

HANTS 알고리즘을 이용한 MODIS NDVI 시계열 영상의 구름화소 문제 해결에 관한 연구

A Study of cloud-free MODIS NDVI time series reconstruction using HANTS algorithm

허용*, 변영기, 김용일, 유기윤

Yong Huh*, Young-gi Byun, Yong-il Kim, Ki-yun Yu

서울대학교 공간정보연구실

{hy7808*, kko071, yik, kiyun}@snu.ac.kr

요약

식생지수 시계열 자료를 이용한 식생 및 토지피복 모니터링을 수행하기 위해서는 구름으로 인한 누락 및 왜곡된 식생지수 문제를 먼저 해결해야만 한다. 특히 한반도와 같이 여름철 집중 호우기에 대부분의 영상에 구름이 존재하는 경우 이들 구름화소를 제거하거나 복원하지 않을 경우, 분석 결과에 상당한 왜곡이 발생하거나 특정 시기의 영상자료를 분석에 반영할 수 없는 경우가 발생하게 된다. HANTS 알고리즘은 이 같은 구름 화소 문제를 해결하기 위한 알고리즘으로 연중 식생지수의 변화는 비교적 단순한 반복적 주기함수의 형태를 가지므로 소수의 \cos 함수를 이용한 푸리에 근사식으로 전체 연중 식생지수를 표현할 수 있다는 가정에서 출발한다. 이 때 구름화소로 인한 원식생지수와와의 차이가 특정 임계값을 초과하였을 경우 해당 관측치를 근사과정에서 제외함으로써 구름의 영향을 받지 않은 식생지수 시계열 자료만을 이용하게 된다. 이 과정을 수행하기 위해서는 몇몇 제어변수의 설정이 필요한데, 본 연구에서는 한반도와 같이 특정 시기에 장기간 구름이 분포하는 상황에서 최적의 식생지수 복원을 위한 HANTS 알고리즘의 제어변수를 선정하고 재구축된 식생지수를 평가하였다. 이를 위한 실험으로 2002년 대전 지역의 MODIS Terra 식생지수 시계열 영상을 대상으로 HANTS 알고리즘을 주요 식생피복별로 적용해 보았다.

1. 서론

시계열 자료의 분석은 밴드별 분광특성에 기초한 일반적인 원격탐사 분석과는 달리 생태계의 계절적인 변화, 즉 동적 특성 연구에 자주 이용되어왔다. 원격탐사를 이용한 시계열 자료 분석을 수행하기 위해서는 대상지역의 영상을 동일한 조건으로 장기간 취득함으로써 시간 및 주변 환경인자의 변화에 따른 관측 대상물의 변화를 지속적으로 감지할 수 있어야 한다.

시계열 자료의 분석은 단시기의 다중분광영상 분석과는 몇 가지 차이점을 가

지고 있다(정성창 2000). 첫째, 일정 기간 동안의 방대한 영상자료가 사용됨으로써 시계열 특성을 파악하기 쉬운 장점이 있으나, 그와 비례하여 자료 분석시간이 많이 소모된다. 또한 다중분광영상의 경우 다차원 벡터형태의 입력자료를 단일변수로 표현하기 위한 압축 과정이 필요하다. 둘째, 일반적으로 개별 시점에서 취득된 영상자료가 다중분광자료라 할지라도 식생지수나 주성분분석과 같은 한 개의 일반화된 영상밴드를 시계열 자료화하여 분석에 적용한다. 셋째, 시계열 분석을 위한 자료의 표준화 과정이 필요하다. 즉 동일

한 취득 조건에서의 분광정보로 개별 취득된 영상을 변환하는 과정은 분석결과에 큰 영향을 미치게 된다. 특히 구름과 같은 기상조건에 의한 영향은 반드시 고려되어야 한다.

이런 점을 고려한 결과 지표환경 시계열 자료의 분석사례들은 몇 가지 유사한 접근방식을 따르게 된다. 먼저, 개별 촬영 시점의 지표분광특성을 식생지수로 대표함으로써 시계열 분석자료를 단순화하여 분석을 수행한다. 둘째, 영상 취득 당시의 관측각, 대기 투과율, 태양 고도각 등과 같은 외부 환경에 따른 분광특성의 영향을 최소화하기 위하여 다양한 보정과정을 수행한다. 일반적으로 적용이 용이한 상대 방사보정(relatively radiometric calibration)만을 수행하여도 안정적인 결과를 얻을 수 있다. 하지만 식생지수는 구름이나 대기에 의한 영향을 심각하게 받는 문제점을 가지고 있다. 특히 구름 및 강설로 인한 영향은 지표 분광특성을 왜곡하거나, 취득 불가능한 결과를 가져온다(Verhoef 1996). 그 결과 구름화소의 탐지 및 복원을 위한 다양한 연구가 진행되었다.

HANTS(harmonic analysis of NDVI time series)알고리즘은 단계적이고 반복적으로 푸리에 근사식을 적용함으로써 비구름화소만을 이용한 전체 시계열 자료의 근사식을 구함으로써 구름화소의 탐지 및 복원을 동시에 수행하는 알고리즘으로 다양한 연구에 적용되었다(Azzali 2000). 하지만 알고리즘의 제어변수 및 근사식의 모형을 어떻게 결정하는가에 따라 다양한 결과를 얻게 된다. 특히 한반도와 같이 여름 장마철의 장기간의 구름화소로 인한 식생지수 누락의 경우 어떤 모형과 제어변수를 사용하는 것이 최적의 결과를 얻을 수 있는가에 대한 분석이 필요하다.

본 연구에서는 2002년 대전 지역의 MODIS 식생지수 영상을 이용하여 구름화소를 탐지하고 탐지된 화소를 나머지 화소들을 이용하여 복원하는 재구축 과정을 수행하였다. 이 과정에서 다양한 식생피복별로 알고리즘의 제어변수를 변화시켜보면서 최적 변수값을 도출해보았다.

2. HANTS 알고리즘

HANTS 알고리즘은 식생지수로 표현되는 식생 성장은 계절적인 영향을 강하게 받는 주기적인 형태를 보이며, 푸리에 근사를 통하여 얻게 되는 소수의 cos 삼각함수의 합으로 표현될 수 있다는 개념에서 출발한다(Jakubauskas 2002, Azzali 2000).

연속함수 $y(t)$ 는 푸리에 근사식에 의하여 식(2)와 같은 형태로 표현될 수 있다. 여기서 k 는 전체 식생지수 시계열자료를 근사시키기 위한 모형을 구성하는 삼각함수의 개수를 의미하며, L 은 전체 주기, t 는 관측시점, a_0 는 전체 주기동안의 식생지수의 평균값이 된다.

$$y(t) = \sum_{n=1}^k (a_n \cos \frac{2\pi nt}{L} + b_n \sin \frac{2\pi nt}{L}) + a_0 \quad (2)$$

하지만 사용된 시계열 자료에 구름에 의하여 왜곡된 화소가 존재할 경우, 그 화소는 실제 지표의 식생지수보다 매우 낮은 지수를 가지게 된다. 따라서 푸리에 근사를 통한 근사 함수식과 원자료를 비교하고 특정 임계값보다 그 차이가 클 경우 구름화소에 의한 왜곡으로 추정할 수 있다. 물론 이 방법을 적용하기 위해서는 관측된 모든 시계열 자료와 비교하여 구름에 의하여 왜곡된 화소의 빈도가 매우 낮다는 가정을 충족시켜야 한다.

Fig. 1은 이와 같은 개념을 바탕으로 개발된 HANTS 알고리즘의 프로세스를 보여준다. 모든 시점에서 관측된 원식생지수 시계열 자료, $f(t)$ 를 이용하여 근사식 $F(t)$ 를 얻게 된다. 이 때 근사식의 모형은 사용자에게 의하여 먼저 정의되어야 한다. 이렇게 구한 $f(t)$ 과 $F(t)$ 를 식(2)와 같은 조건을 만족하면서, 이 차이가 가장 큰 시점의 화소를 탐지한다.

$$F(t) - f(t) \geq \theta \quad (1)$$

여기서, t 는 식(1)의 원자료와 근사값의 차이가 최대값을 가지는 시점, 즉

$\text{argmax}\{F(i) - f(i)\}$ 로 정의되며, θ 는 사용자에게 의하여 결정된 임계값이다.

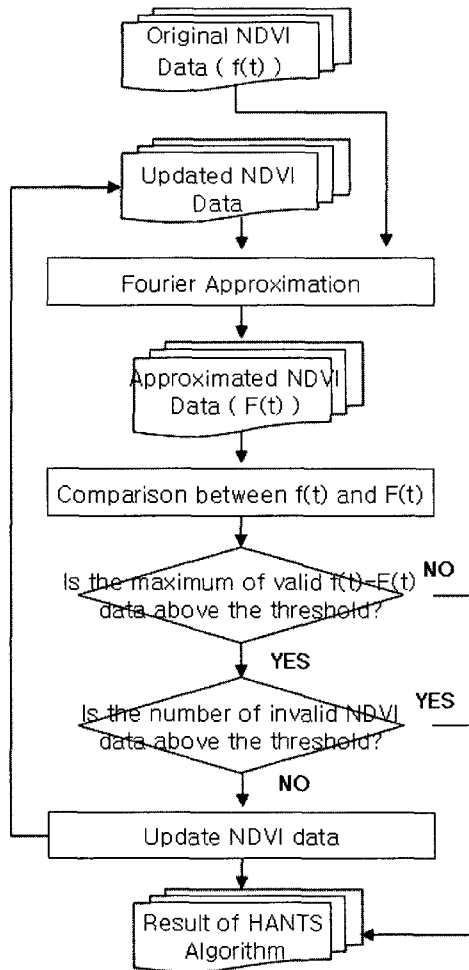


Fig. 1 HANTS 알고리즘의 흐름도

이렇게 탐지된 하나의 구름화소를 제외한 나머지 관측값만을 이용하여 다시 푸리에 근사를 수행함으로써 새로운 $F'(t)$ 를 얻게 된다. 역시 이렇게 얻은 $F'(t)$ 를 식(1)의 $F(t)$ 에 대입한다. 그리고 임계값을 초과하는 관측값이 역시 존재할 경우 구름화소로 판별, 해당 관측값을 제외한 나머지 시계열 자료를 이용하여 새로운 근사식을 반복적으로 구하게 된다.

만약 남아있는 모든 관측값에서 $f(t)$ 와 $F'(t)$ 의 차이가 임계값보다 작을 경우 알고리즘은 중지되며, 최종 결과물을 얻게 된다. 하지만 경우에 따라 너무 많은 관측값들이 연산과정에서 제외되어 복원된 시

계열 자료가 관측자료를 충분히 대표할 수 없게 된다. 이 경우 전체 관측값에서 누락되는 자료의 총 개수를 제한함으로써 문제를 해결하게 된다.

3. 실험 방법

3.1 입력자료

본 연구에서 사용된 분석자료는 2002년 1년 동안 16일 간격으로 생성된 MODIS Terra Vegetation Indices 16-Day L3 Global(MOD13)자료이다. 1999년 발사된 Terra위성에 탑재된 MODIS 센서 영상은 매일 한반도 상공을 이동하면서 지표를 관측한다. 하지만 110도에 이르는 넓은 관측각으로 인하여 관심 대상지역이 위성체의 연직 아래(nadir)에 있는 경우와 관측범위의 외각(off-nadir)에 있는 경우, 반사된 전차기파가 투과해야 하는 경사거리의 변화로 인하여 센서에 도달한 전차기파 에너지 강도에 차이가 발생하게 된다. 또한 매 관측 시점별 대기조건의 영향으로 인하여 동일한 취득조건으로 영상을 보정해야 하는 복잡한 보정 과정이 필요하게 된다.

최대값 대체합성법(MVC : Mximum Value Composite)은 이와 같은 현상은 물론 구름화소 문제를 비교적 간단한 방법을 통하여 일정 부분 해결할 수 있다. NDVI를 이용한 지표관측에 있어서 대기의 영향은 대부분 이상적인 조건에서의 NDVI보다 낮은 관측치로 측정되는 결과를 가져온다. 따라서 일정 기간 동안의 동일한 공간해상력과 좌표체계를 갖는 복수의 NDVI 영상자료를 중첩시켜 동일한 위치의 화소값을 비교하여 이중 가장 높은 값을 택하여 영상자료를 새로 작성하는 방법을 적용함으로써 구름이나 대기의 영향을 보정할 수 있게 된다.

MOD13은 위와 같은 과정을 16일간의 개별 영상자료를 이용하여 구축한 것으로 본 연구의 입력 자료로 사용하였다. 하지만 7,8월경의 집중호우기간의 경우 16일 이상 구름화소가 발생할 경우 여전히 문제점이 남게 된다.

3.2 실험 모형 설계

HANTS 알고리즘을 수행하기 위해서는 몇 가지 변수를 사용자가 정의해야만 하여, 이렇게 정의된 변수에 따라 분석결과 의 정확도가 결정되게 된다.

먼저 식(1)의 임계값, θ 와 식(2)에서 몇 개의 삼각함수를 이용할 것인가, 즉 k 를 결정해야 한다. 또한 HANTS 알고리즘은 근사과정에서 사용되는 최소한의 관측값의 개수가 연산의 종료조건으로 사용되는바, 이 역시 사용자가 결정해야한다.

HANTS 알고리즘을 적용한 많은 선행 사례에서 θ 는 0.05를 일반적으로 사용하고 있어 본 연구에서도 이 임계값을 이용하였다. 하지만 어떤 k 값과 최소 관측 개수를 사용할 것인가에 관해서는 명확한 기준이 제시되어 있지 않다. 따라서 본 연구에서는 다음과 같은 6가지 실험 모형에 따라 각각 자료를 처리하고 그 결과를 비교하였다.

Table 1. 최적 모형을 결정하기 위한 6개의 실험 모형 설계

| | k | 최소 관측 개수 (원자료의 총 개수:23) |
|-------|-----|----------------------------|
| Case1 | 2 | 15 |
| Case2 | 2 | 17 |
| Case3 | 2 | 19 |
| Case4 | 3 | 15 |
| Case5 | 3 | 17 |
| Case6 | 3 | 19 |

Table 2. 실험모형별 RMSE 측정표; 표현된 수치는 모두 NDVI*10000을 적용한 결과이며, Average*는 도심지를 제외한 나머지 지역만의 평균임

| | Case1 | Case2 | Case3 | Case4 | Case5 | Case6 |
|-------------------|--------|--------|--------|--------|---------|--------|
| Urban | 224.57 | 233.86 | 266.21 | 230.00 | 223.53 | 283.48 |
| Agriculture | 253.42 | 270.17 | 301.53 | 289.01 | 255.97 | 363.91 |
| Coniferous Forest | 266.56 | 241.60 | 285.18 | 229.92 | 274.54 | 303.14 |
| Broadleaf Forest | 244.24 | 247.28 | 219.67 | 236.19 | 252.46 | 252.41 |
| Average | 247.20 | 248.23 | 268.15 | 246.28 | 251.625 | 300.74 |
| Average* | 254.74 | 253.07 | 268.79 | 251.71 | 260.99 | 306.49 |

4. 실험 결과

Fig 2.는 Table 1.에서 정의한 조건에 따라 HANTS 알고리즘을 적용한 결과이다. 일반적으로 토지피복별로 연중 식생지수의 변화 양상이 상이하므로 도심지, 농경지, 침엽수림, 활엽수림과 같은 대표적 토지피복을 선정하고 적용해보았다.

14번째 관측시점인 7,8월을 중심으로 원 NDVI 관측값이 급격히 감소한 것을 확인할 수 있으며, 실험 모형별 복원결과와의 차이를 확인할 수 있다. 일반적으로 k 를 2로 설정한 경우에는 복원된 형태가 비교적 단순하며 완만한 형태를 보이는 반면, 3으로 설정한 경우에는 좀 더 원자료의 형태를 세밀하게 유지하지만 구름에 의한 왜곡마저도 복원에 반영시켜버리는 문제가 발생하기도 하였다. 특히 Fig 2.의 좌상단 도표의 경우와 같이 장기간 구름에 의한 왜곡이 발생한 경우에는 보정이 거의 일어나지 않는다는 것을 확인할 수 있었다.

최적 모형을 결정하기위한 정량화 평가하기 위하여 피복별 RMSE를 계산하였다. 이때, RMSE는 일반적인 계산과는 달리 관측값과 복원값의 차이가 NDVI기준으로 0.1보다 작은 관측값만을 이용하였다. 즉, 복원된 NDVI보다 실제 NDVI가 0.1이상 낮은 경우에는 구름의 영향을 받은 것으로 가정하고 해당 관측값을 제외하고 나머지 관측값만을 RMSE 계산에 반영하였다.

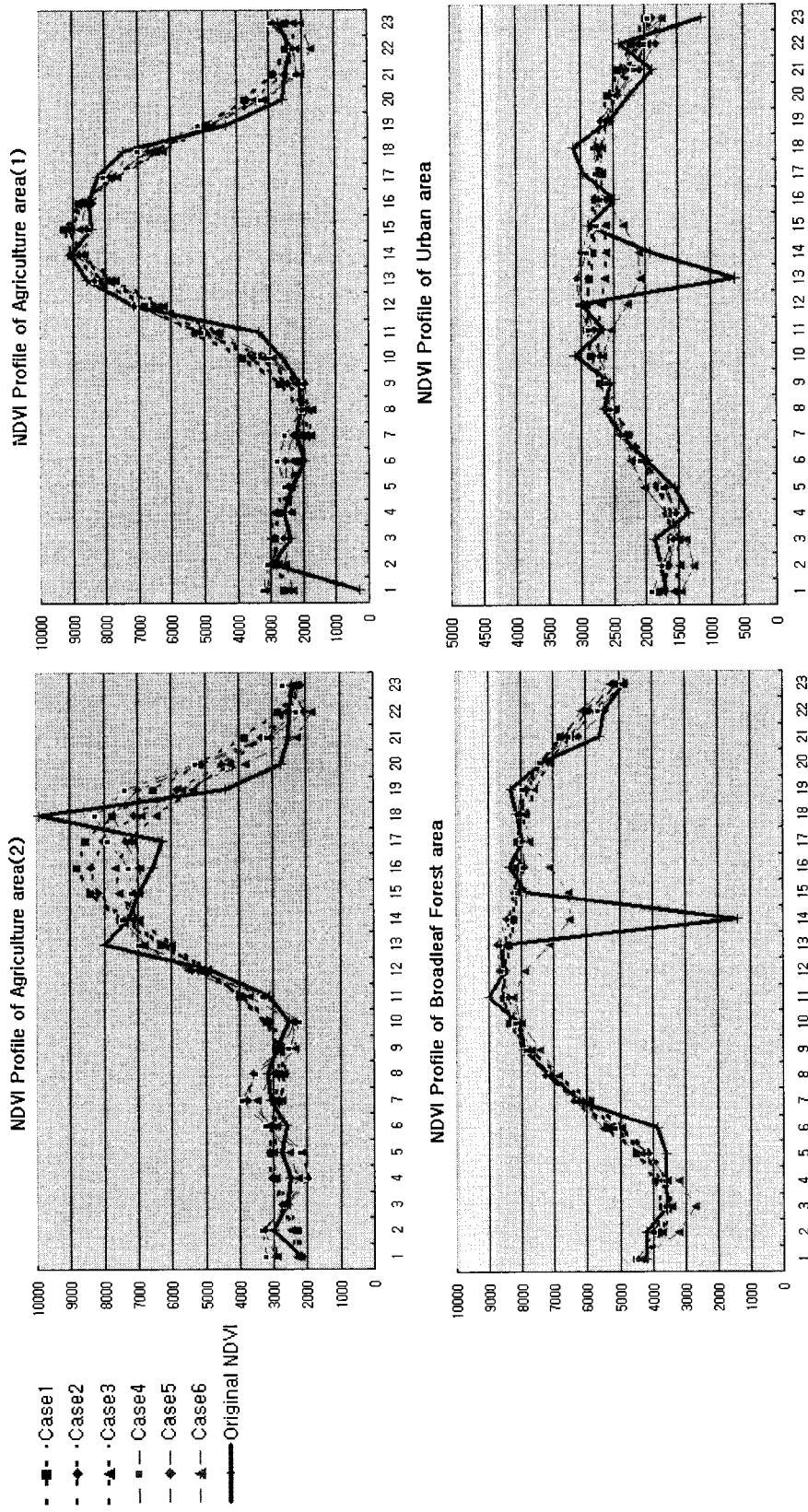


Fig 2. 주요 토지피복별 HANTS 알고리즘 적용 결과

적용결과 유사한 결과를 보이긴 했지만 Case4에서 가장 낮은 RMSE가 측정되었다. 그러나 이 경우 앞에서 언급한 것과 같이 Fig 2.의 좌상단 도표에서는 구름에 의한 왜곡을 전혀 보정하지 못하는 문제점을 가지고 있었다. 즉 구름화소로 인한 NDVI 왜곡이 1회 또는 연속 2회까지 발생하였을 경우에는 Case4가 최적의 복원결과를 가져올 수 있지만 장기적으로 왜곡이 발생할 경우에는 왜곡된 복원결과를 가져오게 된다. 따라서 이와 같은 현상을 고려했을 때, 가장 안정적인 결과를 보이는 것은 Case2, 즉 2개의 삼각다항식에 23개의 NDVI 관측값중 최소 17개 이상을 적용한 HANTS 알고리즘의 적용이 가장 신뢰도가 있다고 판단하였다.

5. 결론

HANTS 알고리즘을 이용하여 2002년 대전지역 MODIS NDVI 시계열 영상에서 발생하는 구름화소의 왜곡을 보정해보았다. 적용결과 구름발생으로 인한 왜곡을 상당부분 보정할 수 있었음을 확인할 수 있었다. 하지만 다음과 같은 점에서 한계를 가지고 있었다. 먼저 정확도 평가에 있어서 구름에 의한 왜곡이 발생한 시점에서의 실제 지표에서의 식생지수를 복원된 식생지수와 비교하지 못하였다. 또한 RMSE를 이용한 평가의 경우 결과적으로 구름이 발생하지 않은 시점에서의 식생지수 관측값과 HANTS 알고리즘의 복원값을 비교한 것이므로 구름에 의하여 왜곡된 식생지수를 얼마나 올바르게 복원하였는가에 대한 평가척도로는 충분치 못하다.

따라서 지상 관측에 의한 실측자료나, 2002년도에는 구름에 의한 왜곡이 발생하였으나 2001년도 또는 2003년도에서 왜곡이 발생하지 않은 지역의 자료를 이용한 간접비교를 통한 향후 연구가 필요하다.

참고문헌

- Verhoef, W., 1996, Application of Harmonic Analysis of NDVI Time Series (HANTS). In Fourier Analysis of Temporal NDVI in the Southern African and American Continents, DLO Winand Staring Centre, Wageningen, The Netherlands, Report 108, pp.19~24.
- 정성창, 2000, 광역 시계열 원격탐사자료 분석의 특성과 응용, Journal of the Korean Society of Remote Sensing, Vol.16, No.1, pp.1~11.
- Azzali, S. and Menenti, M., 2000, Mapping vegetation-soil-climate complexes in southern Africa using temporal Fourier analysis of NOAA-AVHRR NDVI data, Int.J. Remote Sensing, Vol.21, No.5, pp.973~996.
- Jakubauskas, E. Mark, Legates, R. David and Kastens, H. Jude, 2002, Crop identification using harmonic analysis of time-series AVHRR NDVI data, Computers and Electronics in Agriculture, Vol.37, pp.127~139.