# ESTIMATION OF GRASSLAND FRACTIONAL COVERAGE FROM SURFACE MINING AREA USING HIGH RESOLUTION IMAGERY

Nisha Bao1*, Lixin Wu1,2, Shanjun Liu1, Zonglei Cai1
1.Institute for Geo-informatics&Digital Mine Research,Northeastern University,Shenyang110819,China
2.School of Environment Science and Spatial Informatics, China University of mining and technology, Xuzhou 221116)
baonisha@hotmail.com

**ABSTRACT** The aims of this study is to indicate the potential ability of Chinese GF-1 satellite imagery in grassland cover retrieval at the surface mine area of north prairie. The Chinese GF-1 and SPOT6 multi-spectral imagery was selected. The Dimidiate Pixel Model, Partial Least Squares Regression (PLSR) and Support Vector Machine (SVM) models based on the different vegetation indices were established to retrieval grassland cover. The field observation data from sampling plots was used to assess the accuracy of different models. Furthermore, Monte Carlos simulation was conducted to evaluate cross-scale error propagation from SPOT6 to GF-1 data. The results showed that enhanced vegetation index (EVI) from GF-1 can produce high accuracy （$R^2$=0.8149，RPD=2.336，RMSE=8.694） based on SVM model, meanwhile, the normalized difference vegetation index (NDVI) from SPOT-6 data can produce high accuracy（$R^2$=0.8755，RPD=2.870，RMSE=7.032）based on SVM model as well. In terms of the cross-scale error propagation from SPOT6 to GF-1 data, the SVM model outperformed PLS model in grassland cover retrieval. Therefore, the Chinese GF-1 data can provide grassland cover with high accuracy based on SVM model.

**KEYWORDS** : GF-1 data；SPOT6 data；fractional vegetation coverage of grassland；Dimidiate Pixel Model；PLSR；SVR

## 1. INTRODUCTION

Fractional vegetation cover (FVC) refers to the percentage taken by the vertical projected area of vegetation (including leafs, stem and branches) in the total statistical area(Urban et al. 2010). Fractional vegetation cover is an important variable for describing vegetation quality and reflecting ecosystem changes (Laliberte et al. 2007). It is also a sensitive indicator of land degradation and desertification in semi-arid grasslands(Liu et al. 2012). Although the digital camera from ground survey can provide the FVC information in the plant quadrate with higher accuracy for a smaller range, it is unlikely to do so for a larger range. Remote sensing provides the possibility for large scale of the FVC (Xie et al. 2008). The satellite reflectance values must be transformed into plant coverage to make the plant coverage product suitable for estimation and monitoring purposes. Linear spectral unmixing and spectral angle mapper techniques have been widely used to translate reflectance values from remote sensing imagery into plant coverage values. More advanced multivariate methods used to retrieve plant coverage from satellite data encompass partial least squares regressions (PLSR) and machine-learning algorithms such as support vector machines (SVM), which have been evaluated as a valuable tool to cope with non-linear relations and highly correlated predictor variables (Lehnert et al. 2015).

Mining actives, particularly surface coal mine at the semi-grassland has a large impact on the surrounding landscape ecosystem function(Demirel et al. 2011). The monitoring and estimation of plant coverage is essential, which would indicate the condition of grassland affected by human activates including mining and grazing. Fractional vegetation cover (FVC) is an important surface parameter for characterizing land surface vegetation cover as well as the most effective indicator for assessing mine affected environment(Zou et al. 2010). With the development of the remote sensing data with high resolution, the accuracy of estimation vegetation coverage has been improved. Wide field view (WFV) sensor on board the Chinese GF-1, the first satellite of the China High-resolution Earth Observation System, is acquiring multi-spectral data with decametric spatial resolution, high temporal resolution and wide coverage, which are valuable data sources for environment monitoring(Jia et al. 2016). The objective of this study is to develop a general and reliable fractional vegetation cover (FVC) estimation algorithm for GF-1 WFV data in semi-arid north prairie with land cover of coal mine area, degraded grassland, grazing grassland and fence grassland. The algorithms, including Dimidiate Pixel Model, Partial Least Squares Regression (PLSR) and Support Vector Machine (SVM) models based on the different vegetation indices are expected to estimate FVC from GF-1 WFV reflectance data with spatial resolution of 16 m and SPOT-6 multi-spectral data with spatial resolution 6m.

## 2. STUDY AREA AND DATA

### 2.1 Study area description and data collection

Hulunber meadow steppe is located in gentle hilly area of the greater Khingan Range, with a temperate semi-arid continental climate characterized with a long cold winter, short cool summer, dry windy spring and fall of early frost and sudden drop in temperature. Yimin opencast coal mine is in the central of Hulun Buir meadow steppe. The mining activates started at 1985 with progressive land rehabilitation. The native landscape is plain and temperate grasslands dominated by *Leymuschinensis,lyme grass and Stipagrandis,* and subject to grazing. The soil is chernozemor dark chestnut. This area has undergone some desertification and grassland degradation as the overgrazing, farming and mining. Thus, our field observation was carried out from the grazing exclusion, overgrazing and mine restoration area (Figure 1). The soil substrates on the restoration dump of mine is from the natural soil. Over 20cm of topsoil (chernozem) was used as a growth medium to cover the dump to support vegetation rehabilitation. The restoration dump understory is dominated by native grasses. There is no specified treatment on the restoration dump.


Figure 1 the photos of grass sampling plots under different land use

Ground sampling and observation experiment was from August 3 to July 28, 2015. Ground sampling is aligned with remote sensing image. Sampling range extended 10km which take Yimin opencast coal mine as the center, and the sampling area is 1600 $km^2$. Based on the characteristics of the vegetation in the study area, random sampling method was used. At the same time using GPS to record each kind of geographic coordinates, vegetation type, and geomorphic environment, a total of 64 samples (size 1 m×1 m)(Figure 2). SONY SELP 165 digital camera was used to take photos above quadrat 1m.

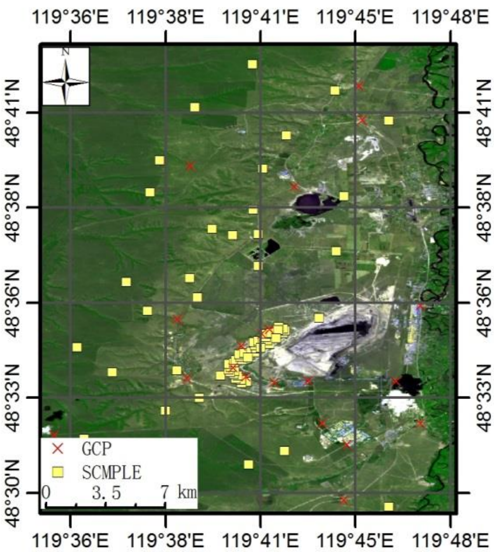


Figure.2 Distribution of the samples and GCP

## 2.2 Remote sensing data preprocessing

GF-1 and SPOT6 multispectral data covering the study area were obtained at July 5th and July 28th 2015. The parameters of two types of imagery are as shown in table 1. The image preprocessing included radiometric calibration, atmospheric correction, and geographic registration. The parameters and the spectral response function of GF-1 were downloaded from China Central For Resources Satellite Data and Application (http://www.cresda.com/CN/). First-order polynomial method was selected for geographic registration using the ground contrl points. The RMSE value of geographic registration of GF-1 and SPOT6 data are below one pixel.

Table.1 Characteristic of for SPOT and GF-1 remote sensing data

| sensor | band | resolution/m | width/km |
|---|---|---|---|
| GF-1 | blue（0.45-0.52um）、green（0.52-0.59um）、red（0.63-0.69um）、near-infrared（0.77-0.89um） | 16 | 800 |
| SPOT6 | blue（0.45-0.52um）、green（0.53-0.59um）、red（0.62-0.69um）、near-infrared（0.76-0.89um） | 6 | 60 |

The normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), soil adjusted vegetation index (SAVI) and modified soil adjusted vegetation index (MSAVI) were selected to estimate vegetation coverage (Perry and Lautenschlager 1984, Huete 1988). The characteristics of the four vegetation index are shown in table 2. NDVI has been widely used in vegetation monitoring, which is the highest correlated with vegetation coverage. EVI base on red band and near-infrared band joining blue band, can reduce the aerosol scattering and the influence of soil background. SAVI is soil adjusted vegetation index, can reduce the influence of soil background. MSAVI is modified soil adjusted vegetation index, is developed on the basis of SAVI.

Table.2 vegetation index

| Vegetation index | formula | advantage |
|---|---|---|

| | | |
|---|---|---|
| NDVI | $\frac{\rho_{NIR}-\rho_{RED}}{\rho_{NIR}+\rho_{RED}}$ | eliminate the agle of the sun, the satellite observation Angle, and the influence of the terrain change |
| EVI | $2.5\times\frac{\rho_{NIR}-\rho_{RED}}{\rho_{NIR}+6.0\rho_{RED}-7.5\rho_{BLUE}+1}$ | overcome the atmospheric attenuation of spectral reflection and weaken the influence of soil background |
| SAVI | $\frac{\rho_{NIR}-\rho_{RED}}{\rho_{NIR}+\rho_{RED}+L}\times(1+L)$ | weaken the influence of soil background |
| MSAVI | $\frac{2\rho_{NIR}+1-\sqrt{(2\rho_{NIR}+1)^2-8(\rho_{NIR}-\rho_{RED})}}{2}$ | |

## 3. METHODS

### 3.1 Calculation of plant coverage from digital images

The digital images were taken from each measuring plot, which preprocessed by geometric correction and subset. We derived the plant coverage within the greenness value of the digital photo. We calculated an image from the RGB values in which green vegetation was enhanced using the following rules (Baxendale et al. 2016): (1) Extracting vegetation by setting up the green band is greater than the red and blue band; (2) removing influence by shadow with setting up the green band minus red band is greater than a threshold.

### 3.2 Dimidiate pixel model

The dimidiate pixel model (DPM) is a simple but practical model for remote sensing estimation(Sohn and Mccoy 1997). The model hypothesizes that the 1-pixel surface consists of two parts: the vegetation-covered and the vegetation-uncovered surfaces. According to the principle of dimidiate pixel model $R$, the information observed from the remote sensor can be expressed as the sum of $R_v$, the information contributed by the green vegetation components, and $R_s$, the information coming from the soil components. The mixed pixel model simplifies the pixel information $R$ into $Rv$ and $Rs$, representing the vegetation information and non-vegetation information, respectively.

$$R = R_v + R_s \qquad (1)$$

Let $fc$ be the coverage percentage of the vegetation in one pixel, the FVC of this pixel, then the percent of the non-vegetation coverage is 1–$fc$. If let $R_{veg}$ be the pixel information under all covered by vegetation, $R_v$ can be expressed as the product of $R_{veg}$ and $fc$.

$$R_v = fc \times R_{veg} \qquad (2)$$

Similarly, if there is no vegetation coverage (pure pixel all covered by soils) and let $R_{soil}$ be the pixel information, then $Rs$ is the product of $R_{soil}$ and 1–$fc$.

$$R_s = (1 - fc) \times R_{soil} \qquad (3)$$

Substituting Equations 2 and 3 into Equation 1, the following equation can be obtained:

$$fc = \frac{(R - R_{soil})}{(R_{veg} - R_{soil})} \qquad (4)$$

Li analyzed each vegetation index and pointed out that dimidiate pixel model has the best correlation with NDVI.

The formula of dimidiate pixel mode can be obtained:

$$FVC = \frac{NDVI - NDVI_{soil}}{NDVI_{\mathrm{veg}} - NDVI_{soil}} \qquad (5)$$

where, $NDVI_{soil}$ and $NDVI_{veg}$ represent the NDVI values of the bare soil or non-vegetation coverage pixel and a pure vegetation pixel respectively. Theoretically, the Nsoil value of most bare soil surfaces approaches zero. However, the range of Nsoil values is between -0.1 and 0.2 due to the impact of many factors. According to the frequency statistics and the actual situation of vegetation coverage in study area, the values of $NDVI_{soil}$ and $NDVI_{veg}$ are select 0.05 and 0.95[17].

### 3.3 PLSR and SVM

The PLSR and SVM model were established by the vegetation index from remote sensing imagery and vegetation coverage observed from digital image. In order match the vegetation index and vegetation coverage in spatial location, the 3*3 window from remote sensing image was generated by selecting the field sampling plots as center. Then the average vegetation indices value from 3*3 windows was calculated, which corresponds to the vegetation coverage from field sample plot. Finally, 44 samples were selected to establish the model, and 20 samples were used as test data to validate the predictive performance of the model. The root-mean-square error (RMSE), the relative percent deviation (RPD) and the $R^2$-value of the linear relationship between predicted and observed plant coverage values were calculated to evaluate the model. When the RPD < 1.75 that the established model is not available; When the RPD is in 1.75 to 1.75 that the established model is basically available; When the RPD is in 2.25 to 3 that the established model is generally successful; When the RPD > 3 that the established model is very successful(Boulet et al. 2013).

The PLSR is a bilinear modeling technique where information in the original x data is projected onto a small number of underlying (“latent”) variables called PLSR components. The y data are actively used in estimating the “latent” variables to ensure that the first components are those that are most relevant for predicting the y variables. Interpretation of the relationship between x and y data is then simplified as this relationship is concentrated on the smallest possible number of components. In this study, the x represents various vegetation indices, and y represents vegetation coverage. This regression model was implemented in R 3.2.1 version, using the package “pls” for the PLSR modeling.

SVM regression was the second approach to estimate plant coverage values using the same feature space as for PLSR as predictors. One of the important benefits of SVM regression models is the very good balance between estimation accuracy and overfitting (Schwieder et al. 2014).SVM is a kernel-based learning method from statistical learning theory. It is possible to derive a linear hyperplane as a decision function for non-linear problems and then apply a back-transformation in the non-linear space using the kernel-based learning method. The radial basis function kernel, which is the typical general-purpose kernel, was used in this study. The support vector machine was performed by using the “e1071 package”, an R interface to library for support vector machines

(LIBSVM).Optimization of the SVM parameters (C, e, and kernel-specific parameter) and the selection of the best preprocessing steps have been done by a systematic grid search of the parameters using leave one-out cross-validation on the training set.

### 3.4 Error propagation and validation of plant coverage predictions

The 6-m-resolution training samples used to fit the regression models at the GF-1 scale encompass estimation errors at the SPOT scale model output. To quantify this error, we used Monte-Carlo simulations as proposed by Gessner (2013) to evaluate cross-scale error propagation. SPOT6 data was selected to fit GF-1 in this study.

(1) The plant coverage values derived from the in-depth analysis of digital photographs taken during field work are selected to establish model with vegetation index derived from SPOT6 data (as shown ① in figure 3);

(2) The plant coverage values were estimated with the model in ① by averaging vegetation index from the SPOT6 data belonging to each respective and co-located GF-1 pixel(as shown ② in figure 3);

(3) Vegetation index derived from GF-1 data was established model with the vegetation coverage in ②(as shown ③ in figure 3);

(4) The plant coverage values were estimated with the model in③ by vegetation index from the GF-1 data.

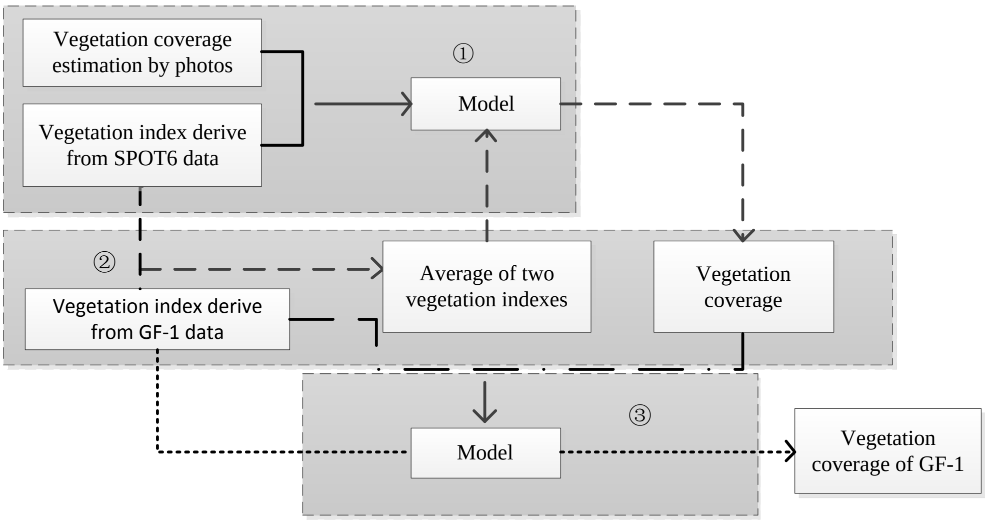


Figure.3 the flow chart of cross-scale error propagation

## 4. RESULT AND ANALYSIS

### 4.1 FVC analysis from sample plots

As shown in figure 4, the extracted greenness footprint for FVC from digital image in 1m*1m sample plot. The FVCs were calculated using digital images from a total of 64 samples. The maximum of vegetation coverage is 73.32%, the minimum is 4.39%, average is 25.37%, the standard deviation is 20.20, and P-value is 1.833 e-07. Statistical analysis of FVC is shown in figure 6. Normal distribution of the QQ, scatter near roughly in a straight line, thus think sample data are approximately subordinate to logarithm normal distribution(Figure 5).

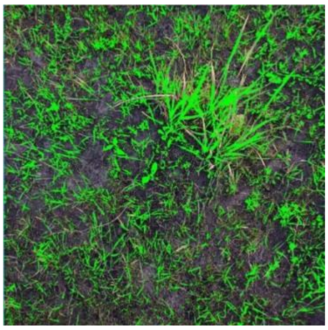

Figure.4The result of estimating FVC by digital image taken in field sampling plot

Table 3 the calibration and validation results of FVC using PLSR and SVM

| Method | Sensor | | Calibration（n=44） | | | | Validation（n=20） | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | NDVI | EVI | SAVI | MSAVI | NDVI | EVI | SAVI | MSAVI |
| PLSR | GF | RMSE | 9.967 | 9.604 | 10.280 | 11.039 | 11.544 | 10.905 | 12.316 | 11.657 |
| | | $R^2$ | 0.7484 | 0.7661 | 0.7319 | 0.6907 | 0.7160 | 0.7644 | 0.6570 | 0.6725 |
| | | RPD | 2.038 | 2.115 | 1.976 | 1.840 | 1.790 | 1.895 | 1.678 | 1.773 |
| | SPOT6 | RMSE | 9.232 | 10.007 | 9.425 | 10.482 | 10.906 | 13.179 | 13.334 | 11.546 |
| | | $R^2$ | 0.7819 | 0.7423 | 0.7725 | 0.7204 | 0.6945 | 0.6525 | 0.5913 | 0.6560 |
| | | RPD | 2.186 | 2.016 | 2.141 | 1.925 | 1.895 | 1.568 | 1.550 | 1.789 |
| SVM | GF | RMSE | 8.439 | 8.694 | 8.920 | 9.018 | 13.518 | 10.295 | 12.061 | 12.967 |
| | | $R^2$ | 0.8230 | 0.8149 | 0.8016 | 0.7975 | 0.7341 | 0.7757 | 0.7437 | 0.6949 |
| | | RPD | 2.407 | 2.336 | 2.277 | 2.252 | 1.528 | 2.007 | 1.713 | 1.593 |
| | SPOT6 | RMSE | 7.032 | 7.601 | 7.213 | 7.896 | 10.230 | 11.655 | 10.237 | 10.026 |
| | | $R^2$ | 0.8755 | 0.8576 | 0.8693 | 0.8431 | 0.7486 | 0.6726 | 0.7474 | 0.7572 |
| | | RPD | 2.870 | 2.655 | 2.798 | 2.556 | 2.020 | 1.773 | 2.018 | 2.061 |

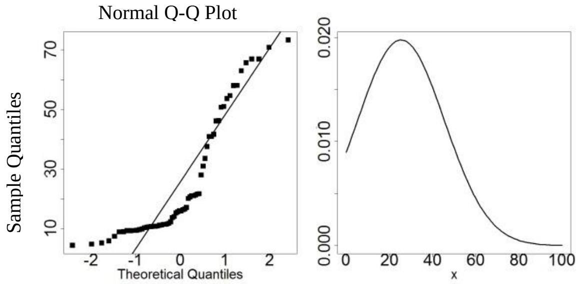


Figure 5 Statistical analysis of samples（Left:QQ figure Right:Normal distribution map）

### 4.2 FVC Model calibration

The results of estimating vegetation coverage with PLSR and SVR are shown in table 3. For GF-1 data, there are better correlative correlation between NDVI and EVI and FVC. The PLS with RPD value around 2 produce relative lower accuracy than SVM with RPD value larger than 2.4. The enhanced vegetation index (EVI) from GF-1 can produce high accuracy ($R^2$=0.8149，RPD=2.336，RMSE=8.694） based on SVM model, meanwhile, the normalized difference vegetation index (NDVI) from SPOT-6 data can produce high accuracy（$R^2$=0.8755，RPD=2.870，RMSE=7.032）based on SVM model as well (Figure 6). However, DPM method produces the lower accuracy than SVM and PLSR model, which generate $R^2$ of 0.62, RMSE of 13.91 for GF-1 data and $R^2$ of 0.64, RMSE of 20.71for SPOT-6 data (Figure 7).

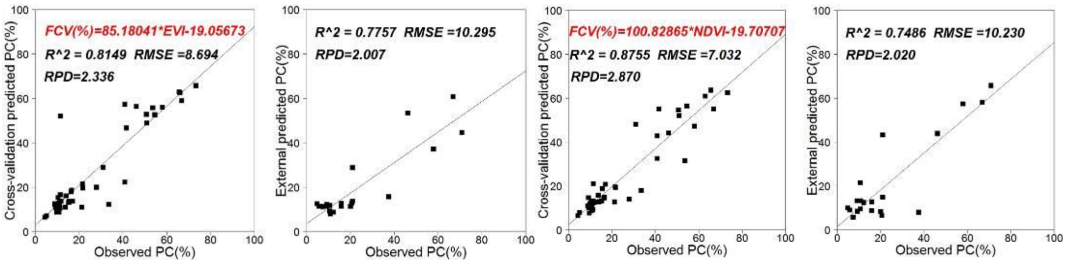


Figure 6 Scatter plots of SVM model（Left:GF-1(EVI) Right:SPOT6(NDVI)）

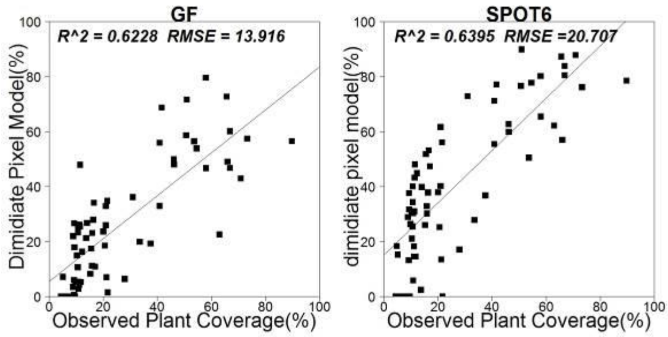


Figure 7 Scatter plots of dimidiate pixel model（Left:GF-1 Right:SPOT6）

### 4.3 Error propagation model validation

As shown in Figure 8, there is significant difference between FVC from DPM and SVM, PLSR, particularly at the level over 60% vegetation cover. For DPM method, all pixels were classified as vegetation ($NDVI_{veg}$) and un-vegetation ($NDVI_{soil}$). Once one pixel contains various type of vegetation or land cover, it would result in the lower accuracy of estimated FVC. The estimation coverage at each level from SVM and PLSR method is similar for SPOT 6 data.

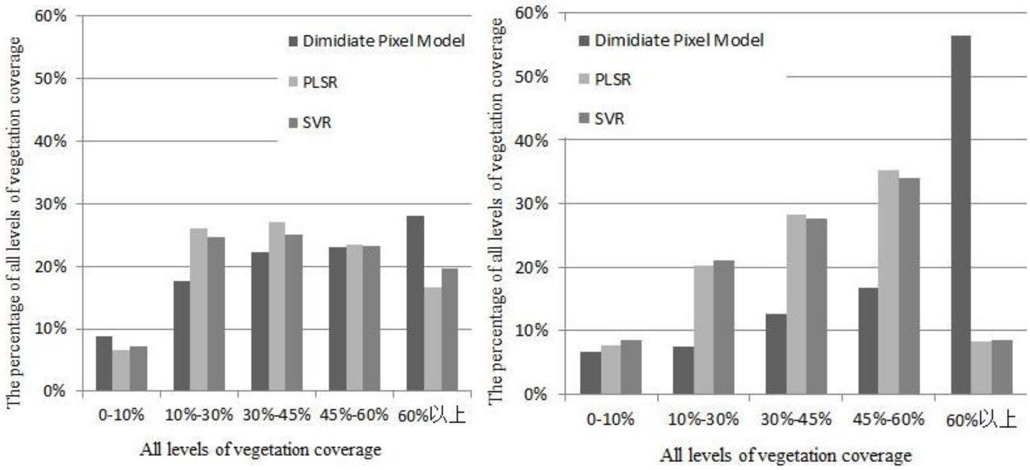


Figure 8 Histogram of vegetation cover in different levels （Left:GF-1 Right:SPOT6）

Monte Carlos simulation was conducted to evaluate the influence of spatial resolution on the precision of estimation vegetation coverage. In Monte-Carlo simulations, the training samples were manipulated using RMSE from the cross-scale error propagation. This step was performed 10,000 times. Cross-scale RMSE of vegetation coverage is mainly at 11-16 for PLSR method and 12-15 for SVM method respectively. The error accumulation would increase with cross-scale error propagation. The frequency value of SVM started to reduce dramatically after the RMSE of 14.Thus, given to different resolution (SPOT6 to GF-1) in FVC estimation, The SVM method can produce lower RMSE value rather than PLSR method.

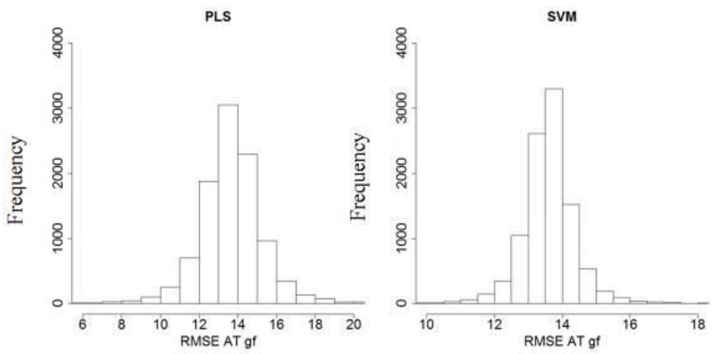


Figure 9 The result of Monte Carlo simulation

## 5. CONCLUSION

In this study, we presented the plant coverage product for the grasslands of the mine affected area in Hulunber meadow steppe to be fully validated against ground data sampled in field. We tested three different methods (DPM, PLSR and SVM regression) in a cross-scale cascade of satellite data (Chinese GF-1 and SPOT-6) with decreasing pixel resolution. It was found that the methods using only satellite reflectance values (PLSR, SVM regression) yielded better results than those based on DPM. Chinese GF-1 data can provide grassland cover with high accuracy based on SVM model. Because the error rates of the SVM regression models were low, the final product may help stakeholders at local and regional scales to define livestock storage capacities or evaluate the mine affected pastures.

Furthermore, the plant coverage data may serve as data sources for future estimations of LAI, biomass and primary production. Such datasets are urgently required to improve the simulation of vegetation dynamics on the mine affected grassland in north prairie.


**ACKNOWLEDGEMENT**

This work was supported by National Natural Science Foundation of China for young researchers (41401233)


**REFERENCE**


Baxendale, C. L., N. J. Ostle, C. M. Wood, S. Oakley, and S. E. Ward. 2016. Can digital image classification be used as a standardised method for surveying peatland vegetation cover? Ecological Indicators **68**:150-156.

Boulet, J.-C., D. Bertrand, G. Mazerolles, R. Sabatier, and J.-M. Roger. 2013. A family of regression methods derived from standard PLSR. Chemometrics and Intelligent Laboratory Systems **120**:116-125.

Demirel, N., M. K. Emil, and H. S. Duzgun. 2011. Surface coal mine area monitoring using multi-temporal high-resolution satellite imagery. International Journal of Coal Geology **86**:3-11.

Gessner, U., M. Machwitz, C. Conrad, and S. Dech. 2013. Estimating the fractional cover of growth forms and bare surface in savannas. A multi-resolution approach based on regression tree ensembles. Remote Sensing of Environment **129**:90-102.

Huete, A. R. 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment **25**:295-309.

Jia, K., S. Liang, X. Gu, F. Baret, X. Wei, X. Wang, Y. Yao, L. Yang, and Y. Li. 2016. Fractional vegetation cover estimation algorithm for Chinese GF-1 wide field view data. Remote Sensing of Environment **177**:184-191.

Laliberte, A. S., A. Rango, J. E. Herrick, E. L. Fredrickson, and L. Burkett. 2007. An object-based image analysis approach for determining fractional cover of senescent and green vegetation with digital plot photography. Journal of Arid Environments **69**:1-14.

Lehnert, L. W., H. Meyer, Y. Wang, G. Miehe, B. Thies, C. Reudenbach, and J. Bendix. 2015. Retrieval of grassland plant coverage on the Tibetan Plateau based on a multi-scale, multi-sensor and multi-method approach. Remote Sensing of Environment **164**:197-207.

Liu, S., D. A. Roberts, O. A. Chadwick, and C. J. Still. 2012. Spectral responses to plant available soil moisture in a Californian grassland. International Journal of Applied Earth Observation and Geoinformation **19**:31-44.

Perry, C. R., and L. F. Lautenschlager. 1984. FUNCTIONAL EQUIVALENCE OF SPECTRAL VEGETATION INDEXES. Remote Sensing of Environment **14**:169-182.

Schwieder, M., P. Leitão, S. Suess, C. Senf, and P. Hostert. 2014. Estimating Fractional Shrub Cover Using Simulated EnMAP Data: A Comparison of Three Machine Learning Regression Techniques. Remote Sensing **6**:3427.

Sohn, Y., and R. M. Mccoy. 1997. Mapping desert shrub rangeland using spectral unmixing and modeling spectral mixtures with TM data. Photogrammetric Engineering & Remote Sensing **63**:707-716.

Urban, M., S. Hese, M. Herold, S. Pöcking, and C. Schmullius. 2010. A fractional vegetation cover remote sensing product on pan-arctic scale with link to geotiff image. Supplement to: Urban, Marcel; Hese, Sören; Herold, Martin; Pöcking, Stefan; Schmullius, Christiane (2010): Pan-Arctic land cover mapping and fire assessment for the ESA Data User Element Permafrost. Photogrammetrie Fernerkundung Geoinformation, 4, 283-293, doi:10.1127/1432-8364/2010/0056. PANGAEA.

Xie, Y. C., Z. Y. Sha, and M. Yu. 2008. Remote sensing imagery in vegetation mapping: a review. Journal of Plant Ecology-Uk **1**:9-23.

Zou, X. C., W. Su, Y. Y. Chen, L. Li, and D. L. Li. 2010. Estimation of Fractional Vegetation Cover in Opencast Coal Mine Dump Area Using Landsat TM Data. Sensor Letters **8**:81-88.