparison of three different classification algorithms for mapping crops in Hokkaido, Japan, using TerraSAR-X data. In the study area, beans, beets, grasslands, maize, potatoes and winter wheat were cultivated. Although classification maps are required for management and for the estimation of agricultural disaster compensation, those techniques have yet to be established. Some supervised learning models may allow accurate classification. Therefore, comparisons among the classification and regression tree (CART), the support vector machine (SVM) and the random forests (RF) were performed. SVM was the best algorithm in this study, achieving overall accuracies of 89.1% for the same-year classification, which is the classification using the training data in 2009 to classify the test data in 2009, and 78.0% for the cross-year classification, which is the classification using the training data in 2009 to classify the data in 2012.

1. Introduction

Land-cover classification is one of the most common applications of remote sensing. Crop type classification maps are useful for estimating the amount of crops harvested or agricultural disaster compensation. Furthermore, the ability to generate crop type classification maps without concurrent training data is useful for reducing labour costs for the management of the agricultural field and early information gathering. Optical remote sensing is still one of the most attractive options for obtaining biomass information, and new sensors are available with fine spatial and spectral resolutions (Sarker and Nichol 2011). In addition, some optical satellites, such as Landsat, have been used for crop type classification (Hartfield *et al.* 2013; Mishra and Crews 2014). Significant information about soil and vegetation parameters has also been obtained through microwave remote sensing, and these techniques are increasingly being used to

manage land and water resources for agricultural applications (Fontanelli *et al.* 2013). Unlike passive systems, synthetic aperture radar (SAR) systems are not dependent on atmospheric influences or weather conditions; thus, they are especially suitable for a multi-temporal classification approaches (Bargiel and Herrmann 2011). The number of studies on rice monitoring and mapping using SAR data has increased, and there are strong correlations between the backscattering coefficients and the plant height and age. There are examples of uses for crop growth monitoring of beets (Vyas *et al.* 2003), maize (Beriaux *et al.* 2013; Blaes *et al.* 2006), and wheat (Fontanelli *et al.* 2013; Lievens and Verhoest 2011; Mattia *et al.* 2003; Sonobe *et al.* 2014c). Furthermore, SAR data have been used to identify specific crop types, such as paddy fields (Choudhury and Chakraborty 2006; Kuenzer and Knauer 2013). The basic idea of these studies is to use multi-temporal SAR data within a vegetation period to clarify the change pattern with the time series (Costa, 2004), and it may be applied for other crop types. The backscattering coefficient is a function of the geometry and dielectric properties of the target and the amount of biomass in agricultural fields. Therefore, different types of temporal changes can be distinguished with multi-temporal SAR data. The first large backscatter intensity change occurs as a result of ploughing and seeding. Smaller changes occur due to variations of biomass and plant water content, and, for X-band SAR data, changes in the plant structure. Furthermore, harvesting causes large backscatter intensity changes (Blaes and Defourny 2003; Sonobe *et al.* 2014a). Sometimes, however, no backscatter intensity change is observed despite geometric changes. This is typically observed for dense vegetation, such as grasslands, for high frequency SAR data, such as C-band (Blaes and Defourny 2003). This indicates the potential of the discrimination between gramineous crops (grass and wheat in this study) and others. Sonobe *et al*. (2014b) shows the potential of X-band SAR data for

mapping winter-wheat planted areas by Otsu's method (Otsu, 1979).

SAR signals acquired under different polarisations show different backscatter responses, providing more information about vegetation (Brisco *et al.* 2013) There is a combination of SAR frequencies, polarisations, and incidence angles that is most suitable for best retrieving soil and vegetation parameters (Ulaby *et al.* 1986).

Multi-temporal dual polarimetric (HH/VV) TerraSAR-X data acquired in StripMap mode were obtained, and the resolutions were 2.75 m in the enhanced ellipsoid corrected format. TerraSAR-X was launched on June 15, 2007, and X-band SAR data are widely available and often operated with dual polarisations. Furthermore, previous studies have proven the high geometric accuracy of TerraSAR-X (Ager and Bresnahan 2009). The robustness of the multi-temporal classification approach with high-resolution TerraSAR-X spotlight data was also shown for a same year classification (Bargiel and Herrmann 2011). However, in order to reduce the labour costs for the selection of training data, which are sometimes collected by in situ surveys, the use of training data selected in another year should be considered.

Within this framework, the main objectives of the present study are to evaluate the potential of Terra-SAR-X data for crop type classification and crop map generation without concurrent training data.

2. Study Area

The experimental area of this study (Figure 1) is the farming area in western Tokachi plain, Hokkaido, Japan (142°55′12″ to 143°05′51″E, 42°52′48″ to 43°02′42″N) at an elevation between 50 and 230 m. The climate of the study area is characterised by warm summers and cold winters with an average annual temperature of 6°C and annual precipitation of 920 mm.

The dominant crops are beans (azuki and soy), beets, grasslands, maize (dent corn and sweet corn), potatoes and winter wheat. A total of 4,955 fields (1,053 beans fields, 709 beet fields, 623 grasslands, 254 maize fields, 831 potato fields and 1,485 winter wheat fields) covered the area in 2009, and 5,074 fields (960 bean fields, 625 beet fields, 644 grasslands, 583 maize fields, 749 potato fields and 1,513 winter wheat fields) covered the area in 2012. The mean size of a fields is 2.16 ha (the maximum area is 18.0 ha and the smallest area is 0.01 ha). The cultivation calendar for the crops in this study area is shown in Table 1.

<Figure 1>

<Table 1>

3. Data and methods

3.1 *Data*

X-band SAR (TerraSAR-X or TanDEM-X) data were acquired on 7 July, 9 August, 31 August and 11 September, 2009, and on 11 July, 2 August, 24 August and 15 September, 2012 (Table 2). The SAR used in this study area was side-looking SARs based on active phased-array antenna technology that flies in a sun-synchronous dawn-dusk orbit with an 11-day repeat at an altitude of 514 km at the equator (Roth *et al.* 2004). Multi-temporal sigma naught data have been revealed to be effective for crop type classification (Bargiel and Herrmann 2011). Therefore, in this study, L1B enhanced ellipsoid corrected products operated in StripMap mode were converted from digital numbers to gamma naught. Then, the mean gamma naught values were calculated for fields for each observation day using field polygons (shape file format) provided by Tokachi Nosai (http://www.tokachi-nosai.or.jp/). These processes were conducted using ERDAS IMAGINE version 14.0 distributed by Intergraph Corporation.

Table 3 represents the numbers of fields of each crop type.

<Table 2>

< Table 3>

3.2 *Classification algorithm and evaluation*

We used multi-temporal backscatter coefficients for crop classification, and the whole processing workflow is illustrated in Figure 2. These classification algorithms were applied using R, which provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering) and graphical techniques, and is highly extensible (R Core Team 2013).

<Figure 2>

In earlier studies, the classification and regression tree (CART) was used to identify crops among alfalfa, corn, cotton, grain, melon orchards and sorghum from Landsat Thematic Mapper (TM) image data, achieving overall accuracies of 87 to 92% for the data acquired in 2008. Furthermore, using training data from one year and applying that data to another year for classification purposes demonstrated that overall accuracies from 71% to 83% are achievable, although accuracies were consistently greater than 85% for some crops (Hartfield *et al.* 2013). In addition to CART, two widely used supervised learning models, the support vector machine (Bovolo *et al.* 2010; Foody and Mathur 2004; Lizarazo 2008; Pal 2008) and random forest (Duro *et al.* 2012; Gislason *et al.* 2006; Kavzoglu and Colkesen 2013; Pal 2005; Rodriguez-Galiano *et al.* 2012), were performed in this study.

SVM builds a model that predicts target values when only the attributes are known. The optimisation problem is solved by mapping the samples into a higher-dimensional space using kernel functions. Instead of modelling probability densities,

SVM uses the marginal sample and most discriminative samples (Cortes and Vapnik 1995). SVM provides sparse models where only a small number of samples are assigned non-zero weights. These samples, called Support Vectors (SV), lie close to the decision surface. The weights or coefficients used in the discriminant function are obtained by maximising a margin criterion (Lizarazo 2008). The Gaussian Radial Basis Function (RBF) kernel was applied (Scholkopf *et al.* 1997), and the two parameters C and γ were tuned using a grid search in this study. The γ parameter defines how far the influence of a single training sample reaches, with low values meaning 'far' and high values meaning 'close'. The C parameter trades off misclassification of training samples against simplicity of the classification boundaries. A low C makes the classification boundaries smooth, whereas a high C aims at classifying all training examples correctly.

RF is an ensemble learning technique that builds multiple trees based on random bootstrapped samples of the training data (Breiman 2001). Each tree is built using a different subset from the original training data, containing approximately two thirds of the cases, and the nodes are split using the best split variable among a subset of m randomly selected variables (Liaw and Wiener 2002). Through this strategy, RF is robust to over-fitting and can handle thousands of input variables (dependent or independent) without variable deletion (Breiman 2001). The output is determined by a majority vote of the trees. Two user-defined parameters are the number of trees (k) and the number of variables used to split the nodes (m); when the number of trees is increased, the generalisation error always converges, and over-training is not a problem. On the other hand, a reduction in the number m of predictive variables results in each individual tree of the model being weaker; therefore, picking a large number of trees is recommended, as is using the square root of the number of variables for the value of m (Breiman 2001).

These classifications algorithms were applied using R (R Core Team 2013), 'rpart' package (Therneau *et al.* 2013), 'randomForest' package (Liaw and Wiener 2002) and 'kernlab' package (Karatzoglou *et al.* 2013). All fields were buffered inward by 10 m, accounting for field shape. The buffer was used to avoid selecting training pixels from the edge of a field, which would create a mixed signal and affect the accuracy assessment.

We used a stratified random-sampling approach to select the fields used for training. Approximately 20% of the crop fields were selected at random as training samples. The number of samples for each crop type was determined based on the percentage of fields in the area. The remaining 80% of fields were used to perform the accuracy assessment.

Classification was performed using the training data in 2009 to classify the test data in 2009 (same-year classification). Furthermore, analysis of the cross-year training and classification was performed using the training data in 2009 to classify the data in 2012 (cross-year classification). Therefore, the crop types of training plots have not changed for the data in 2012. In this study, the selected classification algorithms were classification and regression tree (CART), support-vector machine (SVM) and random forests (RF).

The classification maps were evaluated in terms of their overall accuracy (OA), producer's accuracy (PA) and user's accuracy (UA). Furthermore, the two simple measures of quantity disagreement (QD) and allocation disagreement (AD), which are much more useful to summarise a cross-tabulation matrix than the kappa index of agreement, were used for evaluation (Pontius and Millones 2011). The significant differences among the results were determined at the 95% level of significance using the *Z*-test, which was performed for a pairwise comparison of the proposed methods and

accounted for the ratio between the difference values of two kappa coefficients and the difference in their respective variances (Congalton and Green 2008).

4. Results and discussion

To apply SVM, the optimal values of the two parameters, C and γ, were examined. Table 4 represents the relationships between the two parameters and overall accuracy of the same-year classifications. The higher accuracy observed in the central range of C and γ indicates that nearby same power combination but with opposite sign leads to higher classification accuracy in Table 4. Thus, the parameter pair (C, γ) = (2^{-5}, 2^{6}) was chosen as the optimal parameters in this study.

<Table 4>

For application of RF, the number of trees was tuned, and Figure 3 represents the relationships between the number of trees and the error rate for OOB. Because the results indicate that a number of more than 50 is suitable, 50 was chosen as the number of trees in this study.

<Figure 3>

The corresponding confusion matrixes of classifications using TerraSAR-X data are given in Table 5, and SVM was the best classification algorithm for the both classifications. Although it is impossible to compare the results with earlier studies due to the different study area and the crop types, the overall accuracies of the same year classification are close to the results using backscatter data of three ENVISAT/ASAR data and TerraSAR-X data for crop classification in the North China Plain (Jia *et al.* 2012).

Figure 5 represents the weather conditions in 2009 and 2012. The harvesting periods of winter wheat were approximately the same, regardless of the difference in the climate conditions. Thus, the PAs and UAs in the cross-year classification were more

than 88 % for wheat. For other crops, in 2012, due to the higher air temperature in August to September, the crop growth was advanced 2-5 days earlier than in the normal year, whereas the crop growth was delayed 2-7 days in 2009, according to the announcements by Tokachi Subprefecture (http://www.tokachi.pref.hokkaido.lg.jp/ss/nkc/). In particular, the difference in the growing conditions was large for maize (8 days). To make matters worse, in September the acquisition date in 2012 was 4 days later than that in 2010. The PAs and UAs in the cross-year classification were very low. Nevertheless, the overall accuracy of the cross-year classification using RF or SVM was close to that of Hartfield et al. (2013), which indicated 71-83% using Landsat Thematic Mapper (TM) image data.

We used the *Z*-test to compare the accuracy of the classification methods because the same samples and the same assessment points were used for each classification. The CART, SVM and RF classifications were compared to determine whether they produced significantly different results, as shown in Table 6. Based on a comparison of the overall accuracies, the SVM and RF algorithms were the most accurate for same-year classification because the difference between SVM and RF was not meaningful ($p < 0.05$), as shown in Table 4. However, in the case of cross-year classification, these algorithms differed from each other, and the SVM algorithm was the most accurate (Figure 4).

<Table 5>

<Table 6>

<Figure 4>

5. Conclusions

To generate classification maps, in this study, three StripMap images from TerraSAR-X were used, and three algorithms, CART, SVM and RF, were applied. SVM was the best

classification algorithm for both classifications in terms of OA, QD and AD, and the difference of the cross-year classification result was meaningful ($p < 0.05$). Using the training data from one year and applying those data to another year for classification purposes resulted in an overall accuracy of 78.0%.

These results allow for the automatic and consistent crop type classifications for the six defined classes. The approach offers possibilities to generate crop classification maps to estimate the amount of crops harvested or agricultural disaster compensation with little human power, which has significant cost. Interpretation of the entropy-alpha decomposition may improve the accuracy of the classification due to understanding of the scattering mechanism. In future studies, the potential of the entropy-alpha decomposition for crop type classification will be tested.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 26・5253. The field data were provided by Tokachi Nosai (http://www.tokachi-nosai.or.jp/).

Finally, many thanks to Prof. Michael Collins and the three anonymous reviewers for their important insights, which improved this article.

References

Ager, T.P. and P.C. Bresnahan. 2009. Geometric precision in space radar imaging: Results from TerraSAR-X. In *ASPRS 2009 annual conference*, 9-13. Baltimore, Maryland, USA.

Bargiel, D. and S. Herrmann. 2011. Multi-temporal land-cover classification of agricultural areas in two European regions with high resolution spotlight TerraSAR-X data. *Remote Sensing* 3, no 5: 859-877.

Beriaux, E., C. Lucau-Danila, E. Auquiere and P. Defourny. 2013. Multiyear independent validation of the water cloud model for retrieving maize leaf area index from SAR time series. *International Journal of Remote Sensing* 34, no 12: 4156-4181.

Blaes, X. and P. Defourny. 2003. Retrieving crop parameters based on tandem ERS 1/2 interferometric coherence images. *Remote Sensing of Environment* 88, no 4: 374-385.

Blaes, X., P. Defourny, U. Wegmuller, A. Della Vecchia, L. Guerriero and P. Ferrazzoli. 2006. C-band polarimetric indexes for maize monitoring based on a validated radiative transfer model. *IEEE Transactions on Geoscience and Remote Sensing* 44, no 4: 791-800.

Bovolo, F., L. Bruzzone and L. Carlin. 2010. A novel technique for subpixel image classification based on support vector machine. *IEEE Transactions on Image Processing* 19, no 11: 2983-2999.

Breiman, L. 2001. Random forests. *Machine Learning* 45, no 1: 5-32.

Brisco, B., A. Schmitt, K. Murnaghan, S. Kaya and A. Roth. 2013. SAR polarimetric change detection for flooded vegetation. *International Journal of Digital Earth* 6, no 2: 103-114.

Choudhury, I. and M. Chakraborty. 2006. SAR signature investigation of rice crop using radarsat data. *International Journal of Remote Sensing* 27, no 3: 519-534.

Congalton, R.G. and K. Green. 2008. *Assessing the accuracy of remotely sensed data: Principles and practices*. Boca Raton, Florida, United States: CRC Press.

Cortes, C. and V. Vapnik. 1995. Support-vector networks. *Machine Learning* 20, no 3: 273-297.

Costa, M. P. F. 2004. Use of SAR satellites for mapping zonation of vegetation communities in the Amazon floodplain. *International Journal of Remote Sensing*, 25, no 10:1817-1835.

Duro, D., S. Franklin and M. Dube. 2012. Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests. *International Journal of Remote Sensing* 33, no 14: 4502-4526.

Fontanelli, G., S. Paloscia, M. Zribi and A. Chahbi. 2013. Sensitivity analysis of X-band SAR to wheat and barley leaf area index in the merguellil basin. *Remote Sensing Letters* 4, no 11: 1107-1116.

Foody, G. and A. Mathur. 2004. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 42, no 6: 1335-1343.

Gislason, P., J. Benediktsson and J. Sveinsson. 2006. Random forests for land cover classification. *Pattern Recognition Letters* 27, no 4: 294-300.

Hartfield, K., S. Marsh, C. Kirk and Y. Carriere. 2013. Contemporary and historical classification of crop types in Arizona. *International Journal of Remote Sensing* 34, no 17: 6024-6036.

Jia, K., Q. Li, Y. Tian, B. Wu, F. Zhang and J. Meng. 2012. Crop classification using multi-configuration SAR data in the North China Plain *International Journal of Remote Sensing* 33, no 1: 170-183.

Karatzoglou, A., A. Smola and K. Hornik. Package 'kernlab'. http://cran.r-project.org/web/packages/kernlab/kernlab.pdf.

Kavzoglu, T. and I. Colkesen. 2013. An assessment of the effectiveness of a rotation forest ensemble for land-use and land-cover mapping. *International Journal of Remote Sensing* 34, no 12: 4224-4241.

Kuenzer, C. and K. Knauer. 2013. Remote sensing of rice crop areas. *International Journal of Remote Sensing* 34, no 6: 2101-2139.

Liaw, A. and M. Wiener. 2002. Classification and regression by random forest. *R News* 2, no 3: 18-22.

Lievens, H. and N. Verhoest. 2011. On the retrieval of soil moisture in wheat fields from L-band SAR based on water cloud modeling, the IEM, and effective roughness parameters. *IEEE Geoscience and Remote Sensing Letters* 8, no 4: 740-744.

Lizarazo, I. 2008. Svm-based segmentation and classification of remotely sensed data. *International Journal of Remote Sensing* 29, no 24: 7277-7283.

Mattia, F., T. Le Toan, G. Picard, F. Posa, A. D'alessio, C. Notarnicola, A. Gatti, M. Rinaldi, G. Satalino and G. Pasquariello. 2003. Multitemporal C-band radar measurements on wheat fields. *IEEE Transactions on Geoscience and Remote Sensing* 41, no 7: 1551-1560.

Mishra, N.B. and K.A. Crews. 2014. Mapping vegetation morphology types in a dry savanna ecosystem: Integrating hierarchical object-based image analysis with random forest. *International Journal of Remote Sensing* 35, no 3: 1175-1198.

Otsu, N. 1979. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9 no 1, 62-66.

Pal, M. 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing* 26, no 1: 217-222.

Pal, M. 2008. Ensemble of support vector machines for land cover classification. *International Journal of Remote Sensing* 29, no 10: 3043-3049.

Pontius, R. and M. Millones. 2011. Death to kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing* 32, no 15: 4407-4429.

R Core Team. 2013. A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. Accessed October 11. http://www.R-project.org/

Rodriguez-Galiano, V., M. Chica-Olmo, F. Abarca-Hernandez, P. Atkinson and C. Jeganathan. 2012. Random forest classification of mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment* 121: 93-107.

Roth, A., M. Huber and D. Kosmann. 2004. Geocording of TerraSAR-X data. In *20th ISPRS Congress*, 840-844. Istanbul.

Sarker, L. and J. Nichol. 2011. Improved forest biomass estimates using alos avnir-2 texture indices. *Remote Sensing of Environment* 115, no 4: 968-977.

Scholkopf, B., K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio and V. Vapnik. 1997. Comparing support vector machines with gaussian kernels to radial basis function classifiers. *IEEE Transactions on Signal Processing* 45, no 11: 2758-2765.

Sonobe, R., H. Tani, X. Wang, N. Kobayashi and H. Shimamura. 2014a. Random forest classification of crop type using multi-temporal TerraSAR-X dual-polarimetric data. *Remote Sensing Letters* 5, no 2: 157-1564.

Sonobe, R., H. Tani, X. Wang, N. Kobayashi, A. Kimura and H. Shimamura. 2014b. Application of multi-temporal TerraSAR-X data to map winter wheat planted areas in Hokkaido, Japan. *Japan Agricultural Research Quarterly* 48, no 4: 465-470.

Sonobe, R., H. Tani, X. Wang, N. Kobayashi and H. Shimamura. 2014c. Winter wheat growth monitoring using multi-temporal TerraSAR-X dual-polarimetric data. *Japan Agricultural Research Quarterly* 48, no 4: 471-476.

Therneau, T., B. Atkinson and B. Riple. Package 'rpart'. http://cran.r-project.org/web/packages/rpart/rpart.pdf.

Ulaby, F.T., M.K. Moore and A.K. Fung. 1986. *Microwave remote sensing active and passive*. Norwood: Artech House.

Vyas, S., M. Steven, K. Jaggard and H. Xu. 2003. Comparison of sugar beet crop cover estimates from radar and optical data. *International Journal of Remote Sensing* 24, no 5: 1071-1082.

Table 1. Cultivation calendar for the crops in this study area.

	May	June	July	August	September	October
Beans	sowing	sprouting				harvesting
Beet	sowing	sprouting				harvesting
Grassland	appearance of ears of grain	1st harvesting		2nd harvesting		
Maize	sowing		appearance of tassel			harvesting
Potato	planting	sprouting			harvesting	
Wheat	appearance of ears of grain			harvesting		

Table 2. Characteristics of the satellite images used.

Satellite	Mode	Polarisation	Acquisition date	Orbit	Pixel spacing (m)	Incidence angle (°)
TerraSAR-X	StripMap	HH, VV	7 July, 2009	Ascending	2.75	42.3
TerraSAR-X	StripMap	HH, VV	9 August, 2009	Ascending	2.75	42.3
TerraSAR-X	StripMap	HH, VV	31 August, 2009	Ascending	2.75	42.3
TerraSAR-X	StripMap	HH, VV	11 September, 2009	Ascending	2.75	42.3
TanDEM-X	StripMap	HH, VV	11 July, 2012	Ascending	2.75	42.3
TanDEM-X	StripMap	HH, VV	2 August, 2012	Ascending	2.75	42.3
TanDEM-X	StripMap	HH, VV	24 August, 2012	Ascending	2.75	42.3
TanDEM-X	StripMap	HH, VV	15 September, 2012	Ascending	2.75	42.3

Table 3. Crop type and number of fields.

Crop type	No. of fields		
	Data acquired in 2009		Data acquired in 2012
	Training data	Test data	
Beans	211	842	960
Beet	142	567	625
Grassland	124	499	644
Maize	50	204	583
Potato	167	664	749
Wheat	297	1188	1513

Table 4. Relationships between the overall accuracy and the parameters of SVM.

		C							
		2^1	2^2	2^3	2^4	2^5	2^6	2^7	2^8
γ	2^{-7}	0.823	0.790	0.807	0.825	0.833	0.838	0.848	0.877
	2^{-6}	0.836	0.812	0.829	0.839	0.847	0.853	0.856	0.882
	2^{-5}	0.853	0.870	0.874	0.882	0.884	0.891	0.884	0.881
	2^{-4}	0.870	0.878	0.885	0.886	0.890	0.887	0.882	0.878
	2^{-3}	0.877	0.885	0.887	0.887	0.887	0.881	0.878	0.870
	2^{-2}	0.883	0.884	0.886	0.883	0.881	0.877	0.863	0.862
	2^{-1}	0.884	0.884	0.885	0.881	0.873	0.869	0.868	0.866
	2^{0}	0.876	0.878	0.871	0.869	0.868	0.866	0.865	0.863

High

Low

Table 5. Accuracy results.

Algorithm / Class	classification and regression tree (CART)		random forest（RF）		support vector machine（SVM）	
	Test data acquired in 2009	Data acquired in 2012	Test data acquired in 2009	Data acquired in 2012	Test data acquired in 2009	Data acquired in 2012
PA						
Beans	0.678	0.460	0.802	0.542	0.859	0.630
Beet	0.785	0.562	0.937	0.734	0.951	0.946
Grassland	0.810	0.814	0.896	0.856	0.834	0.873
Maize	0.426	0.058	0.466	0.122	0.608	0.110
Potato	0.858	0.818	0.893	0.874	0.886	0.907
Wheat	0.962	0.887	0.960	0.962	0.960	0.963
UA						
Beans	0.801	0.522	0.818	0.571	0.853	0.564
Beet	0.878	0.680	0.863	0.873	0.822	0.847
Grassland	0.884	0.681	0.871	0.804	0.929	0.813
Maize	0.554	0.810	0.792	0.826	0.832	0.821
Potato	0.674	0.420	0.879	0.490	0.899	0.668
Wheat	0.890	0.931	0.938	0.952	0.944	0.960
OA	0.812	0.652	0.878	0.732	0.891	0.780
QD	7.013	16.456	2.548	12.712	2.926	9.953
AD	11.756	18.388	9.637	14.131	7.997	12.042

Note: PA, producer’s accuracy; UA, user’s accuracy; OA, overall accuracy; QD, quantity disagreement; AD, allocation disagreement.

Table 6. *Z*-test results for the classification.

classification		SVM	RF
same-year	CART	10.22	8.42
	SVM		1.8
cross-year	CART	15.49	9.61
	SVM		5.83

Figures

Figure 1. The study area (background map shows the TerraSAR-X data acquired on 2 May, 2009).

Figure 2. Overview of the data processing.

Figure 3. Relationships between the number of trees and the error rate for OOB samples.

Figure 4. Crop type classification map in 2012 using SVM.

Figure 5. Weather conditions in 2009 and 2012.

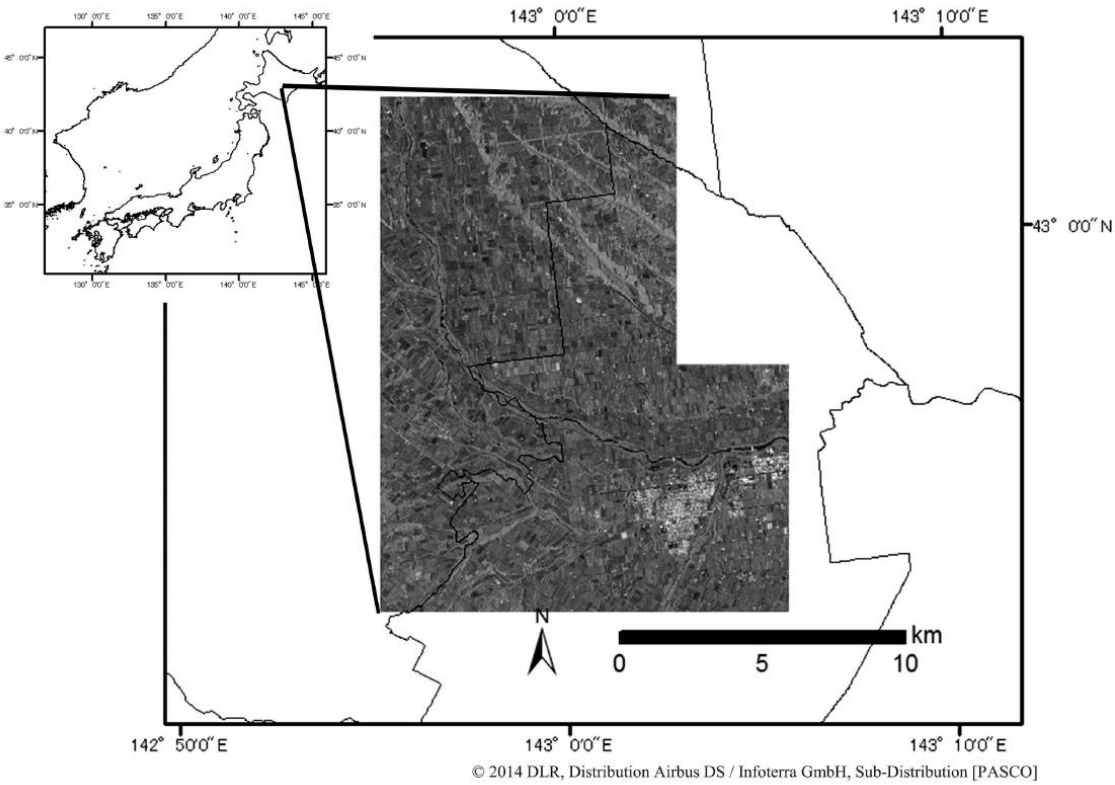

Figure 1. The study area (background map shows the TerraSAR-X data acquired on 2 May, 2009).

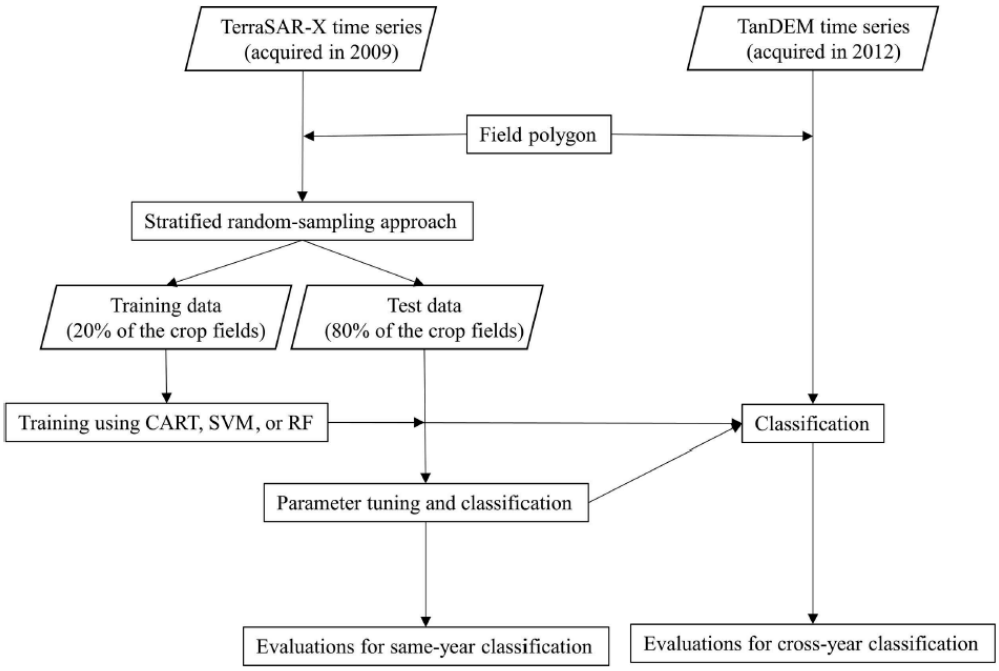

Figure 2. Overview of the data processing.

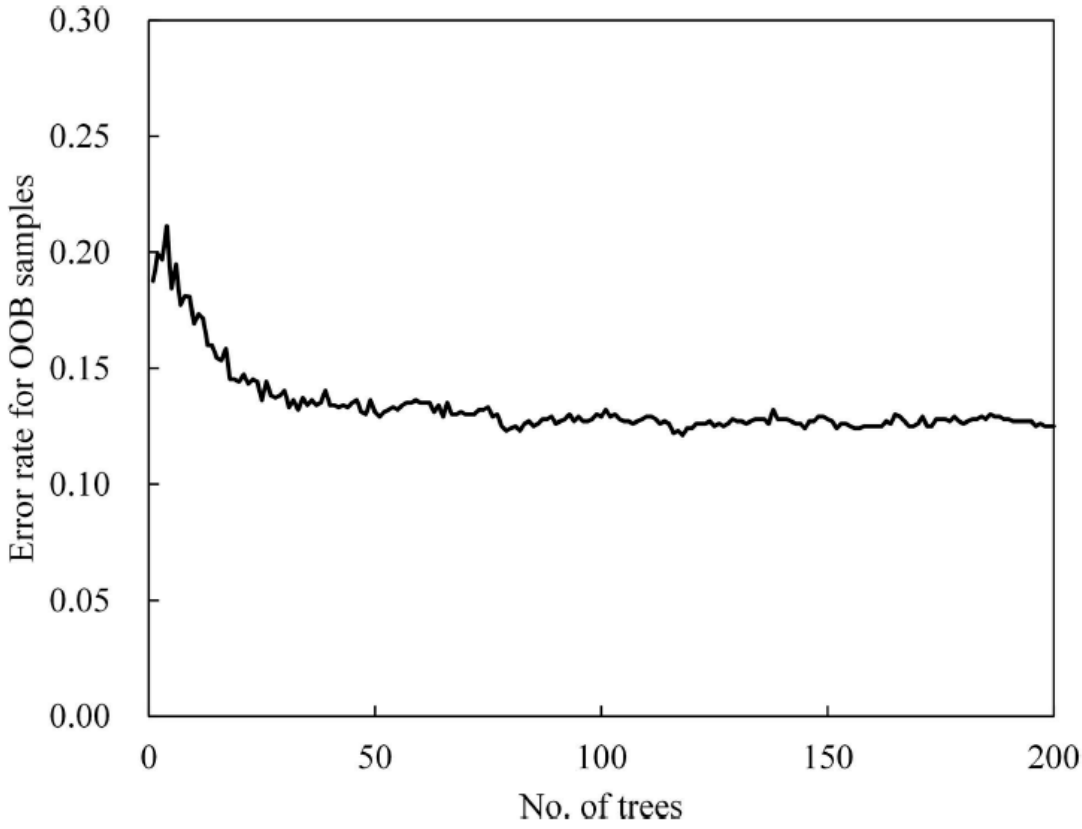

Figure 3. Relationships between the number of trees and the error rate for OOB samples.

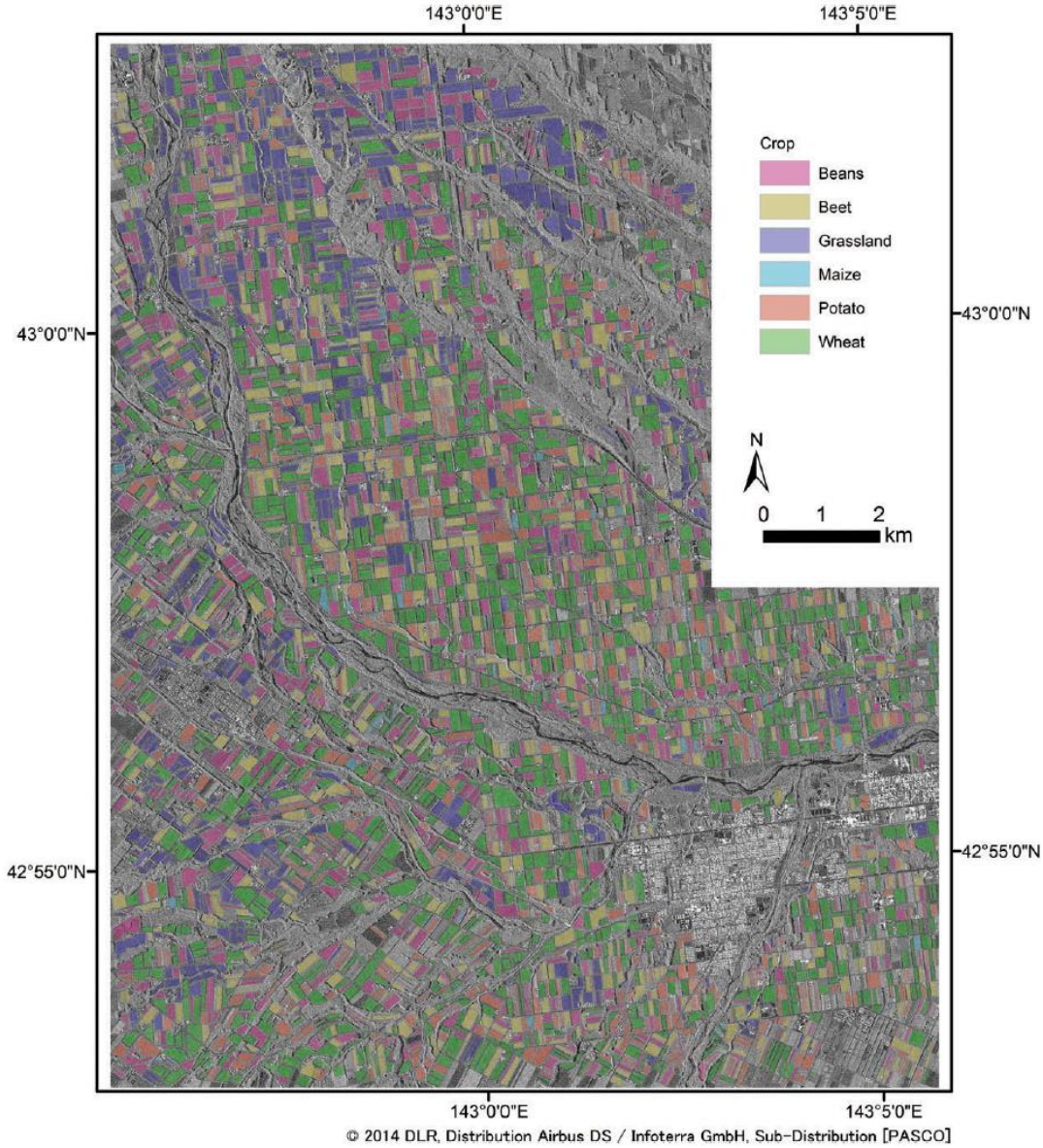

Figure 4. Crop type classification map in 2012 using SVM.

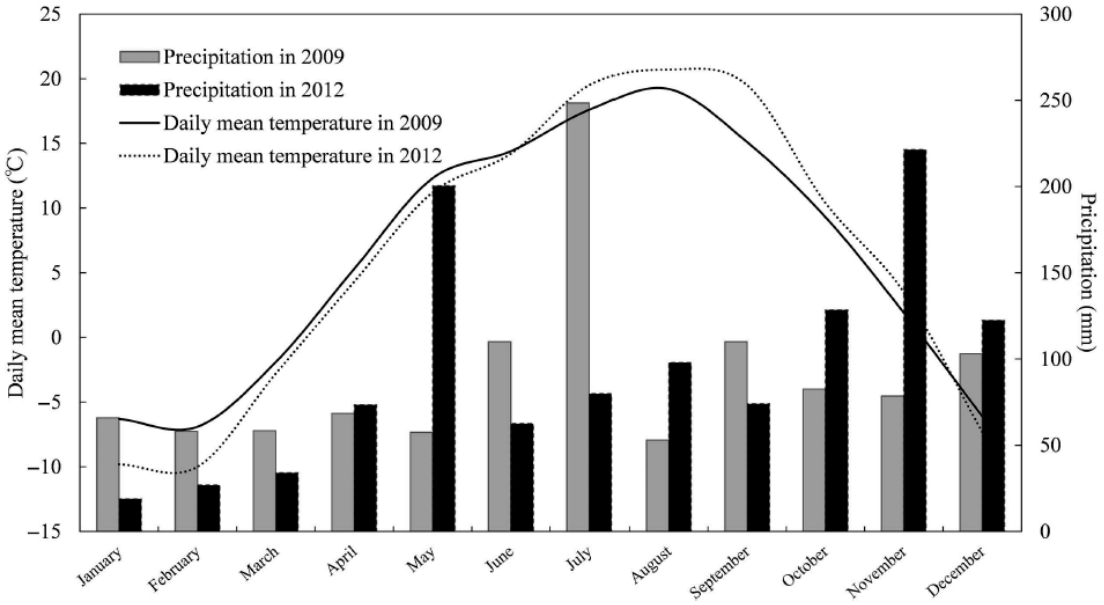

Figure 5. Weather conditions in 2009 and 2012.