Determination of scaling characteristics of AVHRR data with wavelets: application to SGP97

N. A. BRUNSELL* and R. R. GILLIES†
*Department of Plants, Soils, and Biometeorology, Utah State University, Logan, UT 84322-4820, USA
†Department of Aquatic, Watershed and Earth Resources, Utah State University, Logan, UT 84322-5210, USA

(Received 2 August 2001; in final form 4 April 2002)

Abstract. A wavelet multi-resolution analysis was conducted using vegetation and radiometric temperature data derived from the Advanced Very High Resolution Radiometer (AVHRR) data from the Southern Great Plains 1997 (SGP97) Hydrology Experiment. The wavelet coefficients were used to investigate whether the Normalized Difference Vegetation Index (NDVI) and radiometric temperature fields exhibit self-similar scaling behaviour. Using the first six moments of the wavelet coefficients through eight levels of dyadic decomposition, the NDVI data are shown to be statistically self-similar with a slope of approximately −0.72 in each of the horizontal, vertical, and diagonal directions of the image over scales ranging from 2000 to 256 000 m. The radiometric temperature data are also shown to exhibit self-similarity with slopes ranging from −0.9 in the diagonal direction to −1.0 in the vertical direction over the same scales. These slopes indicate long range behaviour and may imply a methodology for statistically assimilating remotely sensed data into large-scale meso and global climate models.

1. Introduction

Satellite remote sensing potentially offers a valuable resource in the modelling and validation of terrestrial systems as well as the understanding of hydrological and meteorological processes, due in part to a large spatial coverage and varying temporal resolutions. Remote sensing could be used in several ways in conjunction with global and meso-scale modelling, the most likely of which is assimilating satellite data into such models (e.g. Houses et al. 1998). In another study, Santanello and Carlson (2001) showed some improvement in model (MM5) performance by incorporating a percentage cover of vegetation as derived from the Advanced Very High Resolution Radiometer (AVHRR) sensor and applying an aggregation scheme of the fields up to the resolution of the model. At the same time, this issue of assimilation exemplifies the lack of current understanding when translating data to larger spatial scales. Satellite data such as AVHRR are collected at a spatial scale of 1 km², whereas climate models generally run at a grid resolution of the order of 10–100 km². This is one aspect of the ‘scaling problem’ referred to as aggregation, and it is a direct
Surface heterogeneity exists at all scales. However, measurements of this heterogeneity can only sample a finite range of temporal and spatial scales, resulting in data having some resolution. The resolution of a spatial measurement represents the integrated volume or pixel size, and is the smallest observable measurement within a given data set. For a good review of the subject of small-scale heterogeneity and pixel size see Cracknell (1998). In practical terms, it becomes necessary to understand at what scale the underlying physical process is accounted for, and what scale is the most beneficial for the specific application.

In many areas of science, the investigator has some say over how data are collected. In the field of satellite remote sensing, however, the resolution of a given sensor is defined *a priori*. This poses some difficulty, when the dominant scales of heterogeneity for a given process are not known, and the scale of measurement is inflexible. Given all of this, it becomes increasingly important to understand and document what the dominant scales are within a given image, and what physical process is responsible for the manifestation of the observations at that scale. In addition to determining the dominant length scales, it is also necessary to examine the changes in spatial patterns as the resolution of interest is changed. This requires the scaling characteristics of the data, which are necessary to identify correctly the appropriate method to assimilate the remotely sensed data into large-scale climate models.

Recent research (Kumar and Foufoula-Georgiou 1993a, b, Hu et al. 1998, Pelgrum et al. 2000) suggests that wavelet decompositions are powerful tools in analysing the ‘scaling behaviour’ of geophysical variables (for the purposes of this paper, any remotely sensed data field will be referred to as a geophysical field). The term ‘scaling behaviour’ refers to the statistical variation of a signal across different resolutions; e.g. self-similar (fractal), multifractal, etc. The scaling characteristics are then the parameters necessary to describe that behaviour. A key advantage of wavelet transforms over other forms of analysis is that wavelets allow for the breakdown of a signal into a scale frequency space. This permits easy determination of the relative contribution of different spatial scales present within the signal. In addition, this will provide information towards the choice of an appropriate technique to apply in aggregating data to coarser resolutions for assimilation.

Self-similarity is observed when data follow a power law spectrum (reference section 2). An additional benefit of wavelet analysis is that, if a process exhibits self-similar scaling behaviour, the wavelet coefficients obtained through a wavelet transform preserve that self-similarity (Kumar and Foufoula-Georgiou 1993a, b). Due to the preservation of this behaviour, wavelet coefficients allow for a convenient method of analysing the fluctuations between spatial resolutions within a data set, and thus the scaling exponents of the power law spectrum. If a process does not show self-similar behaviour, this methodology still permits analysis of the actual scaling behaviour through a multi-scaling framework.

Hu et al. (1998) used multi-resolution wavelet analysis to study the scale variation of soil moisture. In this study, soil moisture images were decomposed into average large-scale and detailed small-scale fluctuation components. The results suggested that the variation in soil moisture could be analysed as independent large- and small-scale processes. The small-scale fluctuations were analysed with moments versus scale plots. A surprising result of the research was that small-scale fluctuations
exhibited simple scaling, whereas the overall nature of the quantity exhibited multi-scaling characteristics.

Kumar and Foufoula-Georgiou (1993a, b) used multi-resolution wavelets to analyse the spatial characteristics of precipitation. The results of this study were also that the small-scale fluctuations exhibited simple scaling over a small range of scales, while over the remaining scales multi-scaling was exhibited.

Pelgrum et al. (2000) compared the average wavelet variance with an autocorrelation analysis for Daedalus data from the Jornada Experimental Range in southern New Mexico. The data consisted of surface reflectance, the Normalized Difference Vegetation Index (NDVI) and brightness temperature. For the semi-arid environment both analyses resulted in very small length scales (less than 30 m), which has implications for detecting variance when using satellite data. However, there was no examination of the issues of aggregation or variance added at scales which are measured by many satellite sensors.

The previous use of wavelets to examine the scaling characteristics of remotely sensed data has focused almost exclusively on smaller scales using airborne data. The research conducted here examines the issue of scaling characteristics at larger scales as determined by satellite sensors. The objectives of this research were to conduct a length scale analysis of large-scale satellite vegetation and radiometric temperature fields over the Southern Great Plains 1997 (SGP97) study area, as well as to investigate the presence of self-similarity in these data fields derived from the AVHRR sensor. This paper is organized as follows: section 2 contains a review of relevant scaling and wavelet theory, section 3 presents a description of the SGP97 Hydrology Experiment study area which was used for the analysis, section 4 gives the results of a wavelet-based fractal analysis of NDVI and radiometric temperature data, and the final section contains some concluding remarks.

2. Methodology

2.1. Wavelet transform and multi-resolution analysis

In order to estimate the scaling behaviour in the AVHRR data, the first step was to calculate the length scale of the data. The length scale is defined in this case as the scale with the highest wavelet variance as determined by the scalogram. The continuous wavelet transform of a real signal $f(x)$ where $x$ is a real variable is defined as

$$Wf(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x) \psi(\frac{x-b}{a}) dx$$

(1)

where $a$ is the resolution (scale) and $b$ is the spatial position within the original signal, $\psi$ is the mother wavelet, and the overbar designs the complex conjugate. The mother wavelet is any function which meets the so-called admissibility condition, i.e. it integrates to zero. In addition, the mother wavelet is usually chosen so that it has good localization in both time and frequency spaces, and exhibits some characteristics which are commensurate with the expected structure of the data series.

The wavelet transform is a highly redundant transform in scale space. Thus, a discrete wavelet transform reduces some of the redundancy by only conducting the transform at specific discrete resolutions, usually done on a dyadic ($2^n$) grid. A multi-resolution analysis is a method of projecting the original signal on to coarser and coarser resolutions. In practice the multi-resolution analysis is conducted over a
finite level of scales $j=0, \ldots, J$. The multi-resolution analysis results in a series of approximation signals at scale $J$ (coarsest resolution), and a detailed signal at all other levels of decomposition. Thus, the original signal at scale $j=0$ can be recreated by summing the coarsest resolution and the detailed coefficients at each scale

$$\text{Approximation}_{j=0} = \text{Approximation}_{j=J} + \sum_{j=1}^{J-1} \text{Detail}$$

(2)

The wavelet variance is then designated as $1/N \sum |Wf(a, b)|^2$ (Percival 1995) where $N$ is the number of data points in the original data set. The quantity $|Wf(a, b)|^2$ has been referred to as the scalogram (Kumar and Foufoula-Georgiou 1997) and, by extension, the cross-scalogram can be represented as $Wf(a, b)Wg(a, b)$ where the overbar is again the complex conjugate.

2.2. Self-similarity of wavelet coefficients

Following the terminology of Gupta and Waymire (1990), a random function $Y(x)$ of some variable $x$ is said to be a scaling function when the function at a different scale $Y(lx)$ modified by the scale factor ($l$) can be represented as follows

$$Y_l(x) = Y(lx)$$

(3)

If the relationship in equation (3) is a power law relationship, this leads to the following probability rule for the random field $Y(x)$

$$P[Y(lx) < y_1, \ldots, Y(lx) < y_n] = P[Y(x) < y_1, \ldots, Y(x) < y_n]$$

(4)

Simple scaling (also known as self-similarity) is then defined when the scaling exponent ($s(p)$) is a linear function of the order of moment $p$ (Hu et al. 1998). If the moments of the above exist, a result of simple scaling is

$$E[Y^p(l)] = l^{sp}E[Y^p(1)]$$

(5)

Gupta and Waymire (1990) demonstrated an important consequence of simple scaling: a log–log linearity between the logarithm of the moments with the logarithm of the scaling factor $l$ (achieved by taking a log transform of equation (5)) as

$$\log(E[Y^p(l)]) = s(p)\log(l) + \log(E[Y^p(1)])$$

(6)

Departures from simple scaling are termed multi-scaling, which is indicated if $s(p)$ is a non-linear function of the order of moment (usually determined in a least-squares framework). This results in a spectrum of scaling exponents, rather than the single exponent observed in a process which exhibits simple scaling behaviour. A general characteristic of multi-scaling of a geophysical variable is a concave (downward) curve when $s(p)$ is plotted against the order of moment on the log–log scale (Hu et al. 1998).

3. Description of data

The SGP97 Hydrology Experiment campaign was an interdisciplinary study conducted during the summer of 1997. It was designed, in part, to examine surface soil moisture at a variety of spatial scales. The location of the site covered much of the state of Oklahoma, and ranged from 97° W to 99° W and 34.5° N to 37° N (figure 1). During the course of the study, a variety of surface, airborne, and satellite measurements were collected.

This analysis only concerns the AVHRR satellite overpass on 2 July 1997.
image covers a larger extent (512 × 512 pixels, nadir resolution of 1.1 km) than the SGP97 study area. This extent was chosen after a preliminary Fourier spectral analysis (not presented) which indicated the possible presence of extremely large-scale processes. The image was georeferenced to the Universal Transverse Mercator (UTM) zone 14 N with the WGS1980 spheroid and NAD83 datum.

Atmospheric corrections to the red and near-infrared reflectance data were conducted using the Modtran radiative transfer model using a local radiosonde profile from the SGP97 data set. The red and near-infrared surface reflectance values were then used to calculate the NDVI

$$\text{NDVI} = \frac{\rho_{\text{nir}} - \rho_{\text{red}}}{\rho_{\text{nir}} + \rho_{\text{red}}}$$

The observed values of NDVI ranged from −0.195 to 0.820 with a mean value of 0.584 and a standard deviation of 0.107. Figure 2(a) presents the data. There is a general pattern of lower values in the western portion of the region and higher values in the eastern location.

The radiant temperature data were also atmospherically corrected with an incorporation of surface emissivity using Modtran following the method outlined in Brunsell and Gillies (2001). This correction resulted in a range of 24–54 °C with a mean of 38 °C. In general, the spatial pattern observed from the radiometric temperature is the opposite to that of the NDVI, with generally higher values in the west and lower in the more vegetated eastern region of the state.
4. Scaling behaviour of AVHRR NDVI and radiometric temperature

The 2 July 1997 AVHRR corrected data were used to examine the scaling behaviour in the SGP97 region. The multi-resolution analysis was conducted for eight levels of decomposition using the Daubechies 2 wavelet. The Daubechies wavelet was chosen primarily as it is a continuous wavelet. The decomposition resulted in one average or residual image ($J=8$) and detailed coefficients in each of the horizontal, vertical, and diagonal directions at each level of decomposition.

The detail images each contain half the resolution of the preceding level of decomposition. In order to view the effects of these changes, figure 3 presents the inverse wavelet transform at each level of decomposition with the coefficients at all other levels set to zero for the NDVI image. This shows the relative contribution of each scale to the observed signal. Figure 4 shows the same analysis for the radiometric temperature field.

Length scales were investigated by calculating the wavelet variance for both NDVI and radiometric temperature as presented in figure 5. Examination of figure 5
Figure 3. Multi-resolution decomposition levels of 2 July 1997 AVHRR NDVI. (a) Presents the coarsest resolution \( j = 8 \) and becomes progressively finer until \( (b) \ (j = 1) \) and presenting only the inverse wavelet transform at each level of reconstruction. This is a representation of what would be observed if only that scale was present in the signal.

NDVI field reveals a slight peak at the 128,000 m scale, while there is no clearly identifiable peak in the radiometric temperature field. One interpretation of self-similarity is the lack of any dominant length scale (i.e. a peak on figure 5). This lack of dominant scale in the wavelet variance is justification for analysing the self-similar nature of the observed radiometric temperature and vegetation index over the SGP region.

In order to examine the self-similar nature of the data sets, the first six moments were calculated for each level of decomposition in each direction based upon the wavelet coefficients produced by the multi-resolution analysis. The slope \( s(p) \) in equation (6) was estimated via linear regression from a plot of moment versus resolution. Table 1 presents the results of the regressions. In general, as the order of moment increases, the magnitude of the slope increases and \( r^2 \) decreases. The fit of
the linear regressions are all very high with the lowest being 0.986. The linearity of the regressions indicates the presence of self-similarity and there is no reason to question whether the fields exhibit self-similarity over the entire range of the analysis.

The estimated slope terms from the moment versus resolution regressions were then used to examine the scaling behaviour in each of the three directions that the wavelet coefficients were calculated. Linear regression techniques were again used to examine the relationship between order of moment and estimated slope terms (table 2). The values along with the regression lines are presented in figure 6.

For the NDVI data, each direction of decomposition exhibits self-similarity over the entire range of data analysed. It is interesting to note that the slope is approximately $-0.72$ in each of the three directions. This is the same order of magnitude found by Hu et al. (1998) in their examination of soil moisture variability in Little Washita using the ESTAR microwave data collected during Washita '92. The interpretation of the large values of the slope is that the NDVI field exhibits self-similarity
Figure 5. Wavelet variance of (a) NDVI and (b) radiometric temperature.

over the entire range of scales examined and the field exhibits correlation between data points over large scales.

The radiometric temperature field also exhibits self-similarity, but is more variable in the different directions the slope was calculated. The lowest magnitude is in the diagonal direction with $-0.90$ and the highest is in the vertical direction with $-1.0$. Again, the large values indicate the presence of large-scale correlation over the range of scales examined in this study.

5. Conclusions

The self-similar nature of vegetation and radiometric temperature fields was examined with the use of wavelet multi-resolution analysis. At relatively large scales, statistical self-similarity is observed through all levels of decomposition. Undoubtedly, there is a large amount of surface heterogeneity which is not observed at the scale of AVHRR data, therefore it is important to note that nothing can be said about the fractal nature of vegetation (as determined by AVHRR NDVI) at smaller scales. However, the AVHRR platform is a suitable candidate, due to the
Table 1. Regression results including slope $s(p)$, intercept $a$, and goodness of fit $r^2$ for order of moment versus resolution in horizontal, vertical, and diagonal directions for (a) NDVI and (b) radiometric temperature.

<table>
<thead>
<tr>
<th>Order of moment</th>
<th>Horizontal</th>
<th></th>
<th></th>
<th></th>
<th>Vertical</th>
<th></th>
<th></th>
<th></th>
<th>Diagonal</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$s(p)$</td>
<td>$a$</td>
<td>$r^2$</td>
<td></td>
<td>$s(p)$</td>
<td>$a$</td>
<td>$r^2$</td>
<td></td>
<td>$s(p)$</td>
<td>$a$</td>
<td>$r^2$</td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-1.01</td>
<td>9.10</td>
<td>0.993</td>
<td>1</td>
<td>-1.06</td>
<td>9.58</td>
<td>0.996</td>
<td>1</td>
<td>-1.11</td>
<td>9.71</td>
<td>0.992</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-1.91</td>
<td>17.87</td>
<td>0.994</td>
<td>2</td>
<td>-1.99</td>
<td>18.56</td>
<td>0.995</td>
<td>2</td>
<td>-2.05</td>
<td>18.41</td>
<td>0.994</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-2.71</td>
<td>25.89</td>
<td>0.993</td>
<td>3</td>
<td>-2.80</td>
<td>26.75</td>
<td>0.993</td>
<td>3</td>
<td>-2.87</td>
<td>26.17</td>
<td>0.993</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-3.40</td>
<td>33.21</td>
<td>0.991</td>
<td>4</td>
<td>-3.49</td>
<td>34.14</td>
<td>0.989</td>
<td>4</td>
<td>-3.57</td>
<td>33.13</td>
<td>0.990</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-4.03</td>
<td>40.02</td>
<td>0.989</td>
<td>5</td>
<td>-4.11</td>
<td>40.96</td>
<td>0.983</td>
<td>5</td>
<td>-4.19</td>
<td>39.49</td>
<td>0.985</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-4.61</td>
<td>46.48</td>
<td>0.986</td>
<td></td>
<td>-4.69</td>
<td>47.45</td>
<td>0.978</td>
<td></td>
<td>-4.76</td>
<td>45.49</td>
<td>0.979</td>
<td></td>
</tr>
<tr>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-1.09</td>
<td>13.45</td>
<td>0.997</td>
<td>1</td>
<td>-1.25</td>
<td>15.29</td>
<td>0.997</td>
<td>1</td>
<td>-1.17</td>
<td>13.79</td>
<td>0.997</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-2.18</td>
<td>27.45</td>
<td>0.994</td>
<td>2</td>
<td>-2.43</td>
<td>30.36</td>
<td>0.998</td>
<td>2</td>
<td>-2.26</td>
<td>27.28</td>
<td>0.998</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-3.21</td>
<td>41.23</td>
<td>0.992</td>
<td>3</td>
<td>-3.51</td>
<td>44.99</td>
<td>0.999</td>
<td>3</td>
<td>-3.27</td>
<td>40.34</td>
<td>0.997</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-4.17</td>
<td>54.51</td>
<td>0.990</td>
<td>4</td>
<td>-4.52</td>
<td>59.07</td>
<td>0.999</td>
<td>4</td>
<td>-4.18</td>
<td>52.72</td>
<td>0.996</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-5.06</td>
<td>67.27</td>
<td>0.988</td>
<td>5</td>
<td>-5.46</td>
<td>72.68</td>
<td>0.998</td>
<td>5</td>
<td>-4.98</td>
<td>64.26</td>
<td>0.995</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-5.90</td>
<td>79.61</td>
<td>0.986</td>
<td></td>
<td>-6.35</td>
<td>85.93</td>
<td>0.998</td>
<td></td>
<td>-5.70</td>
<td>75.15</td>
<td>0.994</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Slope $s(p)$, intercept $a$ and $r^2$ for the regressions of estimated slope versus order of moment in horizontal, vertical, and diagonal directions for (a) NDVI and (b) radiometric temperature.

<table>
<thead>
<tr>
<th></th>
<th>$s(p)$</th>
<th>$a$</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horizontal</td>
<td>−0.716</td>
<td>−0.438</td>
<td>0.993</td>
</tr>
<tr>
<td>Vertical</td>
<td>−0.721</td>
<td>−0.502</td>
<td>0.991</td>
</tr>
<tr>
<td>Diagonal</td>
<td>−0.724</td>
<td>−0.555</td>
<td>0.991</td>
</tr>
<tr>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horizontal</td>
<td>−0.961</td>
<td>−0.242</td>
<td>0.997</td>
</tr>
<tr>
<td>Vertical</td>
<td>−1.018</td>
<td>−0.360</td>
<td>0.997</td>
</tr>
<tr>
<td>Diagonal</td>
<td>−0.907</td>
<td>−0.421</td>
<td>0.994</td>
</tr>
</tbody>
</table>

Figure 6. Slope versus order of moment plots for horizontal, vertical, and diagonal wavelet coefficients. Lines are linear regression predictions for (a) NDVI and (b) radiometric temperature.

high temporal coverage (twice daily), for the application of assimilation of satellite-derived geophysical variables into global climate and large-scale meteorological and hydrological models.

Understanding the nature of scale in remote sensing is an important issue when extrapolations to larger resolutions is necessary (i.e. in the case of assimilation of data into meso-scale and global climate models). The knowledge that statistical self-similarity is observed in remotely sensed vegetation index and radiometric temperature over the SGP97 study region is useful for assimilation of these fields into large-scale models as well as for comparison of model output with the observed
satellite data. This suggests that one may have reasonable confidence in the aggregation procedure necessary for maintaining the statistical properties present within the original data, such that one is accounting for the observed results of the underlying physical processes.

Moreover, the high values of the slopes may lead to a larger scale view of the dynamics observed by coarser resolution sensors such as the AVHRR. In other words, these values indicate the existence of larger scale correlation over the entire range of scales examined in this paper. It should be noted that through the robustness of the scaling exponents (as determined by the $r^2$ values) the implication is that the observed properties of the data can be maintained through the aggregation procedure.

The knowledge of how biophysical variables vary in space is an important issue in many areas of science. The use of remotely sensed data to ascertain the observed scaling behaviour has implications for any large-scale ecological and the climatological modelling where model output is necessary at scales larger than that at which the data were collected. The results of this work (e.g. the statistical variability of vegetation and temperature with changes in spatial resolution) contribute to the basis for understanding how to assimilate remotely sensed data into these larger models. As a further example in the ecological field, research in phenology is examining the spatial distribution of the growing season length as a function of environmental forcing (M. White, personal communication). Thus, an understanding of the spatial scaling of the input parameters ideally will represent more realistically underlying physical processes. This will lead to more pragmatic agricultural considerations as to when to plant, length of growing season, etc. Furthermore, the incorporation into such models of representative scaled parameters may well lead to further insights into non-linear processes that affect the overall physiological mechanisms that underlie plant atmosphere systems.

Acknowledgment

This study was funded through a grant from NASA Cooperative Agreement no. NCC8-162 in support of the Global Energy and Water Cycle Experiment (GEWEX) Continental Scale International Project (GCIP) and the Utah Agricultural Experiment Station, Utah State University, Logan UT 83322-4810.

References


