

URBAN VEGETATION MAPPING BASED ON THE HJ-1 NDVI RECONSTRUCTION

Feng Li^{1*}, Liu Han², Zhu Liujun³, Huang Yinyou⁴, Guo Song¹

¹ Department of Geographic Information Science, Hohai University, Nanjing, 211000, China

² Geography Graduate Group, University of California, Davis, 95616, United States

³ Department of Civil Engineering, Monash University, 3800, Australia

⁴ Department of Geographic Information Science, Nanjing University, Nanjing, 210023, China

Commission VIII, WG VIII/8

KEY WORDS: HJ-1 NDVI, Urban vegetation mapping, Nanjing city, Filtering algorithm

ABSTRACT:

HJ-1A/B NDVI (HJ NDVI) time-series data possess relatively high spatio-temporal resolution which is significant for the research on urban areas. However, its application is hindered by noise resulting from the restrictions of imaging quality and limits of the satellite platform. The NDVI noise reduction is necessary. Some noise-reduction techniques including the asymmetric Gaussian filter (AG), the double logistic filter (DL), the Savitzky-Golay (S-G) filter and the harmonic analysis (Hants) of NDVI time-series have been used to carry out the NDVI time series reconstruction, and based on the comparison results of different filter, S-G filter is the optimal in the application on urban areas. Finally, urban vegetation mapping is carried out based on the new HJ NDVI.

1. INTRODUCTION

Vegetation is of particular interest as it presents a versatile resource for effectively managing and moderating a variety of problems associated with urbanization. The spatial distribution and abundance of urban vegetation are recognized as a key factor influencing numerous biophysical processes of the urban environment [1,2]. As the understanding of ecosystem services is evolving, researchers are becoming increasingly aware of the importance of urban vegetation. In recent years, remote sensing technology provides a deal of convenience for urban vegetation information acquisition [3,4,5] and mapping [6].

NDVI time series data acquired by satellite sensors can reflect terrestrial vegetation growth status, seasonal aspect, and inter-annual variation accurately. It has been widely used in global and regional ecological environment variables monitoring and simulation, dynamic changes of vegetation cover research, vegetation phenology feature recognition and information extraction, and many other fields [1, 2]. The most used NDVI time series is from NOAA AVHRR, SPOT VEGETATION and TERRA / AQUA MODIS and other sensors [7]. However, limited by their relatively low spatial resolution, these NDVI time series products can no longer satisfy the fine scale researches especially for urban research. Therefore, NDVI time series products with high spatial-temporal resolution are necessary for urban vegetation research. The satellite of HJ-1 A/B was launched by China in 2008. Since the launch date, the remote sensing images of HJ have constituted nearly eight years of time-series data making it possible for urban application research due to relative high spatial and temporal resolution. However, limited by restrictions of HJ satellite platform, there are much noise in HJ images which can lead to errors for application research. Therefore, it is necessary to carry out the HJ NDVI time series reconstruction. For nearly three decades, many researches focus on the establishment of high-quality NDVI time series using different filtering methods such as Savitzky-Golay (S-G) [8], Gaussian fitting method (AG) [9], the double logistic curve fitting method (DL) [10] and the harmonic analysis of NDVI time-series (Hants) [11]. However, these researches focus on the filtering effect of specific filter. Though there are researches which discussed the pros and cons of

different filtering algorithms through quantitative comparative analysis [7, 8, 12, 13], it is lack of the specific application objective and an evaluation framework in urban area using relatively high-resolution NDVI time series.

In this letter, taking the typical vegetation region in Nanjing City as the study area, four filtering methods including S-G, AG, DL and Hants are used for HJ NDVI time series reconstruction. For comparing the filtering results, the evaluation framework from three points of view is proposed including pure vegetation sample points, the classification accuracy of the typical vegetation, as well as the relationship with MODIS NDVI time-series to test the reliability of these four filters. This letter aims to provide a reference for NDVI time series reconstruction in fine spatial scale application.

2. RESEARCH AREA AND DATA SOURCE

2.1 Research Area

The urban area of Nanjing City, Jiangsu Province (Fig. 1) is selected as the case study area. Nanjing, is located in the largest economic zones of China, the Yangtze River Delta, as part of the downstream Yangtze River.

It's extremely scorching in summer and frigid in winter, and the temperature gap turns out to be wide each year. Summer witnesses the largest precipitation. Due to the specific ecosystem, vegetation phenology is significant in this area. And the study area is characterized by a great variety of plants and tree species, including high trees, short-cut trees, shrubs and grassland which means that the several phenological patterns occur in this area.

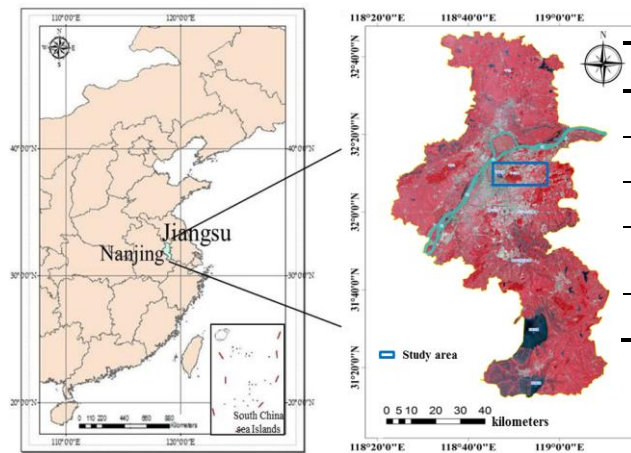


Fig.1. The location of study area

2.2 HJ-1A/B images, field data and data pre-processing

77 HJ-A/B images with the uniform interval of 5 days are selected for this research. These images were acquired on the whole of year 2013 which cover the study area with the size of 771×391 pixels. The images have a spatial resolution of 30 m and a breadth of 700 km. Radiometric calibration and atmospheric correction are applied to the images firstly, then, taking HJ satellite CCD image of May 1, 2013 as a benchmark, relative registration is carried out and registration error is controlled in 0.5 pixels.

To better understand the vegetation in the study area and collect some pure vegetation samples for the evaluation of filters' performance a field survey was conducted on April 2014. Specifically, typical urban vegetation samples in Nanjing, i.e. shrub, grassland, evergreen coniferous tree, broadleaved deciduous tree, and evergreen and deciduous broadleaved mixed tree, are collected. The size of each sample is larger than $90 \text{ m} \times 90 \text{ m}$ to avoid the possible error in NDVI time series caused by the pixel offset in the process of remote sensing image registration. The detailed information of a field survey is shown in Fig. 2.

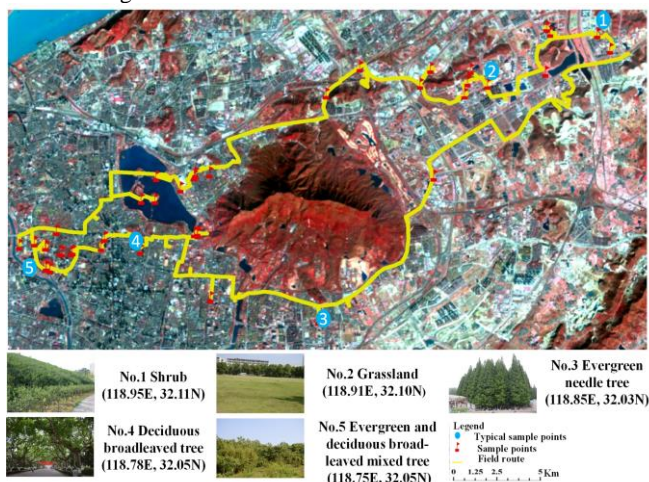


Fig. 2. The route of field survey and the sample points

Combining remote sensing data and meteorological data, "HJ Quality Assurance (HJ QA) is developed analogy to the MODIS Quality Assurance (MODIS QA) which indicates the signal to noise ratio (SNR) of the original remote sensing data [12]. The values of quality level are shown in Table 1; the larger the value of quality level, the higher the signal to noise ratio (SNR) of the original HJ data. In this research, the quality level of grassland is 5 and evergreen coniferous tree is 1.

Table 1 Sample information of pure vegetation points

No.	Vegetation Type	Location	Quality Level	Remarks
1	Shrub	118.95E 32.12N	4	NanJing Univ. Xianlin Campus
2	Grassland	118.90E 32.10N	5	Nanjing Normal Univ. Xianlin Campus
3	Evergreen coniferous Tree	118.85E 32.03N	1	Nanjing Univ. of Science and Technology
4	Deciduous Broadleaved Tree	118.76E 32.06N	3	Southeast Univ. Gulou Campus, Baima Park
5	Mixed Tree	118.76E 32.05N	2	Qingliang Mountain Park

3. METHODOLOGY

In this section, the evaluation framework from three points of view is proposed to evaluate the filtering results.

3.1 Pure Vegetation Sample Points Filtering Analysis

The way of pure vegetation sample points is necessary for evaluating the filters' performance because it is easy to compare with the normal curve and filtering curve of pure pixels. The pure pixels of evergreen and deciduous broad-leaved mixed tree are chosen for this analysis because leaves of this kind of tree germinate in the spring and aging in the autumn, as well as its chlorophyll content can vary with the seasons which will result in the curve of NDVI time series first increase and then decrease.

To evaluate the filtering performance, the statistic methods of mean NDVI, absolute error of the average NDVI, and the correlation coefficient are used. The formula and explanations of the three methods are shown in Table 2. Mean can be used to compare the changing extent of the NDVI value before and after filtering. The higher the value, the better the filtering effect. MAE assesses the fidelity of different filtering algorithms. The lower the value, the better the filtering effect. Correlation Coefficient represents the extent of liner dependence between filtering results and original NDVI time series. The higher the value, the better the filtering effect.

Table 2 The formula and explanations of the statistics

Statistics	Formula	Explanations
Mean	$\frac{\sum_{i=1}^N NDVI_{pi}}{N}$	It stands for the mean NDVI ($p = \text{Raw, S-G 3, S-G 5, AG, DL, Hants}$) before or after the filtering. N is the number of images within a time series; $NDVI_i$ represents value of the i -th image's NDVI.
Mean Absolute Error (MAE)	$\frac{\sum_{i=1}^N NDVI_i - NDVI_{pi} }{N}$	It stands for the average absolute error of NDVI ($p = \text{S-G 3, S-G 5, AG, DL, Hants}$) between the filtering result and the original value. N is the number of images within a time series; $NDVI_i$ represents value of the i -th image's NDVI.
Correlation Coefficient	$\frac{Cov(NDVI_p, NDVI)}{\sqrt{D(NDVI_p)D(NDVI)}}$	It stands for the correlation coefficient of NDVI before and after filtering of a particular method ($p = \text{S-G 3, S-G 5, AG, DL, Hants}$).

3.2 Vegetation Classification Analysis

Phenology feature in terms of NDVI time series of land cover has been widely used for vegetation mapping in various scales [16]. And, the noise containing in NDVI time series is one of the main causes for the vegetation classification error. In turn, the result of vegetation classification can be regarded as an indicator to evaluate the quality of NDVI time series. HJ NDVI time series and hyperspectral data have similar data form, therefore, vegetation classification based on NDVI time series is carried out as the classification based on "simulated hyperspectral data"[17]. Training samples are determined according to the field survey. Principal component transformation is applied to the NDVI time series and three

principal components are selected as the input for supervised classification. Classification accuracy and kappa coefficient are obtained for evaluating the filtering performance of different filters.

3.3 Filtering Results Comparison with MODIS NDVI time series

MODIS NDVI time series has relatively good quality and has been widely used for the evaluation of other NDVI products [7]. Therefore, MODIS NDVI time series in 2013 was used to verify the reliability of the filtering results of HJ NDVI time series. 23 scenes are selected from HJ NDVI time series, the overlap error of which is less than or equal to 3 days in order to have the similar acquisition time with MODIS NDVI time series. However, images of MODIS have a resolution of 250 m. Since HJ-1 satellites' spatial resolution is relatively high, resampling is a compromising approach for this problem. The NDVI map of HJ will be resampled from a 30 m to 250 m resolution referred by the resampling techniques proposed by Ma et al [13]. The pixels with little vegetation area excluded for further analysis by overlaying the resampled vegetation map. And the correlation analysis between these two NDVI time series can be carried out.

4. EXPERIMENTS AND RESULTS

4.1 Experimental Setup

Four kinds of filters including S-G, AG, DL and Hants are used in the experiments, among which the first three can be processed using TIMESAT software [14]. The details of these filters can be found in [8-12, 15]. Parameters setting is shown in Table 3. The parameters of iteration times and the size of sliding window for S-G are important. Here, the size of sliding window is set 3 and 5 respectively according to the ideal filtering effect. The parameters of iteration times and upper envelope fitting intensity should be set for AG and DL. And the parameters of period and maximum error tolerance are needed for Hants. All values of these parameters are provided after extensive experimenting to compare the four filters' performance [17].

Table 3 Parameters setting for different filters

No.	Filter	Parameters
1	S-G	Iteration times (1~3) : 3 size of sliding window : 3
2		Iteration times (1~3) : 3 size of sliding window : 5
3	AG	Iteration times (1~3) : 3 upper envelope fitting intensity (1~10) : 5
4	DL	Iteration times (1~3) : 3 upper envelope fitting intensity (1~10) : 5
5	Hants	period: 73 maximum error tolerance: 1000

4.2 Pure Vegetation Sample Points Filtering Results Analysis

The "bimodal phenomenon" appears in the HJ NDVI time series because of the quality of HJ image, as shown in Fig. 3. All these four filters can reduce the noise to some extent. However, S-G is sensitive to abnormal value, followed by the AG filter, DL and Hants. For S-G, the filtering curves will fluctuate with the abnormal curves when there is much noise and the more significant of filtering effect the smaller of the sliding window size (Fig. 3). NDVI values after filtering by S-G, AG and DL increases significantly; whereas, NDVI values of the filtering by Hants are increased slightly. In addition, comparing the smooth effects of the filtering results, Hants can

get the smoothest result, followed by the AG, DL and S-G. Although Hants can eliminate the "bimodal phenomenon" of HJ NDVI time series, the peak of HJ NDVI time series is also reduced which means that the information of vegetation phenology in the NDVI time series will be weakened.

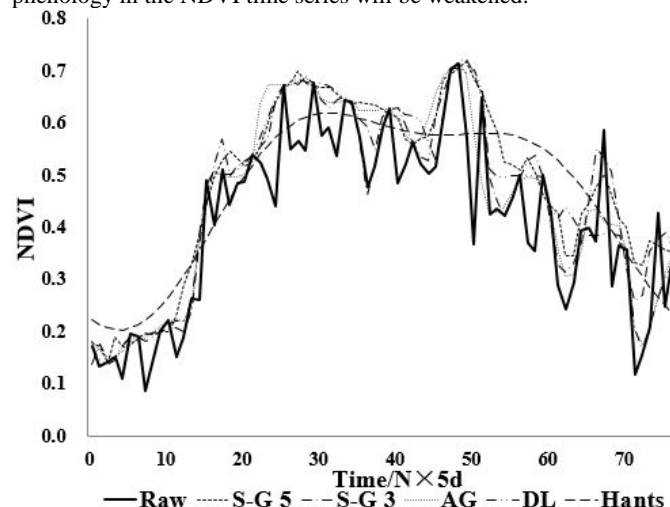


Fig. 3: The original and filtering curves for mixed tree sample points

The NDVI time series of five typical vegetation types, corresponding to the five points mentioned in section II, are constructed. For the pure vegetation pixel points, the mean NDVI values of all vegetation types increases after filtering, as shown in Fig. 4(a). According to the comparison results, the increasing extent of S-G with sliding window size of 5 (S-G 5) is most significant. In addition, the increasing extent of NDVI values after filtering for the evergreen coniferous tree, deciduous tree, evergreen and deciduous broad-leaved mixed tree, shrub and grassland increased rate are 16%, 14%, 14%, 12%, and 9% respectively. The reason can be explained by the characteristics of NDVI value of trees, shrubs and grasslands as well as the principles of filters. Under normal circumstances, the mean NDVI values of trees is larger than those of shrubs and grasslands, while the intensity of the noise does not vary with the NDVI values. The larger the original NDVI values, the smaller the absolute value of the deviation. For the filters, they are subject to repair the reduction introduced by ground water, snow-covered ground, clouds and aerosols and so on [1]. Therefore, for trees, the larger deviation is needed to be repaired than shrubs and grasslands. That is the reason that the NDVI increasing extent of trees, shrubs and grasslands shows the diminishing trend.

According to the comparison of the average absolute error between filtering results and the original NDVI values, AG and DL show more stable results than S-G and Hants, as shown in Fig. 4(b). The average absolute error of filtering results of Hants is the largest because Hants has the strongest smooth effect. In addition, for S-G 5, the mean absolute error of filtering results is lower, the value of quality level of the sample points is larger.

The correlation coefficients of the filtering results and the original data are all greater than 0.78, as is shown in Fig. 4(c). Considering the quality level of sample points, the value of quality level is larger, the correlation between the filtering results and the original data is higher. In this research, for S-G 5, the correlation of filtering results and original data increases when the value of quality level of the sample points is larger.

Based on the quantitative analysis, all filters can reduce the noise to improve the quality of NDVI time series for urban vegetation. 1) The results of S-G 5 show the largest increasing

extent; 2) The correlation coefficients of AG, DL and the S-G 3 show the similar correlation coefficients and are higher than those of filter results of S-G 5 and Hants; The former can more obviously show the growth characteristics of vegetation; 3) Due to Hants' strong smoothness, it can reduce the characteristics of vegetation phenology of the original data besides the noise.

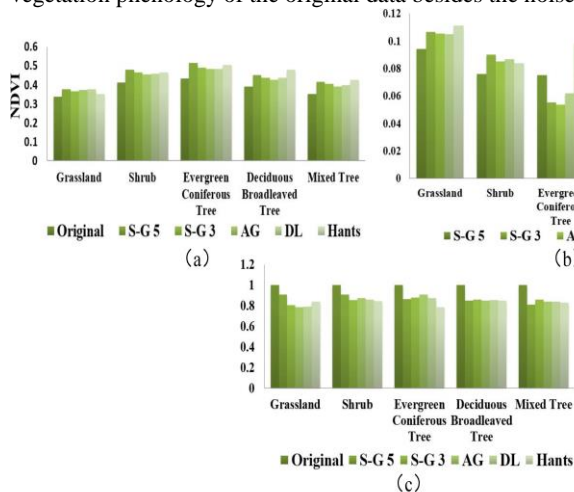


Fig. 4: a) Mean values of the original and filtered NDVI time series; b) MAE of the filtered NDVI time series; c) Correlation coefficient of the original and filtered NDVI time series

4.3 Vegetation Classification Results Analysis

The results of overall accuracy and kappa coefficient for five types of vegetation can be seen in Table 4. O indicates the original data.

Table 4 Overall accuracy and Kappa for five types of vegetation

	O	S-G 3	S-G 5	AG	DL	Hants
Overall accuracy	Shrub	0.55	0.60	0.59	0.58	0.53
	Grassland	0.49	0.57	0.56	0.54	0.50
	Evergreen Coniferous	0.56	0.63	0.69	0.59	0.65
	Deciduous Broadleaved	0.62	0.71	0.75	0.65	0.63
	Mixed	0.58	0.65	0.68	0.56	0.59
	Kappa	0.42	0.57	0.59	0.53	0.43

Some conclusions based on these results are as follows: 1) The classification accuracy of grassland and shrub is lower than that of other vegetation types which means the NDVI of grassland and shrub can be regarded as the coverage of green space; 2) The filter window sizes of S-G have different results. Smaller windows have the stronger ability of preserving detailed information, at the same time, the noise is preserved. Whereas the larger windows have the stronger ability of smoothing, but the big window may cause distortion. In this research, for evergreen coniferous tree and deciduous broadleaved tree, the filter of S-G 5 is better than S-G 3 and for other two types, the filter of S-G 3 can get the better results; 3) The filtering results of AG and DL filter have a great similarity; 4) The filtering results of Hants are similar to the original data. The classification accuracy and kappa for all the five filters are shown in Fig. 5, in which the classification accuracy of S-G 5 is the highest, followed by the S-G 3, the AG, the DL and the Hants which means the S-G 5 is the best one for removing the noise and remaining the details. Overall, the results of 5 filters are better than those of original data which mean that the filtering is an effective way to reduce the noise for improving the classification accuracy of the urban vegetation.

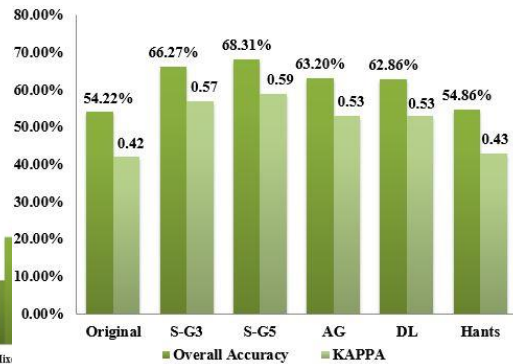


Fig. 5: Overall accuracy and KAPPA of the original and filtered NDVI time series

4.5 Filtering Results Comparing Analysis

As shown in Fig. 6, all the filtering results show the higher correlation coefficients comparing with the MODIS NDVI time series which are larger than 0.85 even when the original correlation coefficient is poor (less than 0.80). Comparing with the four filters, the S-G 5 has the highest reliability because it significantly improves its correlation coefficient (greater than or equal to 10%). In addition, comparing with the filtering results of pure vegetation sample points and vegetation classification, the filter of S-G 5 in this evaluation method still reflects the best result.

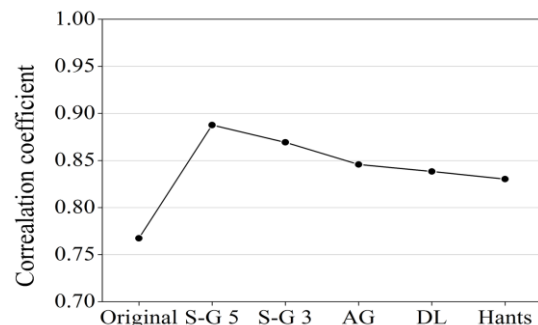


Fig. 6: Correlation coefficient of the MODIS NDVI time series and the HJ NDVI time series

5. CONCLUSIONS

The complexity of vegetation in urban area is more significant than that in non-urban area. Therefore, the high spatio-temporal remote sensing data is necessary. In this research, in order to improve the quantity of HJ NDVI time series in urban areas, the NDVI noise reduction is carried out based on the filters of S-G, AG, DL and Hants. According to the evaluation framework with three evaluation methods, the four filters' performance can be compared and the filter of S-G 5 can get the best results.

In addition, the existing NDVI noise-reduction research revealed the superiority of the filters of AG and DL over others in natural regions [12], whereas, the filter of S-G has the strongest ability of NDVI noise reduction for urban area.

In future work, by increasing SNR of HJ NDVI time series, high quality HJ NDVI time series can be obtained which can better explore the potential application of HJ satellite data.

Acknowledgment

This research was supported by "the National Natural Science Foundation of China (No. 41301446)" and "the Fundamental Research Funds for the Central Universities (2015B11314)". The authors would like to thank all the organizations for sharing data.

References

- [1] S. A. Ackerman, K. I. Strabala, W.P. Menael, et al. Discriminating Clear Sky from Clouds with MODIS. *Journal of Geophysical Research*, 1998, 103(32):141-157.
- [2] F. Yu, K. P. Price, J. Ellis, et al. Satellite Observations of the Seasonal Vegetation Growth in Central Asia: 1982-1990. *Photogrammetric Engineering & Remote Sensing*, 2004, 70(4): 461-469.
- [3] R.T. Tooke, N.C. Coops, N.R. Goodwin, J.A.Voog. Extracting urban vegetation characteristics using spectral mixture analysis and tree classification. *Remote Sensing of Environment*, Vol.113, 2009, pp.398- 407.
- [4] J. Tigges, T. Lakes, P. Hostert. Urban vegetation classification: Benefits of multitemporal RapidEye satellite data. *Remote Sensing of Environment*, Vol.136, 2013, pp.66-75.
- [5] R.L. Pu, S.Landry. A comparative analysis of high resolution IKONOS and World View-2 imagery for mapping urban tree species. *Remote Sensing of Environment*, Vol.124, 2012, pp.516-533.
- [6] M. Immitzer, C. Atzberger, & T. Koukal. Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. *Remote Sensing*, Vol.4, 2012, pp.2661-2693.
- [7] R. Fensholt, K. Rasmussen, T. T. Nielsen, & C. Mbow. Evaluation of earth observation based long term vegetation trends - Intercomparing NDVI time series trend analysis consistency of Sahel from AVHRR GIMMS, Terra MODIS and SPOT VGT data. *Remote Sensing of Environment*, 113, 2009, pp.1886-1898.
- [8] J. Chen, P. Jönsson, T. Masayuki, et al. A Simple Method for Reconstructing a High-quality NDVI Time-series Data Set Based on the Savitzky-Golay Filter. *Remote Sensing of Environment*, 2004, 91:332-344.
- [9] P. Jönsson, L. Eklundh. Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*, 2002, 40 (8), pp.1824-1832.
- [10] P. Beck, C. Atzberger, K.A. Høgda et al. Improved Monitoring of Vegetation Dynamics at Very High Latitudes: A New Method Using MODIS NDVI. *Remote Sensing of Environment*, 2006, 100, 321-334.
- [11] W. Verhoef. Application of Harmonic Analysis of NDVI Time Series (HANTS). In Azzali S, MenentiM (eds).
- [12] N.H. Jennifer, G. J. Mcdermid. Noise Reduction of NDVI Time Series: An Empirical Comparison of Selected Techniques. *Remote Sensing of Environment*, 2008, 113 (1), pp. 248-258.
- [13] H.Y. Ma, J.X. Huang, D.H. Zhu, J.M. Liu, et al. Estimating regional winter wheat yield by assimilation of time series of HJ-1 CCD NDVI into WOFOST-ACRM model with Ensemble Kalman Filter. *Mathematical and Computer Modelling*, 2013, 58, pp. 759-770.
- [14] P. Jönsson, L. Eklundh. TIMESAT A Program for Analyzing Time series of Satellite Sensor Data: User's Guide for TIMESAT 2. 3. Sweden: Malmö and Lund, 2006.
- [15] A. Savitzky, M. J. E. Golay. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 1964, 36, pp.1627-1639.
- [16] Z. H. Xue, P. J. Du, and L. Feng. Phenology-Driven Land Cover Classification and Trend Analysis Based on Long-term Remote Sensing Image Series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(4), pp.1142-1156.
- [17] L. Feng, L.J. Zhu, H. Liu, Y.Y. Huang, P. J. Du. Urban Vegetation Classification based on Phenology using HJ-1A/B time series imagery. *Joint Urban Remote Sensing Event-JURSE 2015*, DOI: 10.1109/JURSE.2015.7120477.