ESTIMATION OF GRASS YIELD IN LARGE REGION ON GEOGRAPHICALLY WEIGHTED REGRESSION MODEL

Luo Chengfeng [a], *, Yu Xiujuan [b], Liu Caijuan [c], Du Yingkun [c]

[a] Institute of Photogrammetry and Remote Sensing, Chinese Academy of Surveying and Mapping, Beijing 100830, P.R. China – chfluo@ casm.ac.cn

[b] National Calibration Center for Electro-optical Distance Meter, Chinese Academy of Surveying and Mapping, Beijing 100830, P.R. China –xjyrspxf@163.com

[c] Geomatics of College，Shandong University of Science and Technology，Qingdao 266590, P.R. China –liucaijuan15@163.com

[d] College of Information Science and Engineering ，Shandong Agricultural University，Taian 271018, P.R. China - dyk9118@126.com

Commission VI, WG VI/4

KEY WORDS: Geographically Weighted Regression (GWR) Model；Grass Yield；QingHai Province；Remotely Sensed Data

ABSTRACT:

The grass yield embodies its productivity，and also is ground for developing animal husbandry production management. Now the remote sensing technology has been becoming an efficient and feasible mean to estimate the grass yield. In the study, the thought about Geographically Weighted Regression (GWR) was involved in estimating the grass yield. The special characteristics of samples measured on field were considered, and then each sample has a local function covering area around. And the parameters for the function are decided by the weighted function which is associated with the spatial distance between the sample and others around. GWR is a good solution to the model without spatial stationarity, as a consequence a significant model-fitting degree comes out. Based on GWR model an ideal production of grassland can be estimated. In this study, Qinghai province, about 0.72 million square kilometres, was taken as an example. The province is an important one on the Qinghai Tibet Plateau. Here the grassland not only closely relates with the local animal husbandry economy, but also directly affects the regional ecosystem security. Landsat TM data in 2013 and samples on field were used to estimate the production. As input parameters, OSAVI and FVC have high correlation coefficient more than 97% with grass yield. There were 201 samples involved in modelling, and the accuracy is 87.27%, above about 47% than that of multiple linear regression model, a widely used traditional statistic model. Another 220 samples were used to verify the results, and here the accuracy can reach 81.3%. Out results indicated that in 2013 the yield of grass in Qinghai province is $1.018*10^8$ ton. The difference between our data and that from professional sector is less than 10%.

1. INTRODUCTION

As the natural resources for livestock production forage, the productivity of grassland can reflect the capabilities of pasture carrying directly. There have many studies about estimating grass yield by remote sensing technology（Jin Y X，2011；Benie G.B.，2005；Liu X Y，2010）. Issues in focus include how to estimate the grass yield quickly and efficiently, monitor the productivity of grassland in a large region and know the current situation and development trend well. Such issues has becoming key topic in ecology and grassland academia field. It is also practical problem for grassland management urgently needed to be solved (Xu B, 2007). To estimate grass yield there are two models used widely, biological - physical model and statistical empirical model（Lv H Y，2010）. The former is always adopted to work for large scale region with coarse resolution data, like NOVAA and MODIS. And the result has low accuracy. When it comes to estimation with high accurate for local region, the modelling is becoming more complex and affected by the natural characteristics of the geographical closely（Chen J，2009）. Establishing the model based on ground truth data and remote sensing data, the statistical empirical model is relatively simple and practicable. Because the spatial heterogeneity in nature always is ignored in practice,

to get high estimation accuracy is difficult（Tao W G，2007）. According to the first law of geography: the closer the distance between the features, the greater similarity （ Miller H J， 2004）. It is necessary to take the rule into modelling in order to get better result.

According to the first law of geography: the closer the distance between the features, the greater similarity （ Miller H J， 2004）. It is necessary to take the rule into modelling in order to get better result. Geographically Weighted Regression (GWR) model create the conditions for regression analysis of the relationship between spatial features. The grass yield has relation with geographical location, adjacent relationship between features with geographical proximity makes the grass yield spatial correlation. In the study the GWR was adopted to model, which can reflect spatial characteristics of grass yield. Another valuable side is the estimation was applied using imageries with30 m resolution in a large region, about 720,000 square kilometres.

---

* Corresponding author. chfluo@casm.ac.cn , +86-10-63880548, +86 135 2276 0996.

2. DATA AND METHOD

2.1 Study Area

Qinghai Province, one of the four largest provinces in China, extending 8 degree of latitude. It also is one of the national five major pastoral areas, the grassland area here accounting for 50.46% of the total land area. The grassland resources here is very abundance, and available pasture area accounts for 87% of the province's natural grassland area, about 15% of the available grassland area in China (Shen Y C, 1991). Grassland mainly appeared in Southern Qinghai Plateau, Qilian Mountain and the south-eastern margin of the Qaidam basin mountain. There are 9 Grassland categories，including alpine meadow, alpine grassland, alpine meadow grassland, alpine desert, temperate steppe, temperate desert, temperate desert steppe, mountain meadow and lowland meadow. Here alpine meadow and alpine grassland present the main body of natural grassland, about $2948.16 \times 104hm^2$ and accounting for 80.88% of the province's total area of grassland. With remote sensing technology the grass yield can be monitored macro and comprehensively. It has great significance during balancing the livestock development and ecological construction (Yan D L, 2007).
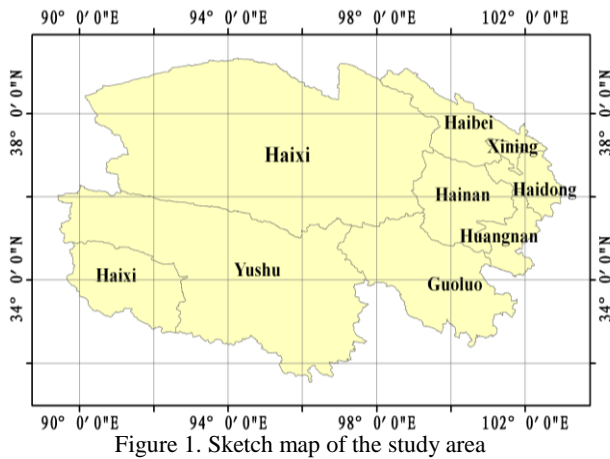

Figure 1. Sketch map of the study area

2.2 Data and Preprocessing

Landsat 8 TM data with resolution 30m was used in the study. According to the growth of grass is best in august each year, the images, receiving before and after a month from august 15th, were selected to work. There involves 46 scenes in all.

The field data comes from the ground field survey by the local grassland supervising department in 2013 august. There are about 600 typical and representative samples totally.

The preprocessing included radiometric calibration, atmospheric correction, terrain correction and removing cloud. Here the FLASH was used to finish atmospheric correction, C-correction model (Wang S, 2013) to terrain correction. The method referred from Li B X (2010) was used to remove the effects of cloud.
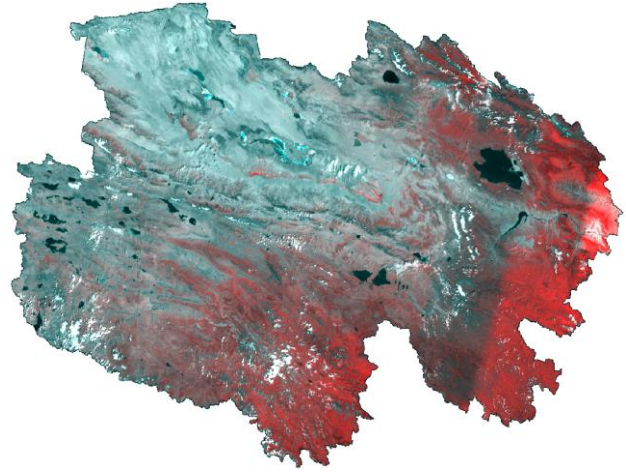

Figure 2. The mosaic image of study area after preprocessing

2.3 Method

GWR model is an expansion on the traditional model, and the spatial characteristics of the data are introduced the model. The geographic location of sample are involved into modelling based on multivariate linear regression. And the regression coefficient is assumed as the function of the location of sample. While modelling, a local function for every observation sample is calculated by GWR. And the coefficients are measured by weighting function among adjacent relationship of the samples. The non-stationary of the model is solved well by GWR. Thereby the goodness of fitting of the model is improved and better simulation results can be made.

GWR model can be expressed as following (Xuan H Y,2007):

$$y_i = \beta_0(u_i,v_i) + \sum_{k=1}^{p} \beta_k(u_i,v_i)x_{ik} + \varepsilon_i, i=1,2,\cdots,n \quad (1)$$

Where（$y_i; x_{i1}, x_{i2}, \cdots, x_{ip}$）is the observation value of dependent variable y and independent variable $x_1, x_2, \cdots, x_p$ in the position (u_i, v_i) （$i=1,2,\cdots,n$）. $\beta_j(u_i,v_j)(j=0,1,\cdots,p)$ is the ith observation point unknown parameter in the position (u_i,v_i), and it is an arbitrary function of (u_i,v_i) l $\varepsilon_i(i=1,2,\cdots,n)$ are error independent and identically distributed, generally are assumed to obey $N(0,\sigma^2)$. According to Tobler's first law of geography, the influence on the estimation of parameters for point i from the observation value near point i on is greater than influence from that far away from i, GWR model based on linear regression model assumes that the regression coefficient is an arbitrary function of observation point location and it brings the spatial character into the model. The model calculates a local equation at each point, and observation value weight is no longer remain constant in the regression process, and the weight is related to the proximity position to i:

$$\hat{\beta}(u_i,v_i) = (X^TW(u_i,v_i)X)^{-1}X^TW(u_i,v_i)Y \quad (2)$$

Where $\hat{\beta}$ is the estimation of β, the first row elements of the independent variable matrix are 1; X is the independent variable matrix about model factors; Y is the vector of the

values of the dependent variable and it is the matrix about the grass yield ground measured data; W is a square matrix of weights relative to the position of (u_i, v_i) in the study area; in practices the calculation method of spatial weights matrices are the Gauss distance, exponential distance and tribute distance. This study chose Gauss distance to determine the weight:

$$W(u_i, v_i) = e^{-\frac{1}{2}(\frac{d_{(u_i, v_i)}}{b})^2} \qquad (3)$$

where $W(u_i, v_i)$ is the geographical weight of the ith observation in the dataset relative to the location (u_i, v_i), $d_{(u_i, v_i)}$ is some measurement of the distance between the ith observation and the location (u_i, v_i), b is a quantity known as the bandwidth.

2.4 Input Parameter

To estimate grass yield with remote sensing data, vegetation index is adopted by many researches（Skinner R H，2011，JOSEM.PARCEIO，1997）. In the same time some weather elements, like temperature, rainfall, dryness also be used as input parameters（Li W J, 2012; Wei Y X, 2012）. Generally correlation coefficient（Xu H L，2013） and estimation accuracy are two important index to compare and choose the input parameters. In this study, five vegetation indices were compared, including normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), soil adjusted vegetation index (SAVI), modified soil adjusted vegetation index (MSAVI) and optimal soil adjusted vegetation index (OSAVI)（Chen P F, 2007）. After several tests with five VIs, weather elements and fraction of vegetation coverage (FVC), finally EVI and FVC was adopted as input parameter to estimate grass yield.

The method mentioned by Brian Johnson（2012）was used to estimate the FVC, and the accuracy for the entire study area in August 2013 is 85.8%.
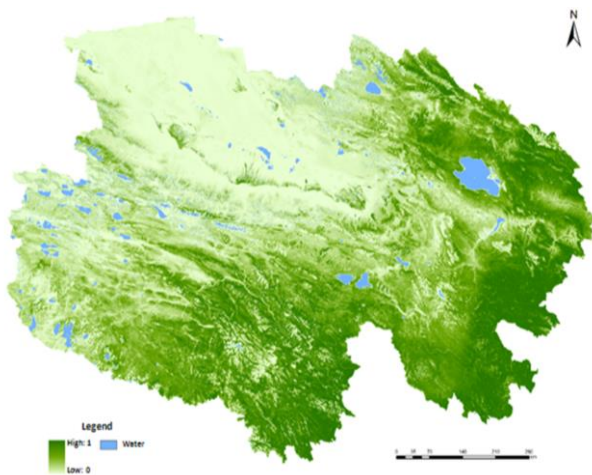

Figure 2. The FVC map of QingHai Province in 2013 August

2.5 Modelling with GWR

To model with GWR, the fresh weight of grass yield measured on the spot is chosen as the dependent variable and other factors include EVI and FVC are independent variables. After modelling, there is an equation for every samples calculated by GWR. While there is only one equation if the binary linear regression model is used.

A total of 201 field samples involved in grass yield estimation model building, the parameters of GWR model change in different geographical locations. As table 2，the parameters of GWR model change with different geographical locations.

Variable	Minimum	Lwr Quartile	Upr Quartile
EVI	-22.94947	-0.29945	1.9236
VFC	-26.45646	-0.4415	1.31991
Constant	-2.71119	2.37819	3.22238
Variable	**Maximum**	**Median**	**Mean**	**Std**
EVI	52.29209	0.5347	1.0831	6.59432
VFC	10.59244	0.2882	0.27507	3.54877
Constant	11.58604	2.8394	2.77774	1.14701

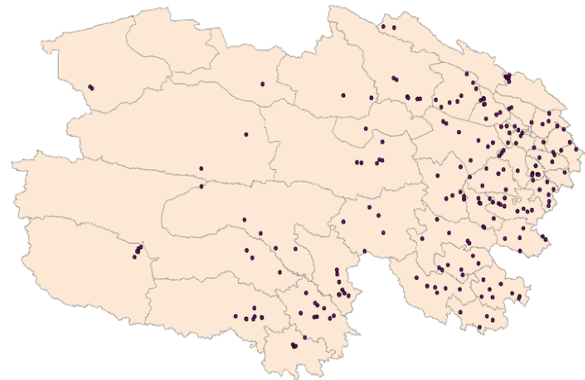Table 1. The range change of parameters from GWR model

Figure 3. The distribution map of the samples for modelling

The multiple linear regression model is as following:

$$Y = 0.684X_1 - 0.006X_2 + 2.706 \qquad (4)$$

Here Bandwidth has a great influence on the operation results of GWR model. the bigger bandwidth the bigger smoothing. As the bandwidth gets larger the weights approach unity and the local GWR model approaches the global OLS model. After several tests and analysis of test results, we use the bandwidth 230km and the corrected Akaike Information Criterion (Hurvich et al, 1998) extensively in GWR for as the measurement of goodness of fit Model establishment and Precision verification With the 201 samples, the modeling accuracy for GWR is 87.3%, while the accuracy for OLS is 40.35%. When it come to the statistical results of model parameters, it can be found the statistical index for GWR are better than that for OLS. Here, AICc is short for Akaike Information Criterion, r is correlation coefficient, and r^2 the coefficient of determination, r^2 Adj Adjusted r-square.

Variable	Sigma	AICc	r	r^2	r^2Adj
OLS	0.193	-86.437	0.815	0.665	0.663
GWR	0.012	749.849	0.987	0.973	0.891

Table 2. The Diagnostic statistics for parameters from GWR

2.6 Grass yield estimating

While estimating the grass yield for the whole region, the input parameters included raster data of the EVI and FVI. The value was calculated pixel by pixel. To find the nearest sample away the current pixel, and take the fitting equation of the nearest sample as basis to figure the grass yield for the pixel. For large region, there is huge amount of data, and the key to improve efficiency is to identify the nearest samples in shortest time. After repeated practice, it was found that building four forks tree index for all samples can improve the operation rate obviously.

The estimation result of grass yield in 2013 for Qinghai province is show as figure 4. Here from southeast to northwest the grass yield is becoming lower. The region with high grass yield distributes in southeast, including Haibei, Haidong, Huang Nan, Guoluo, and the south-eastern region of Yushu. And Haixi Autonomous region is desert and wilderness, so there has lowest grass yield. According the estimation, the total grass yield of Qinghai province is 1.018×108 Ton in 2013. Here the non-grass area was not taken into operate and such area were identified by the thematic map of second grassland survey in Qinghai Province.
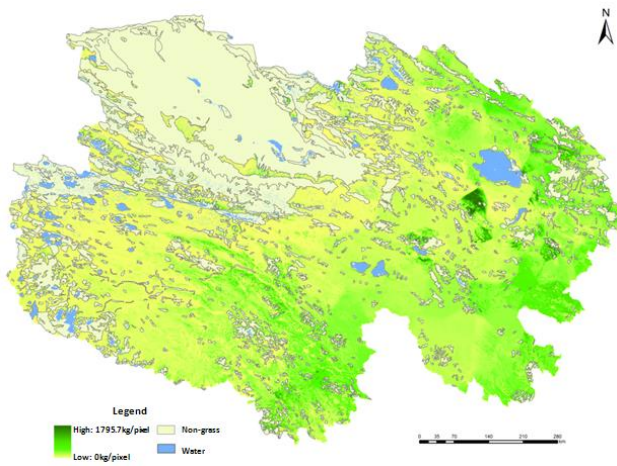

Figure 4. The grass yield map of QingHai Province in 2013

2.7 Precision verification

To verify the accuracy, root mean squared error (RMSE) (Jin Y X, 2011) was used. Theoretically, if RMSE=0, the model is error-free and it is considered an ideal state. In practical applications, the closer predictive value to the critical value, the better the model fitting results. The actual estimation accuracy of the model is 1-RMSE.

There are 210 ground truth samples involved into precision verification and the result value is 81.3%. The total grass yield estimated by the study were also compared with that from local Grassland professional department based on ground survey. The difference is less than 10%.

3. CONCLUSIONS

The location is introduced to building estimation model based on GWR. The parameters for different geographical samples are diverse, which reflect the spatial difference of these points. That means the same factors have different influence to the dependent variable because of deferent geographical position. The spatial difference of the grass yield in a large region can be reflected adequately by GWR. The experiments in the study have shown with GWR the goodness-of-fit of the model can be improved obviously, In the same time the estimation accuracy and verify accuracy are more than 80%. The difference between estimation with remote sensing data and ground survey is less than 10%. When it comes to long-term continuous monitoring in large area, technical method can be used.

To estimate grass yield with GWR model and remote sensing data, the estimation accuracy depends on ground samples. The accuracy is affected directly by spatial distribution characteristics of ground samples, as well as their representation of grassland production capacity. The high accuracy in the study benefited from the typical and representative samples. Now it takes grassland specialized department a lot of manpower and material resources every year to carry out the ground surveys. With the method mentioned above, only using RS data combined with the ground samples, it can be explored whether the ground samples can be reduced and how many samples is minimum requirements within an acceptable estimation error range. Such work will alleviate amount of field work for grassland specialized department.

Based on the grass yield estimated quickly with remote sensing technique, It is very meaningful to carry further research about expertly valuing the service function of the grassland ecosystem near real-time. Both from theoretical and practical view, such study is helpful for such work, like understanding the region ecological importance, scientific management and rational utilization of grass resources , maintaining grassland ecological balance, reasonable arrangements for livestock production, protection grass ecosystems and establish a comprehensive economic accounting system. The following work will focus on ecological evaluation.

REFERENCES

Benie G.B., Kabore S.S., Goita K., et al. Remote sensing-based spatio-temporal modeling to predict biomass in Sahelian grazing ecosystem. Ecological Modeling, 2005, 184:341-354.

Brian Johnson, Ryutaro Tateishi, Toshiyuki Kobayashi. Remote Sensing of Fractional Green Vegetation Cover Using Spatially-Interpolated Endmembers [J], Remote Sensing, 2012, 4: 2619-2634.

Chen J, Gu S, Shen M, et al, Estimating aboveground biomass of grassland having a high canopy cover: an exploratory

analysis of in situ hyper spectral data[J].International Journal of Remote Sensing, 2009, 30（24）: 6497-6517.

Fotheringham A S. Chris Brunsdon, Martin Charlton. Geographically Weighted Regression: the Analysis of Spatially Varying Relationships [M]. U. K. West Sussex: John Wiley & Sons Ltd, 2002.

Hurvich CM, Simonoff JS and Tsai C9L 1998 Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion Journal of Royal Statistical Societ y. B 60 2719293.

Jin Y X, Xu B, Yang X C, et al. Remote sensing dynamic estimation of grass production in Xilinguole, Inner Mongolia[J]. Science China, 2011, 41: 1185–1195.

Josem. Parceio, Howard E. Epstein, W .J. Iam K. Lauenroth, et al, A NPP Estimates From NDVI for the Central Grassland and Region of The United States[J], Ecology, 1997, 78(3): 953-958.

Li B X, Ma Z B, Qi Q W, et al. Cloud and shadow removal from Landsat TM data [J]. Journal of Remote Sensing, 2010, 14(3), 540-545.

Li W J, Jiu C L, Tan Z H, et al. Natural grassland productivity and the livestock-feeds balance in Qinghai Province [J]. Resources Science, 2012, 34(2):367-372.

Liu X Y, Feng Q SH, Liang T G, et al. Spatial-temporal dynamic balance between livestock carrying capacity and productivity of rangeland in Gannan of Gansu Province, China [J]. Chinese Journal of Grassland, 2010, 32(1):99-106.

Lv H Y. Study on the method of grass yield model [D].Beijing: Chinese Academy of Agricultural Sciences, 2010.

Miller H J. Tobler's first law and spatial analysis [J].Annals of the Association of American Geographers, 2004, 94(2):284-289.

Qu C P, Guan D X, Wang Z A, et al. Comparison of grassland biomass estimation models based on MODIS data [J]. Chinese Journal of Ecology, 2008, 27(11): 2028-2032.

Shen Y C, Xiang L P. The physical geography of Qinghai Province [M].Beijing: China Ocean Press, 1991.

Skinner R H, Wylie B K, Gilmanov T G. Using normalized difference vegetation index to estimate carbon fluxes from small rotationally grazed pastures[J].Agronomy Journal, 2011, 103: 972-979.

Tao W G. Remote sensing methods for monitoring grassland production in Inner Mongolia [D].Beijing: Chinese Academy of Agricultural Sciences, 2007.

Wang S, Li X, Zhang Q, et al. Analysis on the applicability of topographic correction models for Landsat images [J].Journal of Huazhong Normal University (Nat. Sci.)2013, 47(4):571-577.

Wei Y X, Wang L W, Shi Y C. Net Primary Productivity of Grassland Resources Monitoring Based on Remote Sensing Data in Qinghai Province, Scientia Geographica Sinica, 2012, 32(5):621-627.

Xu B, Yang X C, Tao W G, et al. remote sensing monitoring upon the grass production in China [J] Acta Ecologica Sinica, 2007, 2007(2): 405-413.

Xu H L, Li J P, Feng J. The pair-wise deletion correlation coefficient testing method and its applications [J]. Acta Meteorological Sinica, 2013, 71(5):901-912.

Xuan H Y, Li S P, Liu S Q. The geographically weighted regression model and its fit methods [J]. Journal of Gansu Sciences, 2007, 19(1):51-52.

Yan J W, Li Ch E , Yuan L, et al. Application Summary of EOS- MODIS data in the monitoring of grassland resources [J]. Pratacultural Science, 2008, 25(4):1-9.

Yan L D, Yin Q J, Zhang H Z.et al. The applied research of remote sensing material in Qinghai grassland resources monitoring and appraisal [J]. Journal of Natural Resources, 2007, 22(4): 640-648.

Zhang Y S, Zhao X Q, Zhou X M. Strategy and counter measure for sustainable development of animal husbandry in Qinghai [J]. Journal of Natural Resources, 2000, 15 (4): 328-334.