2008

Removing Long-Term Errors from the AVHRR Observation Based on Normalized Difference Vegetation Index (NDVI)

Md Zahidur Rahman
CUNY La Guardia Community College

Leonid Roytman
CUNY City College

Atiqur Rahman
CUNY City College

Felix Kogan
NOAA

Runa Jesmin
University of London - King's College

Removing Long-Term Errors from the AVHRR Observation Based on Normalized Difference Vegetation Index (NDVI)

Md Z Rahman [1], Leonid Roytman[2], Atiqur Rahman[2], Felix Kogan[3], and Runa Jesmin[4]
[1] Department of Mathematics, LaGuardia Community College of the City University of New York, 31-10 Thomson Avenue, Long Island City, NY-11101
[2] Department of Electrical Engineering, The City College of the City University of New York, 138 street & convent Avenue, New York, NY-10031
[3] National Oceanic and Atmospheric Administration (NOAA) Silver Spring, MD 20910
[4] Department of Computer Science, King's College London (University of London), Strand London WC2R 2LS, London, UK

ABSTRACT

This paper investigates Normalized Difference Vegetation Index (NDVI) stability in the NOAA/NESDIS Global Vegetation Index (GVI) data during 1982-2003. Advanced Very High Resolution Radiometer (AVHRR) weekly data for the five NOAA afternoon satellites for the China dataset is studied, for it includes a wide variety of different ecosystems represented globally. It was found that data for the years 1988, 1992, 1993, 1994, 1995 and 2000 are not stable enough compared to other years because of satellite orbit drift, and AVHRR sensor degradation. It is assumed that data from NOAA-7 (1982, 1983), NOAA-9 (1985, 1986), NOAA-11 (1989, 1990), NOAA-14 (1996, 1997), and NOAA-16 (2001, 2002) to be standard because these satellite's equator crossing time fall within 1330 and 1500, and hence maximizing the value of coefficients. The crux of the proposed correction procedure consists of dividing standard year's data sets into two subsets. The subset 1 (standard data correction sets) is used for correcting unstable years and then corrected data for this years compared with the standard data in the subset 2 (standard data validation sets). In this paper, we apply empirical distribution function (EDF) to correct this deficiency of data for the affected years. We normalize or correct NDVI data by the method of EDF compared with the standard. Using these normalized values, we estimate new NDVI time series which provides NDVI data for these years that match in subset 2 that is used for data validation.
Keywords: Vegetation, data, stability, satellite, ecosystem

1. INTRODUCTION

For almost two decades, the Advanced Very High Resolution Radiometer (AVHRR) on NOAA polar-orbiting satellites have observed radiances, which were collected, sampled, and stored for the entire world. These data were intensively used by the global community for studying and monitoring land surface, atmosphere, and recently for analyzing climate and environmental changes [1] [2]. AVHRR data, though informative, can not be directly used in climate change studies because of the orbit drift in the NOAA satellites (particularly, NOAA-9, -11, and -14) over these satellites' life time [5] [6]. **P**rice 1991 attributed this drift to the selection of a satellite orbit designed to avoid direct sunshine on the instruments. This orbital drift leads to the measurements of Normalized Difference Vegetation Index (NDVI) are being taken at different local times during the satellites' life time, thereby introducing a temporal inconsistency in the NDVI data. . Consequently, an orbital drift introduces errors in AVHRR data sets for some satellites. It was found that data for the years 1988, 1992, 1993, 1994, 1995 (first eight weeks), and 2000 are not stable enough compared to other years because of satellite orbit drift, and AVHRR sensor degradation. We assume that data from NOAA-7(1982, 1983), NOAA-9 (1985, 1986), NOAA-11(1989, 1990), NOAA-14(1996, 1997), and NOAA-16 (2001, 2002) are the best suited for analysis because of their equator crossing time between 1330 and 1500. As this period of the day is the best time for satellite observation, we consider data for these years as standard.

This paper investigates NDVI stability in the NOAA/NESDIS Global Vegetation Index (GVI) data for the period 1982-2003 [7] [8]. AVHRR weekly data for the five NOAA afternoon satellites NOAA-7, NOAA-9, NOAA-11, NOAA-14,

Remote Sensing for Environmental Monitoring, GIS Applications, and Geology VIII,
edited by Ulrich Michel, Daniel L. Civco, Manfred Ehlers, Hermann J. Kaufmann,
Proc. of SPIE Vol. 7110, 71101G · © 2008 SPIE · CCC code: 0277-786X/08/$18 · doi: 10.1117/12.799900

and NOAA-16 are used for the China region. China has all major types of ecosystems present in the world. These observations were made only under clear skies and thus some regions and seasons may be poorly sampled due to the contamination of clouds. To avoid misinterpretation of signals due to orbit drift and satellite changes, correction must be applied to remove or, at least, reduce these effects from the AVHRR data so one can use a long-term time-series for study. This research introduces a scientific methodology that can be easily implementable to generate the desired long-term time-series. The main goal of this paper is to correct the NDVI data for the years 1988, 1992, 1993, 1994, 1995, and 2000 by the method of empirical distribution functions (EDF) compared to the standard data. We can use the same methodology globally to create vegetation index to improve the climatology. The corrected datasets can be used as proxy to study climate change, epidemic analysis, and drought prediction etc.

1.1 LAND TARGETS

The land targets approximately of 20^0 N to 45^0 N in latitude and 72^0 E to 133^0 E in longitudes were selected for China (Fig. 1).

Fig. 1. Geographical Map of China with the area study (bordered area)

We attempted to select relatively small uniform areas using common knowledge of geography, climate, ecosystem, and human activities. The main cover types are desert, forest, and grassland.

2. DATA AND PROCESSING

Satellite data were collected from the NOAA/NESDIS Global Vegetation Index (GVI) data set [7] [8] which is one of the most widely used satellite products worldwide. The GVI is produced by sampling and mapping the 4-km daily radiance in the VIS (Ch1, 0.58-0.68 μm), NIR (Ch2, 0.72-1.1 μm) in Figure 2 measured onboard NOAA polar-orbiting satellites, to a 16-km map. To minimize cloud effects, these maps, including the NDVI, solar zenith angle, and satellite scan angle, are composited over a 7- day period by saving those values that have the largest difference between VIS and NIR reflectance for each map cell. The weekly GVI data from January 1982 through January 1985 for NOAA-7, from April 1985 through September 1988 for NOAA-9, from October 1988 through August 1994 for NOAA-11, from March 1995 through December 2000 for NOAA-14, and from January 2001 through December 2003 for NOAA-16 were used here.

During 1985-2000, the performance of the channel 1 and 2 differed between NOAA-9, NOAA-11, and NOAA-14 satellites and most importantly, degraded over time for each satellite differently. Since there is no in-flight calibration of channel 1 and 2 of the AVHRR, the question arises as to the validity of the pre-launch calibration coefficients, both in the early days after launch and, perhaps more seriously, after the AVHRR has been in space for a long time. There is a clear evidence in several environmental products, such as the normalized difference vegetation index , global cloud morphology, and earth radiation budget [9] , that are generated from channel -1 and channel-2 AVHRR data to indicate that the performance of the instrument these two channels has deteriorated after launch. The need to correct for this in-orbit degradation has been keenly felt recently since it is now being proposed to use the long-term records of AVHRR-derived environmental products in climate and global change studies [11] [12] [13] [14] [15] and the degradation of the instrument with time is clearly illustrated by the results shown in Figure 3. Therefore, the standard data preparation procedure for the 7-day composite time series now includes a correction of Ch1 (VIS) and Ch2 (NIR) values following Rao and Chen [13].

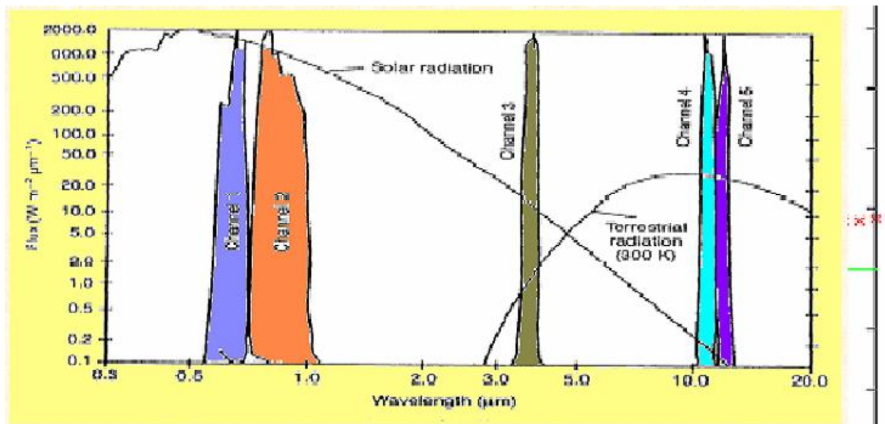

Fig. 2. Normalized spectral response of AVHRR

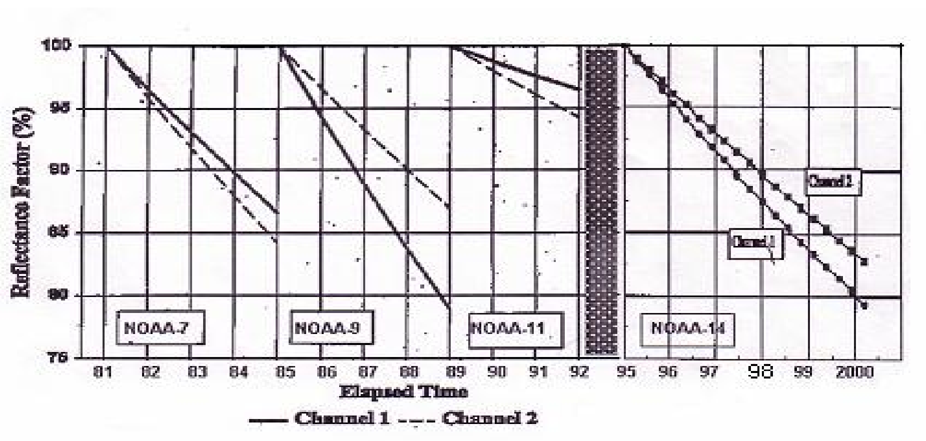

Fig. 3. Degradation of AVHRR Channels 1 and 2 [Rao, 1994]

3. METHODOLOGY

For each satellite, we construct the NDVI time series and also approximate linear trend using least square technique. From trend equation, we estimate two values: the largest difference ((dN_t) between NDVI at the beginning (N_b) and the end (N_e) of satellite life and difference (dN_s) between NDVI at the beginning of the next (n) satellite (N_{bn}) and at the end of the previous (p) one (N_{ep}).

$$dN_t = 100^* (N_e - N_b)/(N_b); \qquad dN_s = 100^*(N_{bn} - N_{ep})/N_{ep} \qquad (1)$$

If the dN_t values are positive then the NDVI time series upward trend and downward for negative value; positive dN_s indicate larger NDVI at the end of the previous satellite and smaller NDVI in the opposite case.

There is no available physical method that can be used to correct for the stability of NDVI. Therefore, we developed a statistical model for the correction of NDVI. The empirical distribution function (EDF) is a statistical technique which is used to generate a normalization data of the years 1988, 1992, 1993, 1994, 1995 and 2000 compared with standard
Empirical distribution function (EDF) approach is based on the physical reality, that each ecosystem may be characterized by very specific statistical distribution, independent of the time of observation. It is the best available technique to normalize satellite data. It allows us to represent global ecosystem from desert to tropical forest and to correct extreme distortions in satellite data related to technical problem. To generate the normalization data, we begin by selecting samples of unnormalized earth-scene data covering as much of the range intensities as possible. For NOAA satellites, the area will be rectangular, extending several thousand pixels from desert to tropical forest (both east to west and north to south). Corresponding to the incoming radiance from any pixel, the instrument will respond with an output, x . One can compile the discrete density function, i.e., the histogram, describing the relative frequency of occurrence of each possible count value, for each year. For year **i**, which is the year to be normalized, let the histogram be $P_{\mathbf{i}}(x)$. An empirical distribution function (EDF) $P_{\mathbf{i}}(x)$ can then be generated; viz.[10],

$$P_{\mathbf{i}}(x)= \sum_{t=0}^{x} p_{\mathbf{i}}(t) \qquad (2)$$

The EDF is also known as a cumulative histogram of relative frequency. It is a non-decreasing function of x, and its maximum value is unity.

For convenience, however, we have chosen the maximum value to be 1; i.e., if the maximum possible output in counts is x, then $P_{\mathbf{i}}(x) = 1$, as shown in Figure 4

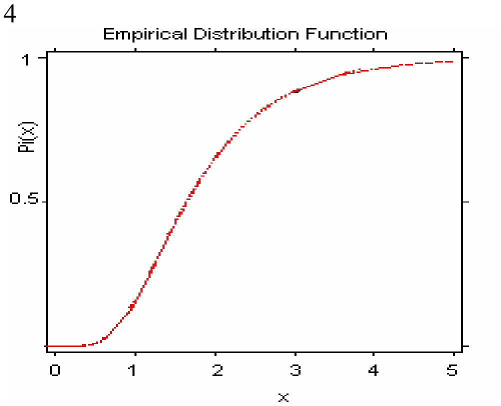

Fig. 4. Empirical Distribution Function

In these terms, the basic premise of normalization is that for each output value x in year i, the normalized value x' should satisfy [10]

$$P_s(x') = P_i(x), \qquad (3)$$

Where the subscript **s** refers to the standard year. In practice, not only is P_s non-decreasing, but it is also monotonically increasing as a function of x' in the domain of x' where there are data. Therefore, it can be inverted, yielding the solution for x', [10]

$$x' = P_s^{-1}(P_i(x)) \quad (4)$$

When it is applied sequentially for every possible count value x, equation 4 generates the normalization data relating each x to an x'. Fig. 5 shows how the procedure is applied in actual practice to generate the normalization data [4] [10]. The figure shows idealized EDF's for the standard and unnormalized years i. In the figure the EDF's are continuous, but in practice they are discrete, being specified only integer values of x.

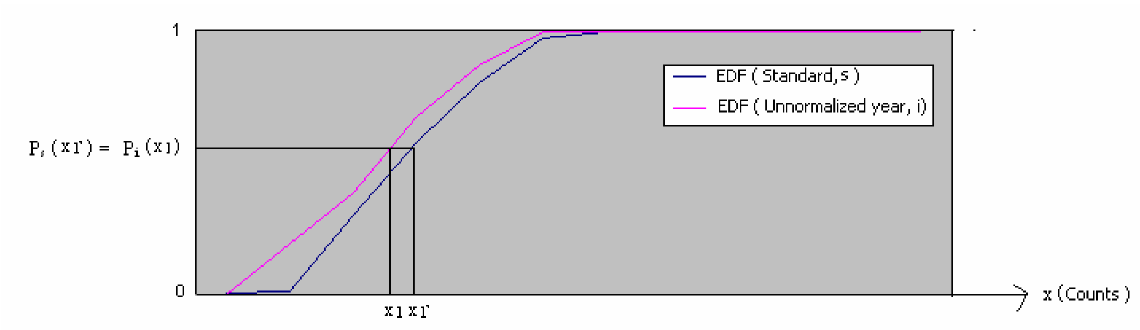

Fig. 5. Example of procedure to generate normalization data

To find x'_1 , the normalized count value corresponding to the unnormalized count value of x_1 , the following is the procedure: First, for the count value x_1 in unnormalized year i, , find the decimal or percentage value from the EDF of year i. In the illustration it is Pi (x_1). Then find the point on the standard year's EDF with the same decimal or percentage value. According to equation 3, that decimal or percentage can also be expressed as $P_s\ (x'_1)$. Finally, use the EDF of the standard year to find the normalized count value x'_1. Since the data are actually discrete, we will need to interpolate within the EDF of the standard year to find the value x'_1. Using this technique, we can generate the normalization data. Therefore, we choose EDF method for the normalization of satellite data.

4. RESULTS AND DISCUSSION

We produce NDVI time series of five NOAA satellites, which is illustrated in Figure 6.

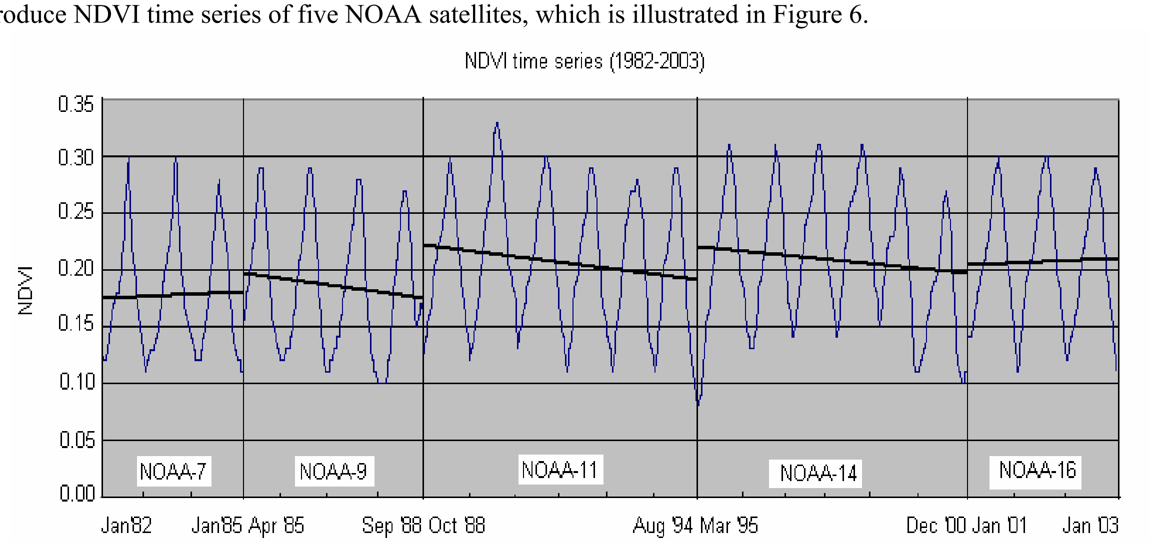

Fig. 6. NDVI time series (yearly old NDVI data) for study area in China.

Data from the afternoon polar orbiters is preferred for producing the NDVI time series because of the high sun elevation angle (low solar zenith angle). However, the equator crossing time drifts to a later hour as the satellites age [6]. Satellite orbit drift results in a systematic change of illumination conditions which is one of the main sources of non-uniformity in multi annual NDVI time series. Figure 6 Shows that the NDVI data of 1988, 1992, 1993, 1994, 1995 and 2000 are nonuniform compared to other years because of satellite orbital drift, and sensor degradation. Therefore, we need to correct the data of those years. We apply EDF for the correction of data of those years. First, EDF construct for unnormalized data and than generate the normalize data compared with standard. Figure 7 shows how the procedure is applied in actual practice to generate normalization NDVI value [4[[10].The figure shows idealized EDF's for the standard and the year of 1988. As EDF are based on cumulative histogram, they are discrete. But in Figure 7 they are shown as continuous function.

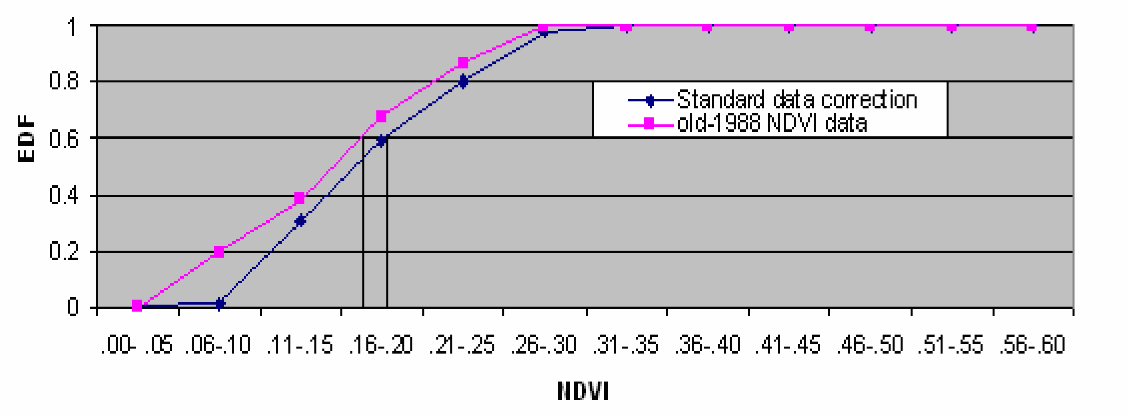

Fig. 7. Illustration of procedure to generate normalization NDVI data.

For example, for the NDVI value 0.16 in year 1988 (Fig. 7) find the value from the EDF of year 1988. In the illustration it the EDF_{88} is 0.6. Then find the point on the standard year's EDF with the same EDF value. According to equation 3, that the EDF value can also be expressed as the $EDF_{standard}$ is 0.6. Finally, use the EDF of the standard year to find the normalized count value 0.18. Since the data are actually discrete, we will need to interpolate within the EDF of the standard year to find the value of 0.18. Therefore,

$$\text{New NDVI value for 1988} = NDVI_{1988} + (NDVI_{standard} - NDVI_{1988}) \text{ or}$$
$$\text{New NDVI value for 1988} = 0.16 + (0.18-0.16) = 0.18$$

Using this technique, EDF's produce to normalize or correct data for the years 1988, 1992, 1993, 1994, 1995, and 2000 compared with standard which are illustrated in following Figures.

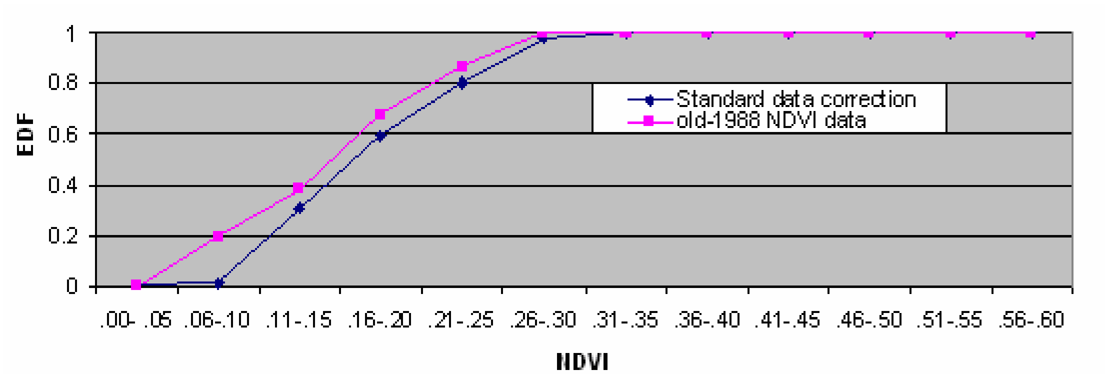

Fig.8. Empirical distribution functions for unnormalized data of 1988 compared with standard data correction sets (subset1)

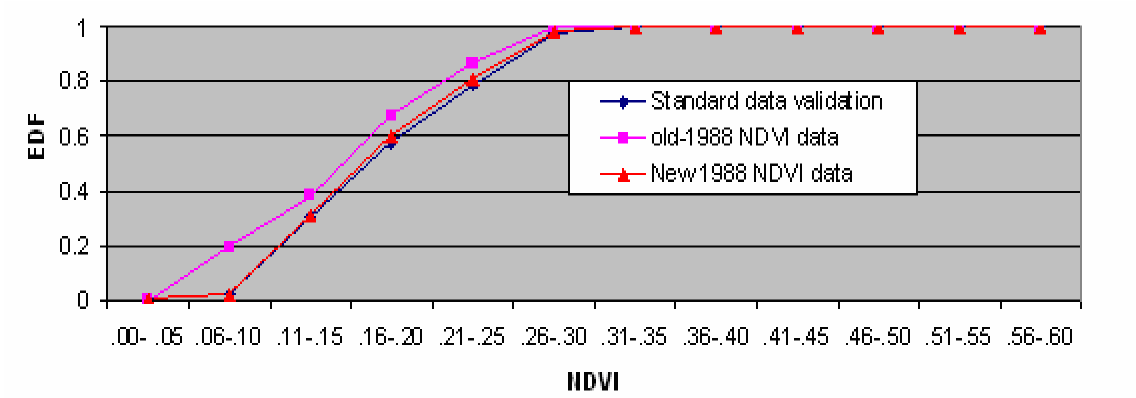

Fig.9. Empirical distribution functions for normalized data of 1988 compared with standard data validation sets (subset 2)

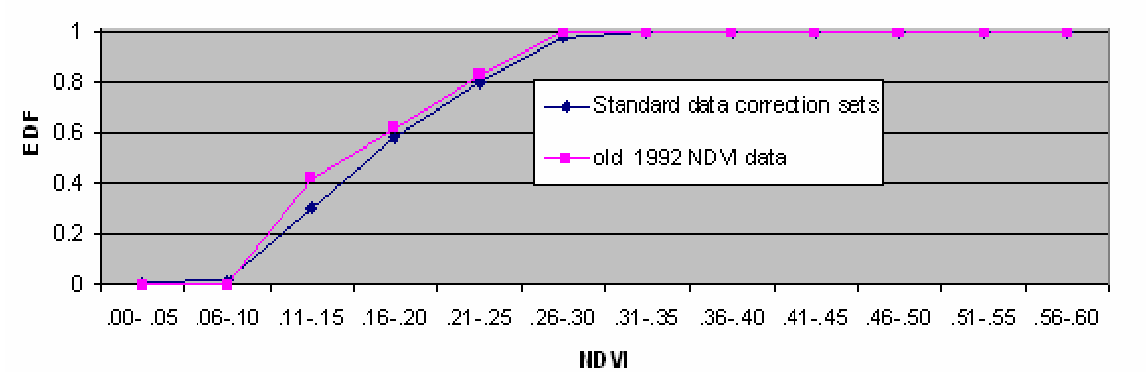

Fig.10. Empirical distribution functions for unnormalized data of 1992 compared with standard data correction sets (subset1)

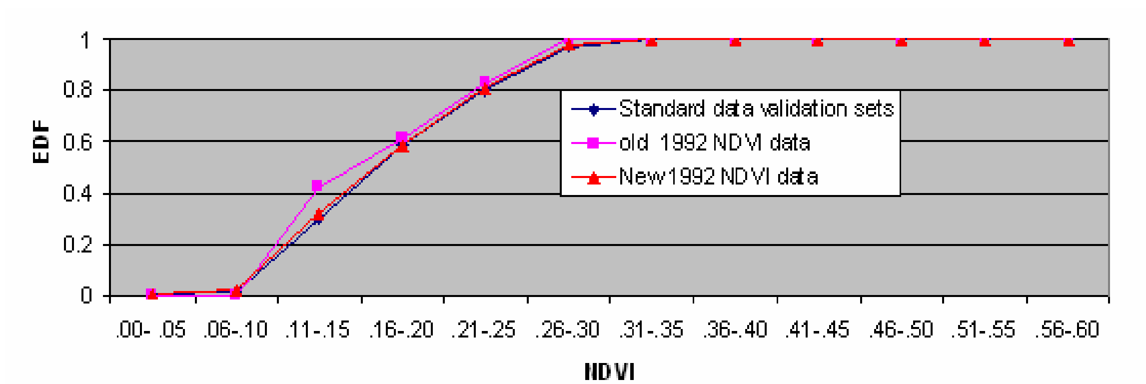

Fig.11. Empirical distribution functions for normalized data of 1992 compared with standard data validation sets (subset 2)

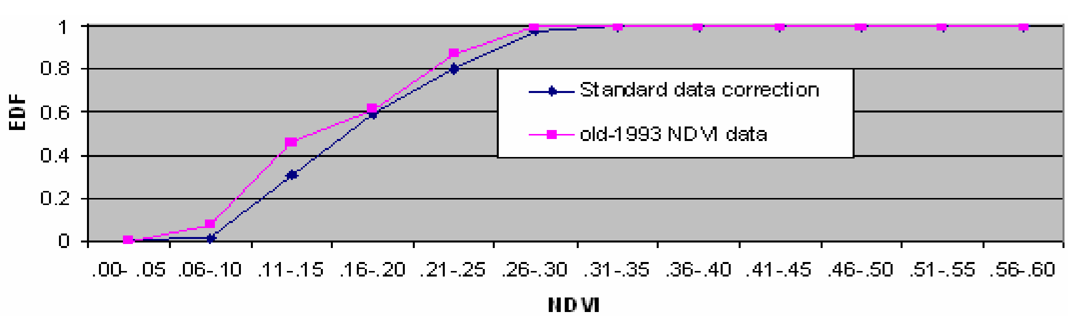

Fig.12. Empirical distribution functions for unnormalized data of 1993 compared with standard data correction sets (subset1)

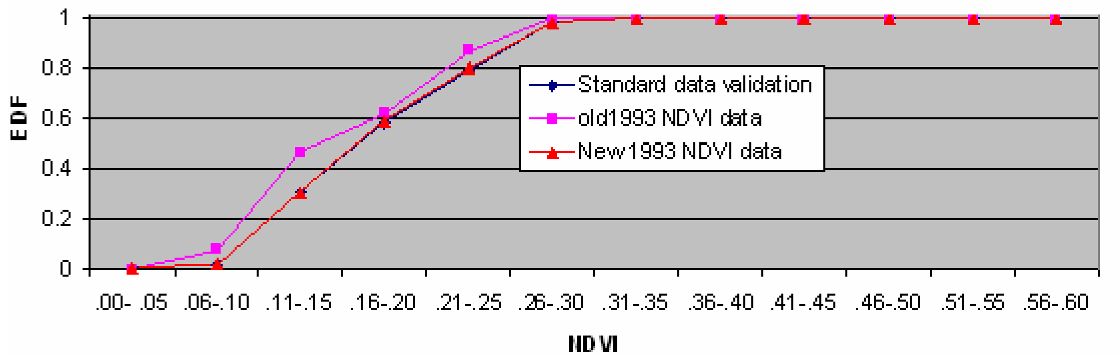

Fig.13. Empirical distribution functions for normalized data of 1993 compared with standard data validation sets (subset 2)

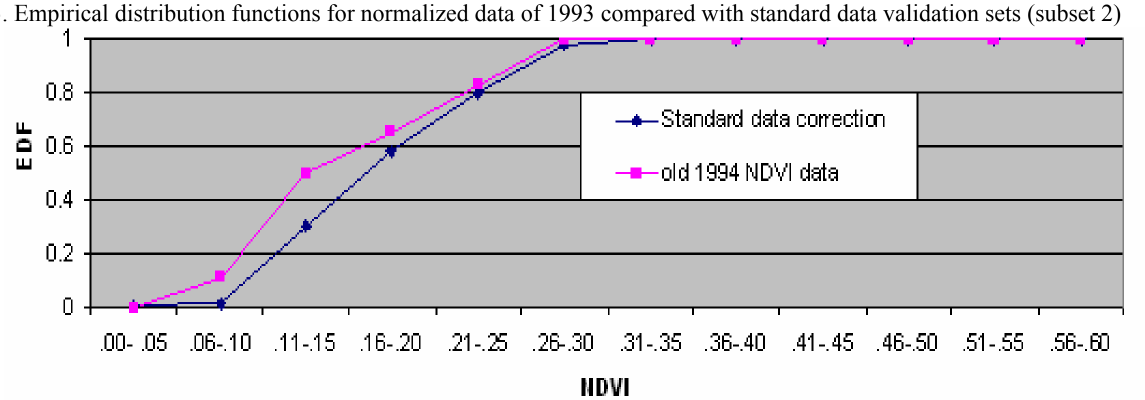

Fig.14. Empirical distribution functions for unnormalized data of 1994 compared with standard data correction sets (subset1)

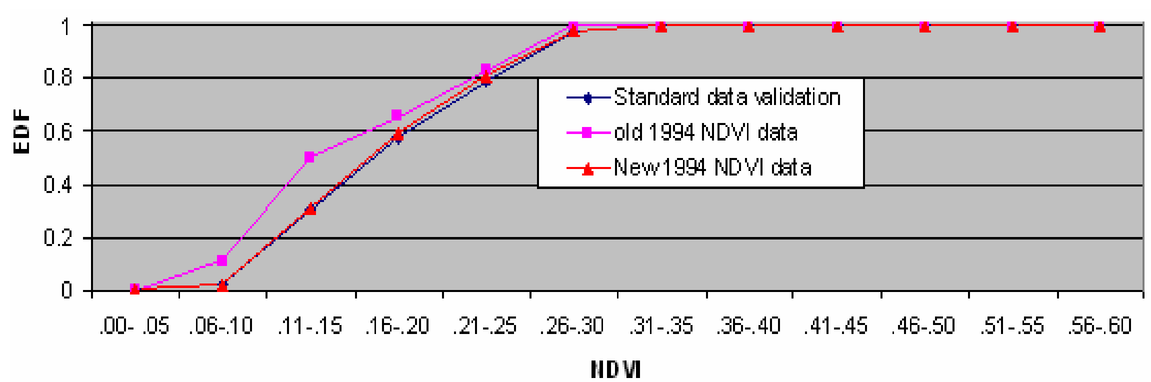

Fig.15. Empirical distribution functions for normalized data of 1994 compared with standard data validation sets (subset 2)

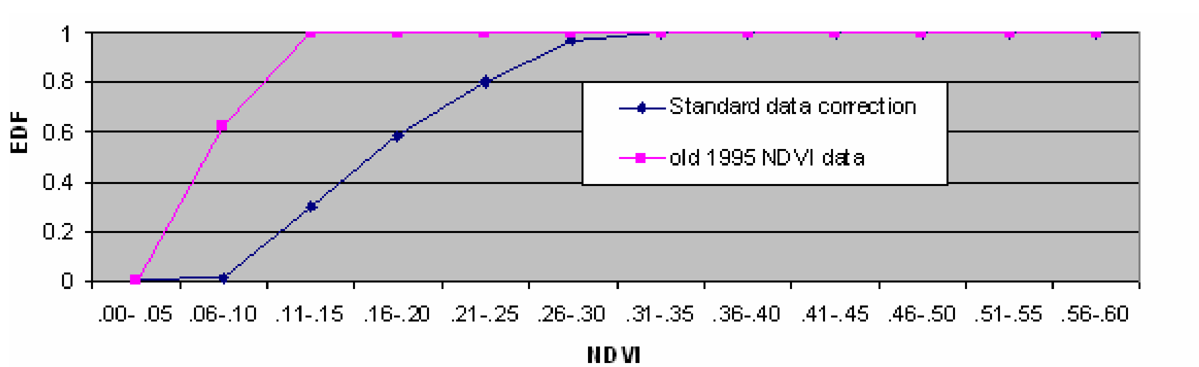

Fig.16. Empirical distribution functions for unnormalized data of 1995 compared with standard data correction sets (subset1)

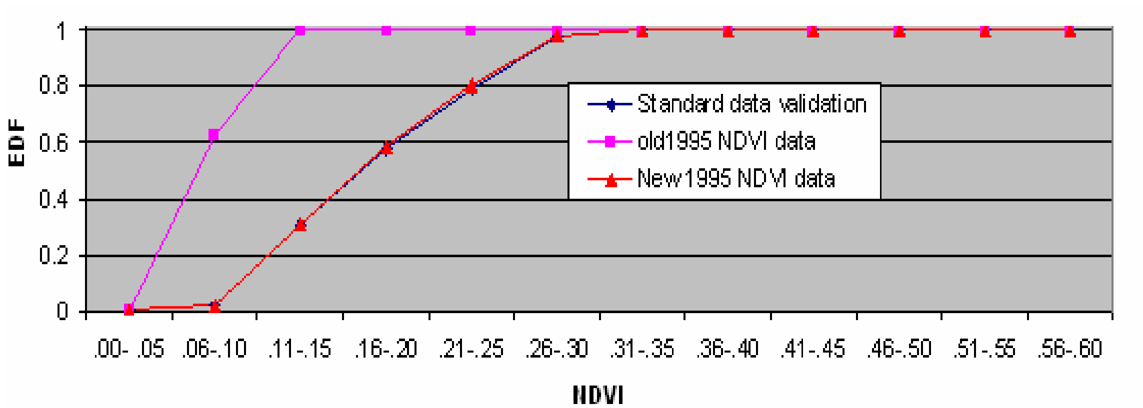

Fig.17. Empirical distribution functions for normalized data of 1995 compared with standard data validation sets (subset 2)

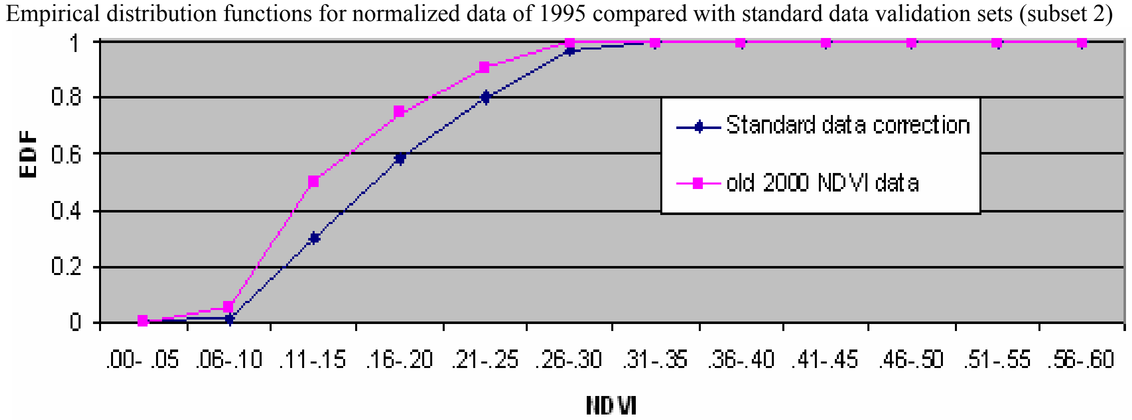

Fig.18. Empirical distribution functions for unnormalized data of 2000 compared with standard data correction sets (subset1)

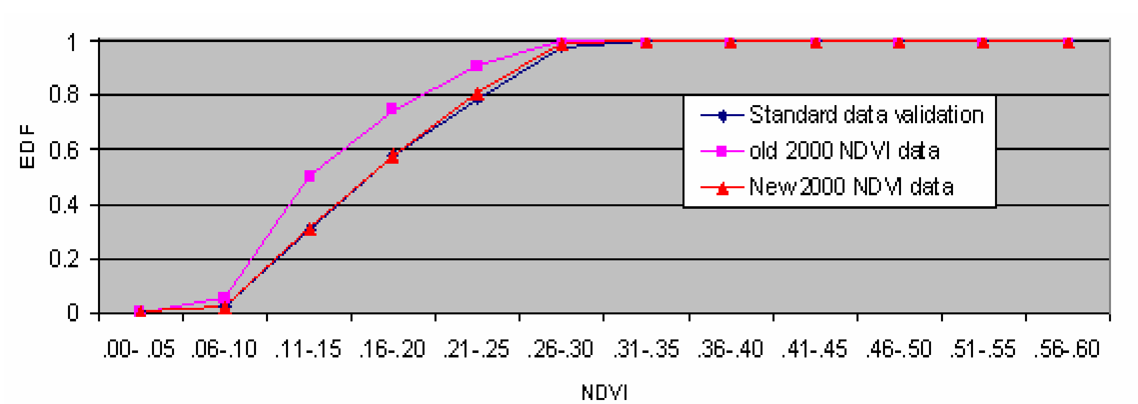

Fig.19. Empirical distribution functions for normalized data of 2000 compared with standard data validation sets (subset 2)

Fig.8-19 shows the EDF's of the normalize data for each of those years and indicates that the normalization was successful in making the EDF's of the two years nearly identical. This implies that the relationships between the EDF's remained essentially the same between two years. Those relationships, in fact, depend only on the relative function between two years. As long as the relative functions remain in the same, the normalization data remain effective. Using normalized value, we produce new NDVI time series for study area in China as shown in Figure 20 which shows improve the NDVI data (pink line) of the year of 1988, 1992, 1993, 1994, 1995, and 2000.

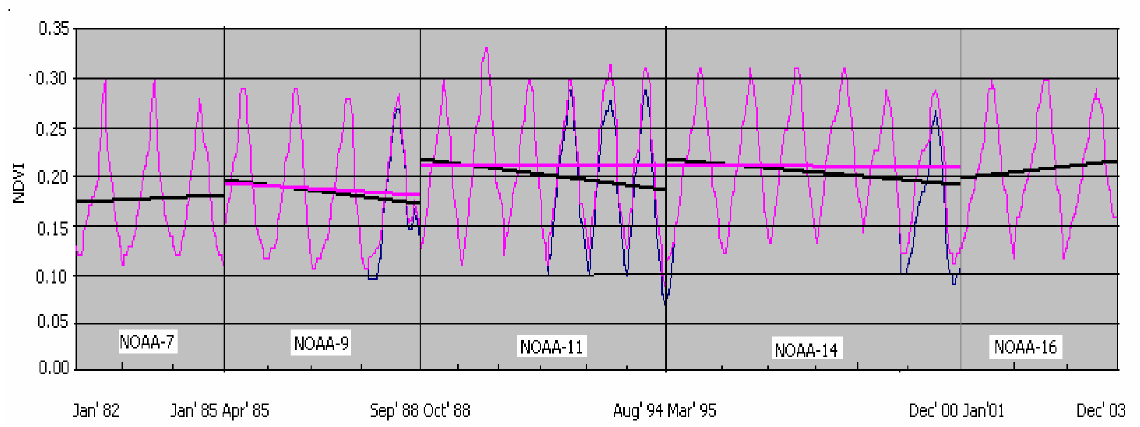

Fig. 20. New NDVI time series (yearly) for study area in China (old NDVI data ——, and new NDVI data ——).

NDVI trends for china and jumps between the satellites are illustrated in Figure 20 and the errors are estimated in Table 1. Figure 20 shows some NDVI trends for each satellite and jump from one satellite to the next one. Considering old NDVI trend (Table 1A), for china, NOAA-9, -11, and -14 have negative trend and NOAA-7, -16 have positive trend. Therefore, NOAA-7, and -16 shows clear tendency to NDVI increase during its three years in operation. However, important is trend rate. Analysis shows that high rate of NDVI change for NOAA-9,-11, and -14 by reduction of NDVI in 1988, 1992-1994, and 2000 due to elevated amount of considerable degradation of satellite orbit.

Regarding NDVI jump from one satellite to the next in Table 1(B), general tendency is a reduction of NDVI between beginning of NOAA-9 and the end of NOAA-7, between beginning of NOAA-16 and the end of NOAA-14. An increase in NDVI is observed only during satellite change from NOAA-9 to NOAA-11, NOAA-11 to NOAA-14, and NOAA-14 to NOAA-16 due to already mentioned orbit drift of satellite.
After correction of NDVI, we also estimate errors (new NDVI) in Table 1 of NDVI trends and jumps between the satellites. This table shows improve the NDVI trends for each satellite and jump from one satellite to the next one. The EDF method is designed to reduce only errors due to orbit drift, the dominant uncertainty in temperature variation during the satellite life time [6]. However, it may be difficult to accurately and completely remove this effect and thus orbit remains as an error source, though at a reduced level. Another large uncertainty lies in NDVI calibration and sensor degradation which includes all errors such as incomplete atmospheric corrections, surface corrections, sensor degradation and volcanic eruptions.

Table 1: Estimation of Errors in (A) NDVI trend at the End of a Satellite Life and (B) Jumps between the Satellites (% to the beginning level)

Target		A					B			
		N-7	N-9	N-11	N-14	N-16	N-7/9	N-9/11	N-11/14	N-14/16
China	Old NDVI	3	-10	-12	-11	7	10	30	16	5
	New NDVI	3	-5	0	0	7	7	19	0	-5

5. CONCLUSIONS

Empirical distribution function improves the time related stability of NDVI for all satellites, especially NOAA-9, -11, and -14 environmental satellites. This is strong evidence that normalization by EDF matching is an effective method for improving stability of NDVI time series. Empirical distribution function (EDF) approach is based on the physical reality, that each ecosystem may be characterized by very specific statistical distribution, independent of the time of observation. EDF approach proposed here shows encouraging results which can be used globally to create vegetation index to improve the climatology. For climate and global change studies, NDVI time series are not stable enough. Following NDVI data distortion due to external forcing (satellite orbit degradation), NDVI data for 1988 (NOAA-9) and 1993, 1994 (NOAA-11), and 2000 (NOAA-14) will likely distort mean values and other statistics and should be tested comprehensively before they are used for monitoring

REFERENCES

[1] Kogan, F.N., **"G**lobal draught watch from space," *Bull. Am. Meteorol. Soc.,* **78,** page 621- 636, (1997)

[2] Kogan, F.N., and X. Zhu, "**E**volution of long-term errors in NDVI time series," *Adv. Space Res*. Vol. **28** (1), page 149-153(2001).

[3] Kogan, F.N., **"D**raught of the late 1980s in the United States as derived from NOAA polar orbiting satellite data," *Bull. Amer. Metor. Soc.* **76,** page 655-668 (1995)

[4] Horn, B.K. P. and Woodham, R.J., "Destriping LANDSAT MSS images by histogram modification," *Computer Graphics and Image processing***. 10**, page 69-83(1979)

[5] Gutman, G. Garik, "On the use of long-term global data of land reflactances and vegetation indices derived from the AVHRR," *J. **G**eophys. Res*., **104,** 6, 6241-6255, 1999.

[6] Price, J. C., **"**Timing of NOAA afternoon passes," , *Int. J. Remote Sens*., **12,** 193-198, 1991

[7] Kidwell, K., "Global **V**egetation **U**ser's **G**uide, NOAA," ***N****atl.* **E***nviron. Satellite* **D***ata and* ***I****nf.* **S**erv. Natl. Clim. Data Cent., Dep. Of **C**ommer., Washington, D. C., 1997.

[8] **K**idwell, **K. B.,** *NOAA Polar Orbiter Data Users Guide.* U.S. Department of Commerce. National Oceanic and Atmospheric Administration, Satellite Data Services Division, Washington, D.C., 1991.

[9] P**. C**racknell, ***T****he Advanced Very High Resolution Radiometer,* Taylor & Francis, page 90-97, 1997

[10] Weinreb, M.P, RXie, J.H.Lienesch, D.S Crosby . "Removing stripes in GOES images by Matching Empirical Distribution Functions," NOAA Technical Memorandum NESDIS **26**, May 1989

[11] Rao, C.R.N. and J.Chen, Nonlinearity corrections for the thermal infrared channels of the Advanced Very High Resolution Radiometer: assessment and recommendations. NOAA Technical Report NESDIS **69** (Washington DC: US Department of Commerce, 1993a

[12] Rao, C.R.N. , "Degradation of the visible and near infrared channels of the Advanced Very High Resolution Radiometer on the NOAA –9 spacecraft," assessment and recommendations for corrections. NOAA Technical Report NESDIS **70** (Washington DC: US Department of Commerce). 1993b

[13] Rao**,** C.R.N. and J**.**Chen, Post-launch calibration of the visible and near-infrared channels of the Advanced Very High Resolution Radiometer on NOAA-7, -9, and -11 spacecraft. NOAA Technical Report NESDIS **78** (Washington, DC: US Department of Commerce). 1994

[14] Rao, C.R.N**.** and J**.**Chen, Postlaunch calibration of the visible and near-infrared channels of the advanced very high resolution radiometer on the NOAA 14 spacecraft, 1**,** *Int. J. Remote Sens.,* **7,** 2743-2747, 1996.

[15] Brest**,** C. L**.,** and W. B. Rossow, "Radiometric calibration and monitoring of NOAA AVHRR data for ISCCP," *Int. J. Remote Sens*., **13**, 235-273, 1992