

Remote Sensing Data Products Oriented Quantitative Computing System

—The GSC Best Practice Data Computing Environment 2018

Liu, Q. H.¹ Zhong, B.^{1*} Tang, P.² Zhang, H. H.³ Li, H. Y.² Wu, S. L.¹
Xin, X. Z.¹ Li, J.¹ Jia, L.¹ Shan, X. J.² Zhang, Z.² Wen, J. G.¹
Du, Y. M.¹ Li, L.¹ Yang, A. X.¹ Li, H.¹ Hu, G. C.¹ Zhao, J.¹
Zhang, H. L.¹ Yu, S. S.¹ Dou, B. C.¹ Wu, J. J.¹

1. State Key Laboratory of Remote Sensing Science, Jointly Hosted by Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences and Beijing Normal University, Beijing 100101, China;

2. Institute of Remote Sensing and Digital Earth (RADI), Chinese Academy of Sciences, Beijing 100101, China;

3. Computer Network Information Center of Chinese Academy of Sciences, Beijing 100190, China

Abstract: The “Remote sensing data products oriented quantitative computing system” was recognized by the Geographical Society of China (GSC) as the GSC Best Practice Data Computing Environment 2018. The system was developed based on the project of quantitative remote sensing system and applications using the space-based, flight-based and ground data together, which was supported by the National High Technology Research and Development Program of China. It aims to establish a platform focused on using the multi-source remote sensing data to produce bio-geophysical parameters with higher spatial and temporal resolutions to support diverse applications. The project focused on developing the algorithms for the data normalization from multiple sources and the bio-geophysical parameters’ retrieving; and a software was developed to integrate these algorithms, which was capable of producing more than 20 bio-geophysical parameters. The software system was called Multi-source Data Synergized Quantitative Remote Sensing Production System (MuSyQ). MuSyQ has been operated for 4 years and it has produced parameters at local, regional or global scales, such as China, the Southeastern Asia, Belt and Road area, and the global areas. The total products developed in the MuSyQ computing environment was over 800 TB. It was one of the best practice data computing environments of GSC in 2018.

Keywords: multi-source data; synergize; quantitative remote sensing; data normalization; production system

Received: 20-07-2018; **Accepted:** 17-09-2018; **Published:** 25-09-2018

Foundation(s): Ministry of Science and Technology of P. R. China (2018YFA0605504)

***Corresponding Author:** Zhong, B. L-4528-2016, State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences and Beijing Normal University, zhongbo@radi.ac.cn

Author ID: Liu, Q. H. S-1647-2016; Zhong, B. L-4528-2016; Tang, P. S-3703-2018; Zhang, H. H. S-9762-2018; Li, H. Y. S-9247-2018; Wu, S. L. S-1572-2016; Xin, X. Z. S-3681-2016; Li, J. R-7298-2016; Jia, L. I-1457-2014; Shan, X. J. S-1589-2016; Zhang, Z. S-3669-2018; Wen, J. G. S-9821-2018; Du, Y. M. S-9953-2018; Li, L. S-4533-2016; Yang, A. X. R-9028-2016; Li, H. S-3644-2018; Hu, G. C. L-6160-2016; Zhao, J. A-3272-2017; Zhang, H. L. L-6168-2016; Yu, S. S. S-4050-2016; Dou, B. C. S-3642-2018; Wu, J. J. S-3712-2018

Article Citation: Liu, Q. H., Zhong, B., Tang, P., *et al.* Remote sensing data products oriented quantitative computing system—the GSC Best Practice Data Computing Environment 2018 [J]. *Journal of Global Change Data & Discovery*, 2018, 2(3): 271–278. DOI: 10.3974/geodp.2018.03.04.

1 Introduction

The Multi-source Data Synergized Quantitative Remote Sensing Production System (MuSyQ) was developed by the team of Liu, Q. H. in the State Key Laboratory of Remote Sensing Science, jointly hosted by Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences and Beijing Normal University, China. The system was awarded the Best Practice Data Computing Environment 2018 of Geographical Society of China (GSC). It is based on the architecture of global bio-geophysical parameters' production using multi-source remote sensing data, integrating a series of key algorithms for multi-source remote sensing data normalization and bio-geophysical parameters' retrieval. It has the capability of the global bio-geophysical parameters' production operationally.

1.1 Data

The datasets operationally used in the MuSyQ mainly include the global remote sensing data with 1 km, 5 km, and 25 km spatial resolutions and the data with 30 m spatial resolution of China from multiple satellites and different sensors. The major data are as follows:

- Data with 30 m and 300 m spatial resolutions: HJ-1/CCD, Landsat data and HJ-1/IRS, as well as GF-1/WFV (16 m), CEBRS/WFI, and ZY3/MUX.
- Data with 1 km spatial resolution: MODIS, AVHRR, MERSI, VIRR, and MERIS.
- Data with 5 km spatial resolution (25 km and 300 km): MTSAT, MSG, FY2, GOES, passive microwave data, and the gravity data.

MuSyQ has collected and processed the global data from 2010 to 2015 with resolution lower than 1 km and the data with 30 m resolution of China and Southeastern Asia in 2013. It has produced 16 kinds of bio-geophysical parameters at global scale for six years, with the total data in this system reaching 800 TB.

1.2 MuSyQ Computing Capability

1.2.1 Server and Network Configurations

The software topology of MuSyQ is shown in Figure 1. The computing component, MyClouds, was constructed on a distributed computing environment with the capability of dynamically expanding the computing ability for remote sensing data processing and analysis, and it was the key component of the computing software for hardware controlling.

The hardware topology of MuSyQ is shown in Figure 2. The hardware resources including computing, storage, and networking were all provided by Computer Information Center of Chinese Academy of Sciences, which served for the platform management.

The hardware of the system includes two categories: one is private and the other is public. The computer clusters composed of blade servers with large amount of storages were used to support the data processing and production of MuSyQ product. The self-developed distributed data management and task scheduling components were equipped. The whole system can support the data computing and accessing as fast as PB level. All the computer clusters were connected through InfiniBand network, with the connecting speed of 1,000 MB.

1.2.2 Software Configuration

The MuSyQ includes four sub-systems: data management, task scheduling, data normalization, and quantitative remote sensing production. All operations are carried on through clients.

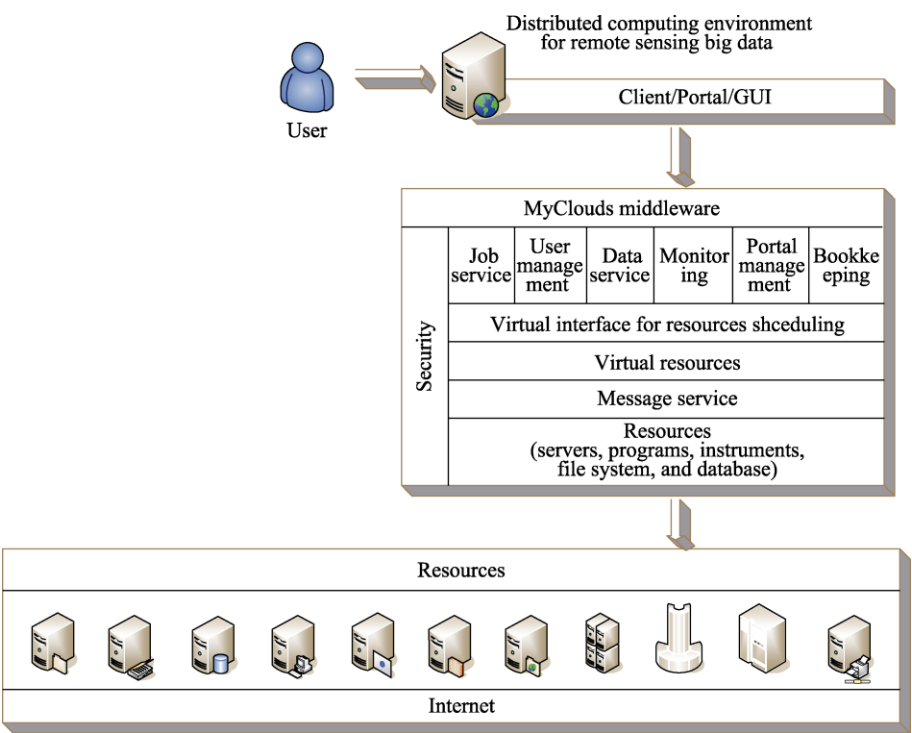


Figure 1 Software topology for the distributed computing environment of remote sensing big data

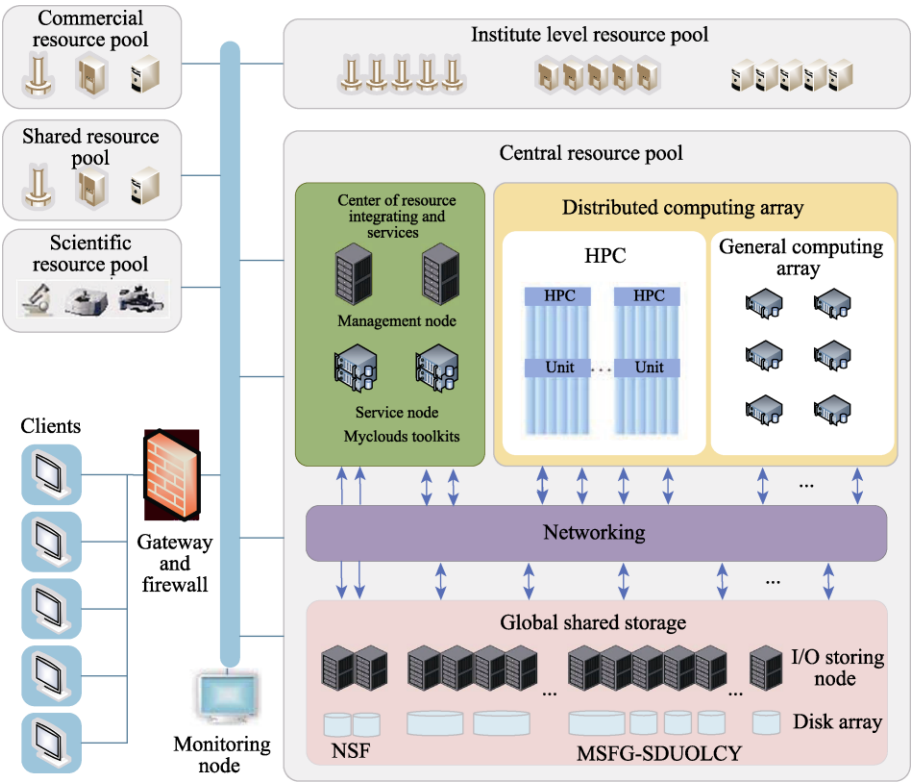


Figure 2 The hardware topology of the MuSyQ for distributed remote sensing data computing environment

The architecture of the software system is shown in Figure 3. All the data in this system is managed by the data management sub-system. Task scheduling sub-system is responsible for transferring the production procedures to the executable scripts, and all the required information is queried from data management sub-system. In addition, task scheduling sub-system is also responsible for scheduling and monitoring the tasks in the clusters. Data normalization sub-system takes in charge of the systematic processing of different data and the systematic processing procedures including geometric registration, spectral matching, radiometric cross-calibration, and atmospheric correction. The normalized data are finally stored into a HDF file, which also contains the metadata. Data normalization sub-system operates at a mode of data driving; once the original data are ready, the procedures are triggered automatically.

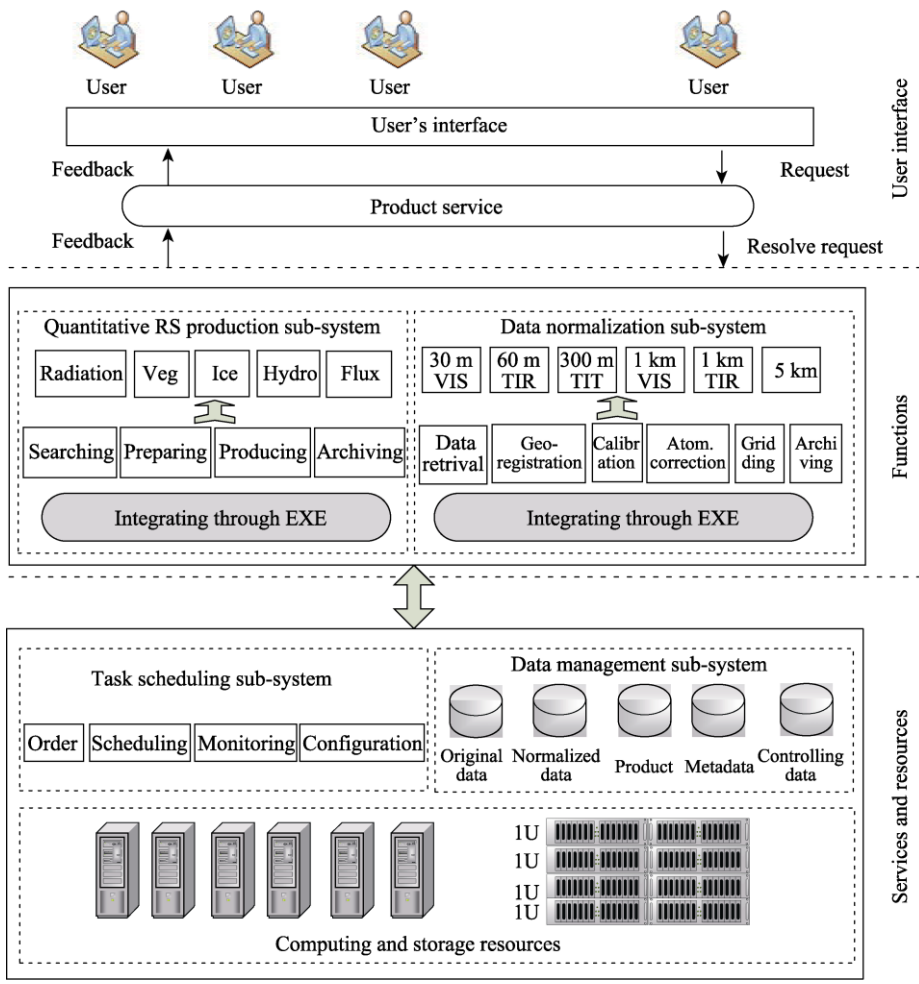


Figure 3 The MuSyQ's architecture.

Quantitative remote sensing production sub-system integrates 26 algorithms belonging to 4 categories of bio-geophysical parameters including vegetation, radiation budget, snow/ice, and water and heat flux. It is driven by user's orders; once an order is submitted, the sub-system sends them to task scheduling sub-system subsequently. The sub-systems are independent and connected through internet.

Table 1 The list of processed data in MuSyQ

NO.	Sensor	Spatial resolution (m)	NO.	Sensor	Spatial resolution (m)
1	HJ1A/HJ1B CCD	30	10	GOES11/ 13/15	5,000
2	Landsat TM/ETM+	30	11	MTSAT1/2	5,000
3	HJ1B IRS	300	12	Himawari-8	5,000
4	Terra/Aqua MODIS	250, 500, 1000	13	MSG2/3	5,000
5	NOAA AVHRR	1,100	14	FY2E	5,000
6	FY3A MERSI/VIRR	1,000	15	GF1 WfV	16
7	FY3B MERSI/VIRR	1,000	16	GF4	50
8	FY3C MERSI/VIRR	1,000	17	GF1 PMS	8
9	MODIS 05/06/07/35	1,000	18	GF2 PMS	5

1.2.3 MuSyQ's Capability for Data Processing

MuSyQ can normalize 18 datasets (Table 1) and produce more than 20 bio-geophysical parameters (Table 2). Up to date, 358 TB of original remote sensing data have been collected and 450 TB of normalized data and bio-geophysical parameters have been produced. These data include the global data (2010–2015) with resolution lower than 1 km and the data with 30 m resolution of China and Southeastern Asia in 2013.

Table 2 The bio-geophysical data products by MuSyQ

NO.	Bio-geo data products	Temporal frequency	NO.	Bio-geo data products	Temporal frequency
1	30 m AOD	Instantaneous	12	1 km VI	5-day
2	30 m FVC	10-day	13	1 km ALBEDO	10-day
3	30 m VI	10-day	14	1 km BRDF	10-day
4	30 m LAI	10-day	15	1 km Emissivity	5-day
5	30 m FPAR	10-day	16	1 km LST	Instantaneous
6	30 m Surface reflectance	10-day	17	1 km FPAR	5-day
7	30 m ALBEDO	10-day	18	1 km NPP	5-day
8	1 km AOD	Instantaneous	19	5 km DSR & PAR	3-hour, 1-day
9	1 km LAI	5-day	20	5 km DLR	3-hour, 1-day
10	1 km FVC	5-day	21	5 km LST	Instantaneous
11	1 km Surface reflectance	5-day	22	300 m LST	Instantaneous

Based on the current hardware configuration of MuSyQ, one week is needed to normalize a whole year's global data and produce 16 bio-geophysical parameters. The normalized data are obtained from 10 sensors including MODIS (2), MERSI (2), VIRR (2), MSG, MTSAT, GOES, and FY2. For a whole year's 30 m resolution data of China, 2 weeks are needed from data normalization to bio-geophysical parameters' producing.

1.2.4 Computing Capability

The computing capability can reach 500 TFlops and the online storage is up to 2.0 PB; furthermore, MyClouds has the potential to incorporate more computing and storing resources from other clouds dynamically to expand the computing ability, such as ALI, Amazon.

1.2.5 Websites

The system can be accessed at the Global Change Research Data Publishing & Repository (<http://www.geodoi.ac.cn>). The report on remote sensing monitoring of the global ecosystem

and environment can be accessed at <http://www.chinageoss.org/geoarc/2017/>.

1.3 Key Discoveries Made by MuSyQ:

- The advantages of the synergy of multi-source remote sensing data, such as higher temporal frequency and accuracy of biophysical parameters, have been realized by MuSyQ.
- The normalization and standardization of multi-source remote sensing data are the key steps for the synergy of multi-source remote sensing data.
- The support for *the report on remote sensing monitoring of the global ecosystem and environment* is an excellent proof of MuSyQ's capability and performance.
- The data environment for handling, processing and computing a large amount of remote sensing data should be efficiently based.

2 Advantages of the System

2.1 Function in Multiple Dimension of Data Computing

Most data processing systems focus on single type of remote sensing data^[1-2]. The MuSyQ can handle multiple sensors' data; this is the unique advantage of the system. Furthermore, the technologies for the normalization and standardization of remote sensing data from multiple satellites and sensors have been developed and brought into practices. These technologies have shortened the gaps of data quality between international and domestic remote sensing data. These technologies include geometric normalization, radiometric cross-calibration, and atmospheric correction.

The distributed computing environment, MyClouds, has been designed and developed specifically for remote sensing data. Compared with the similar computing environment^[3], MyClouds is more economic, flexible, and efficient for remote sensing data. In addition, MyClouds can be constructed not only on HPC servers, but also on cloud servers and ordinary servers.

2.2 High Level Data Products

The SEBERS data released by CRESDA (Center for Resources Satellite Data and Application) at <http://www.cresda.com.cn/CN/> and the FengYun data released by Metrological Remote Sensing Center of China at <http://satellite.nsmc.org.cn/portalsite/default.aspx> are mostly 1B products, although FengYun platform also provides a certain of the bio-geophysical parameters, which are derived by only using single FengYun dataset. The MuSyQ data products are all from level 2 or higher levels of data products.

3 Practicability, Novelty and Innovation

(1) The MuSyQ is the initiative system that launched a set of algorithms to be developed to normalize more than 8 types of remote sensing data, which support the synergy of multi-source remote sensing data for bio-geophysical parameters' retrieval.

The difficulty in synergy of multi-source remote sensing data is the inconsistency of data from different satellites and sensors, which is caused by many different reasons, such as the sensor technology, satellite technology. The inconsistency of data includes the geolocation, radiometric capability, the atmospheric effects, and the spectral settings. During the design and implementation of MuSyQ, a series of technical and scientific measures have been taken to facilitate the collaboratively using of multi-source satellite data, such as the geometric and radiometric normalization, and multi-source data standardization. These solutions have been

integrated into MuSyQ and have been capable of normalizing and standardizing over 18 kinds of satellite data.

(2) The bio-geophysical parameters' production capability by employing multi-source remote sensing data in collaborative way is distinctive.

In MuSyQ, most of the bio-geophysical parameters are produced by using multi-source collaboratively, which is the most distinctive characteristics compared with the other production systems like MODIS. This means of parameter derivation does improve the temporal frequency of all bio-geophysical parameters. Most of the parameters' temporal frequency has been improved from 8/16 days to 5 days, which are very practical for applications. MuSyQ is the first system integrating a whole set of algorithms employing synergy of multi-source remote sensing data and it becomes thereafter a milestone in this research area.

(3) MuSyQ also incorporates a large amount of updated remote sensing data.

These data include HJ-1/CCD, FY3A/MERSI, and so on. The percentage of domestic data is more than 50% among all the data, which were used for producing the bio-geophysical parameters to support *the report on remote sensing monitoring of the global ecosystem and environment*^[4]. It is an extraordinary application of domestic remote sensing data.

(4) MuSyQ is a practical example of integrating of remote sensing data analysis and super-computing.

The distributed computing environment, MyClouds, has been designed and developed specifically for remote sensing big data. With the further technology development of remote sensing data fusion and computer, more and more information will be derived from remote sensing data and will serve the applications for different areas.

4 Promotion and Application

4.1 Direct User

The direct users of the MuSyQ include scientific research institutes, industrial application departments, and national administrative agencies. In addition, with the development of the "Belt and Road" initiative, the MuSyQ products can serve the foreign enterprises in the national project for resources and environment in the process of information acquisition, as well as the promotion and application of the system itself along the countries of the "Belt and Road" area.

4.2 Application Prospect

(1) Promote the application of quantitative remote sensing products in scientific research institutes, industrial departments and international administrative agencies

The MuSyQ quantitative remote sensing products have been applied in relevant scientific research institutes, industrial application departments and national administrative agencies, etc., and have obtained the application certificates. The products have been applied to the Global Ecosystem and Environment Observation Analysis Research Cooperation (GEOARC) for three years, and nearly 30 sets of products have been published. Until June 17, 2018, the download of the report and datasets have reached 124.95 TB. MuSyQ also supported the report of China's Sustainable Development Monitoring by Remote Sensing (2016). When it was released in June 2017, more than 20 media journalists attended the press conference, and since then more than 20 national media and 30 local media have reported it, such as the People's Daily, the People's Daily (overseas edition), Xinhua News Agency, the People's Liberation Army Daily, Guangming Daily, China Daily, etc., and 36 internet media also widely carry it out.

(2) MuSyQ has marketization potential to promote the industrialization development of

remote sensing technology

MuSyQ has the ability to process more than 10 kinds of multi-source remote sensing data and to produce more than 20 kinds of quantitative remote sensing products, which can lower the threshold of quantitative remote sensing application and provide services for local governments, companies, and the public. At the same time, its market share will be inevitably increased and the development of remote sensing industrialization will be further promoted. Recently, the industry sector has put forward a strong demand for quantitative remote sensing product system, and began to deploy related construction tasks. MuSyQ, as well as the data processing technology, and the production ability of multi-source cooperative and quantitative remote sensing products can directly serve these construction projects and have a good prospect of industrialization.

5 Scientific Discoveries

The major scientific discoveries are as follows:

(1) MuSyQ brings the synergy of multi-source remote sensing data into practice systematically and has produced a set of bio-geophysical parameters with 5-day temporal frequency, which is much higher than the similar ones with 8-/16-day temporal frequency. Furthermore, the data missing induced by clouds is less.

(2) The key normalization technologies for more than 10 remote sensing data have been explored and the algorithms have been integrated, which makes an excellent preparation for synergy of multi-source remote sensing data.

(3) The support for *the report on remote sensing monitoring of the global ecosystem and environment* is an excellent proof of MuSyQ's capability and performance.

References

- [1] Masuoka, E., Tilmes, C., Devine, N., *et al.* Evolution of the MODIS science data processing system [C]. Proceedings of 2001 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Sydney, Australia, 2001: 1454–1457.
- [2] MERIS [OL]. <https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/envisat/instruments/meris>.
- [3] Lifka, D., Foster, I., Mehringer, S., *et al.* Xsede cloud survey report [R]. XSEDE Cloud Integration Investigation Team, 2013. <http://www.cac.cornell.edu/technologies/XSEDECloudSurveyReport.pdf>.
- [4] Chinese Academy of Sciences. The report on remote sensing monitoring of global ecosystem and environment (2016) [R]. Economic Daily, July 21, 2017.