Aggregation of a nonlinear land surface model for heterogeneous terrain

HENK PELGRUM
ESTEC, PO Box 229, 2200 AG, Noordwijk, The Netherlands
e-mail: hpeigrum@estec.esa.nl

Abstract The aggregation of a nonlinear land surface model for heterogeneous terrain from the pixel-scale to the scale of a large-scale atmospheric model is investigated. The aggregation process is quantified using a linearization approach based on Taylor expansion. A test case with remote sensing data from the Barrax site applied to the land surface model SEBI is presented. The error due to aggregation is 8% and can be approximated by the linearization approach.

Key words aggregation; EFEDA; energy balance; evaporation; linearization; SEBI; wavelets

INTRODUCTION

Land surface models describe the transport of water vapour and heat between the surface and the atmosphere. Remote sensing data (VIS, NIR, TIR) can be used to derive input data for land surface models. Land surface characteristics that can be derived from remote sensing data are the surface temperature, surface albedo and the normalized vegetation index (NDVI). Land surface models applied with high-resolution remote sensing data will produce detailed maps of the surface energy balance fluxes. The derived surface energy balance fluxes could be used as input in atmospheric models. However the grid cell size of an atmospheric model is much larger than the geometrical resolution of most remote sensing data. The results of the land surface model based on the remote sensing data therefore have to be aggregated to the scale of the atmospheric model. Two important aspects determine the aggregation process:

- the nonlinearity of the land surface model, and
- the heterogeneity of the land surface.

Only if the land surface model is linear and/or the land surface is completely homogeneous can the aggregation be performed by mathematical averaging of the results. However most land surface models are nonlinear and the land surface is never completely homogeneous. Under those conditions aggregation by simple averaging will introduce an error.

This paper describes how to quantify the effects of nonlinearity of the land surface model and heterogeneity of the land surface on the aggregation process. An approach based on Taylor expansion is discussed. A case study is presented where the land surface model SEBI (Surface Energy Balance Index) is used to calculate surface energy fluxes for a heterogeneous site in Barrax, Spain. The results of the SEBI model are aggregated and the effects of both the nonlinearity of SEBI and the heterogeneity of the Barrax site on the aggregation process are quantified.
DATA

Barrax is one of the sites included in the European Field Experiment in a Desertification threatened Area (EFEDA) field campaign (Bolle et al., 1993) held in June 1991. The Barrax site is 28 km north of Albacete. At the Barrax site the main characteristic is the presence of irrigated fields and non-irrigated fields next to each other. Micrometeorological measurements have been made on three different surfaces: bare soil, irrigated maize and fallow land. During the field campaign several aircraft equipped with remote sensing devices flew over the experimental area. The data used in this paper comes from the TMS-NS001 scanner aboard the NASA-ER2 aircraft. The TMS measures in the same spectral bands as the Landsat TM satellite. The TMS-NS001 data were recorded on 29 June 1991 and used to derive the following land surface characteristics: surface albedo $r_0$, surface temperature $T_0$ and NDVI.

AGGREGATION OF SPATIALLY DISTRIBUTED VARIABLES

Aggregation of results obtained from a spatial model from the local-scale to a large(r) scale by means of linear averaging can introduce errors. These errors are the results of (a) the nonlinearity of the model and (b) the heterogeneity of the land surface. Figure 1 illustrates the aggregation of spatially distributed variables.

Two different methods of aggregation from the local-scale to a larger scale can be distinguished: (a) the aggregation of the results which are derived from a distributed model $f$ using distributed input variables (path A in Fig. 1), and (b) the aggregation of
input variables before use in an aggregated model \( F \), thereby producing an aggregated result (path B in Fig. 1).

The aggregation of the results is illustrated by path A in Fig. 1. A distributed model \( f \) uses spatially distributed input variables \( p(x,y) \) as input (Arrow 1). The spatially distributed input data \( p(x,y) \) consists of \( n \) variables: \( p(x,y) = [p_1(x,y), p_2(x,y), \ldots, p_n(x,y)] \), where \( n \geq 1 \). Figure 1 shows an example with two variables \( n = 2 \). The results of the distributed model are denoted as \( f(p(x,y)) \) (Arrow 2). The distributed results \( f(p(x,y)) \) are then aggregated from the local scale to the large scale resulting in the aggregated result \( F \) (Arrow 3).

Along path B in Fig. 1 the spatially distributed input data \( p(x,y) \) are averaged from the local scale to the large scale by taking the arithmetic average, resulting in an averaged input \( \bar{p} \) (Arrow 4). The averaged input \( \bar{p} \) may consist of \( n \) averaged input variables: \( \bar{p} = [\bar{p}_1(x,y), \bar{p}_2(x,y), \ldots, \bar{p}_n(x,y)] \). The averaged input \( \bar{p} \) is then used as input in the aggregated model \( F \) (Arrow 5). The aggregated model \( F \) produces the aggregated result \( F(\bar{p}) \) (Arrow 6).

The difference between \( F \) and \( F(\bar{p}) \) will be zero if the distributed model \( f \) is linear and/or the input data is completely homogeneous. When neither condition is met a difference between \( F \) and \( F(\bar{p}) \) will occur. Since most land surface models are nonlinear and the land surface is never completely homogeneous differences between \( F \) and \( F(\bar{p}) \) are likely to occur. This is the error due to spatial aggregation by averaging spatially distributed variables. In this paper the difference between both aggregated results will be quantified using an approach based on linearization of the land surface model by means of a Taylor series (Hu & Islam, 1997).

### Aggregation analysis by linearization

A continuous model \( f \) depends on only one distributed variable \( p(x,y) \). The dependent model variable \( f(p(x,y)) \) can be approximated with a Taylor series, neglecting third and higher order terms. Then for any value of \( p(x,y) \):

\[
f(p(x,y)) = F(\bar{p}) + (p(x,y) - \bar{p}) \frac{\partial f}{\partial p}|_{p} + \frac{1}{2} (p(x,y) - \bar{p})^2 \frac{\partial^2 f}{\partial p^2}|_{p},
\]

where \( p^* \) is a function of \( p(x,y) \) and \( p^* \neq \bar{p} \) due to neglecting higher order terms. The input variable \( p(x,y) \) is defined over an area \( A \). Integration over \( A \) yields the correct mean of the distributed result:

\[
\frac{1}{A} \int_A f(p(x,y)) \, dA = \frac{1}{A} \int_A F(\bar{p}) \, dA + \frac{1}{A} \int_A (p(x,y) - \bar{p}) \frac{\partial f}{\partial p}|_{p} \, dA + \frac{1}{2A} \int_A (p(x,y) - \bar{p})^2 \frac{\partial^2 f}{\partial p^2}|_{p} \, dA.
\]

The first term on the right hand side is the value of the dependent model variable calculated with the average of \( p(x,y) \) i.e. \( \bar{p} : F(\bar{p}) \). The second term is by definition zero...
and the third term is then the difference between \( \bar{F} \) and \( F(p) \). In the third term both effects of the nonlinearity of the model and the heterogeneity of the land surface are incorporated. To separate both effects the mean value theorem is applied to this term:

\[
\frac{1}{2A} \int_A (p(x,y) - p)^2 \frac{\partial^2 f}{\partial p^2} \, dp \Bigg|_{p_e} \, dA = \frac{1}{2} \int_A (p(x,y) - p)^2 \, dA \Rightarrow \frac{1}{2} \, k \cdot V
\]

(3)

The factor \( k \) is the difference introduced by the nonlinearity of the function \( f \). Note that when \( f \) is a linear function the second derivative will be zero, therefore \( \bar{F} \) and \( F(p) \) will be identical. The nonlinearity term may be found directly from the model, if the function is continuous and therefore differentiable. The nonlinearity term cannot be determined exactly because \( p_e \) is unknown \textit{a priori}. The factor \( V \) accounts for the heterogeneity of the variable \( p(x,y) \) within the landscape. In the case of only one variable, \( V \) is equal to the variance of this variable. When \( p(x,y) \) consists of more variables, covariances also have to be taken into account. If \( p(x,y) \) is completely homogeneous within the area \( A \), the variance of \( p(x,y) \) will be equal to zero. As a consequence \( \bar{F} \) and \( F(p) \) will be identical. A method for calculating the variance and covariance of \( p(x,y) \) for different scales is described below.

**WAVELET ANALYSIS**

Wavelet analysis is a relatively new tool in geophysics (Kumar & Foufola-Georgiou, 1997). In this study a Discrete Wavelet Transform (DWT) will be used to decompose the Barrax dataset into an equally large set of wavelet coefficients. Each wavelet coefficient corresponds with a wavelet \( \psi \) of scale level \( j \) and position \( k \). In this study the Haar wavelet (Haar, 1910) has been chosen. The Haar wavelet was chosen in preference to other wavelets because of its suitability for capturing rapid changes in the data (Mahrt, 1991).

The wavelet coefficients are a measure of the intensity of the local variations of the signal, for the scale under consideration. The value of a coefficient will be large when the size of the wavelet is close to the scale of heterogeneity in the signal. The value of a coefficient will be negligible when the local signal is regular for that particular scale (Ranchin & Wald, 1993). The variance of the wavelet coefficients, the wavelet variance, is thus a natural tool for investigating the spatial scales of variability in remote sensing data (Percival, 1995). The wavelet variance is defined as:

\[
\sigma_{y,j}^2 = \frac{\sigma_{y,j}^2}{N} \left( \frac{1}{N} \sum_{k=1}^{N} D_{j,k}^2 \right)
\]

\( \sigma_{y,j}^2 \) is the sample wavelet variance of dataset \( y \) at scale \( j \) and \( D_{j,k} \) are the wavelet coefficients at position \( k \) and scale \( j \), and \( N \) is the number of elements in the total dataset. The number of data points at scale level \( j \) is given by \( n_j = N/2^j \). In Fig. 2 the wavelet variance for the three land surface characteristics derived from TMS-NS001 are shown: \( r_0 \), \( T_0 \) and NDVI. Most of the information is present at the length scales of 296 and 592 m, which is roughly the mean size of the pivot irrigation systems. The second peak at 2368 m is probably caused by the large-scale structure of irrigated vs non-irrigated surfaces.
RESULTS

The land surface model SEBI (Menenti & Choudhury, 1993) has been applied to the TMS-NS001 data set. The three input variables \( r_0 \), surface temperature \( T_0 \) and NDVI have been derived from the DAEDALUS dataset obtained on 29 June 1991. The results of SEBI have been compared with field measurements of the surface energy balance components. The root mean squared error (RMSE) for the evaporative fraction, \( \Lambda \), is 0.04, where \( \Lambda \) is the ratio between the latent heat flux \( \lambda E \) and the net available energy, i.e. the sum of latent and sensible heat flux \( H \).

For the aggregation analysis the following procedure was followed. First the SEBI model was applied to the original dataset with a resolution of 18.5 m. \( H \) and \( \lambda E \) were calculated and the corresponding \( \Lambda \) derived. These average values are the correct aggregated values, denoted in Fig. 1 (path \( A \)) as \( \overline{F} \). Then the input data \( r_0 \), \( T_0 \) and NDVI were resampled, changing the resolution each time by a factor of 2, resulting in input data at resolutions of 37, 74, 148, 296, 592, 1184, 2368, 4736 and 9472 m. For all these resolutions the SEBI model was applied, then the average \( H \) and \( \lambda E \) values were calculated also resulting in an average \( \Lambda \). In Fig. 1 these aggregated results are denoted as \( \overline{F(p)} \).

The difference between \( \overline{F} \) and \( \overline{F(p)} \) for \( \Lambda \) is plotted in Fig. 3, expressed as an error relative to \( \overline{F} \). The relative error increases with increasing resolution up to a value of 8%. The difference between \( \overline{F} \) and \( \overline{F(p)} \) can also be derived theoretically using equation (3). The factor \( V \) is derived from the wavelet analysis, for which the results are shown in Fig. 2. The factor \( k \), which is a measure of the nonlinearity of the SEBI model, was derived numerically. The only unknown is the value of the input parameter \( p_\alpha \). In this case \( p_\alpha \) consists of three variables: \( r_0 \), \( T_0 \) and the NDVI. Three possible options are chosen for \( p_\alpha \): the mean, median and modal values of \( r_0 \), \( T_0 \) and
NDVI. Figure 3 shows that the median value gives the best result, followed by the mean. The modal value is completely off in predicting no aggregation error, indicating that the modal values of $r_0$, $T_0$ and NDVI are not representative of the whole area.

CONCLUSIONS

The aggregation of a nonlinear land surface model for heterogeneous terrain is determined by two aspects: (a) the nonlinearity of the model, and (b) the heterogeneity of the terrain. Linearization of the model by means of a Taylor series shows that the aggregation can be quantified. The wavelet variance is a good measure of the heterogeneity of the land surface, whereas the nonlinearity of the model can be derived numerically. The Barrax test case showed that the aggregation error can amount to a value of 8%, which is relatively small for a heterogeneous area, compared to the error in field measurements of fluxes.

REFERENCES


