ELSEVIER

Journal of Hydrology 227 (2000) 128–139

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Characteristic length scale of input data in distributed models: implications for modeling grid size

G.A. Artan[a,*], C.M.U. Neale[b], D.G. Tarboton[c]

[a]*Raytheon STX Corporation, EROS Data Center, Sioux Falls, SD 57198, USA*
[b]*Biological and Agricultural Engineering, Utah State University, Logan, UT 84322, USA*
[c]*Civil and Environmental Engineering, Utah State University, Logan, UT 84322, USA*

Received 7 July 1998; accepted 13 October 1999

Abstract

The appropriate spatial scale for a distributed energy balance model was investigated by: (a) determining the scale of variability associated with the remotely sensed and GIS-generated model input data; and (b) examining the effects of input data spatial aggregation on model response. The semi-variogram and the characteristic length calculated from the spatial autocorrelation were used to determine the scale of variability of the remotely sensed and GIS-generated model input data. The data were collected from two hillsides at Upper Sheep Creek, a sub-basin of the Reynolds Creek Experimental Watershed, in southwest Idaho. The data were analyzed in terms of the semivariance and the integral of the autocorrelation. The minimum characteristic length associated with the variability of the data used in the analysis was 15 m. Simulated and observed radiometric surface temperature fields at different spatial resolutions were compared. The correlation between agreement simulated and observed fields sharply declined after a 10×10 m^2 modeling grid size. A modeling grid size of about 10×10 m^2 was deemed to be the best compromise to achieve: (a) reduction of computation time and the size of the support data; and (b) a reproduction of the observed radiometric surface temperature.

Keywords: Distributed models; Scaling; Aggregation; Remote sensing and GIS; Energy balance

1. Introduction

Spatially distributed hydrologic models are often point process models with distributed parameters that require large amounts of data for setting up model parameters. These data can be obtained from digitized maps or/and from remotely sensed imagery (Band et al., 1991; Bathurst and O'Connell, 1992). The spatial resolution of the input data and desired modeling grid size seldom coincide. And since there are no accepted objective methods for data aggregation, the spatial aggregation of model parameters is done ad hoc. Model parameters are then represented by one value at every grid cell, which is often the average value of the aggregated sub-pixels of the digitized/remote-sensed data. However, using the mean value of the model-input parameters does not necessarily yield the mean value of the model response variables (Bresler and Dagan, 1988a; Band et al., 1991).

The characteristic length of an image has been suggested as the appropriate scale of aggregation for data reduction by some researchers (Bian and Walsh, 1993; Brown et al., 1993). The characteristic length of

* Corresponding author. Tel.: +1-605-594-6195; fax: +1-605-594-6529.

E-mail address: gartan@edcmail.cr.usgs.gov (G.A. Artan).

0022-1694/00/$ - see front matter
PII: S0022-1694(99)00176-6

an image is defined as the distance at which, on average, two pixels no longer experience the same surface condition (Simmons et al., 1992). Simmons et al. (1992) used a plot of correlation to estimate the scale of variability of vegetation cover derived from SPOT and Thematic Mapper images of the shrub-steppe area of southeastern Washington with vegetation of blue bunch wheatgrass. They reported a scale of the order of 100 m. Bian and Walsh (1993) used the semi-variogram to determine the scale of vegetation and topography using data from a portion of Glacier National Park in Montana. The site had a complex terrain with a vegetation of Douglas fir (*Pseudotsuga menziesii*) and subalpine fir (*Abies lasiocarpa*) forests at lower elevations and tundra vegetation occupying the area above 2000 m. To identify the characteristic length, Bian and Walsh (1993) used images with different resolution produced by degrading Landsat images with an averaging window. They postulated that the scale of the NDVI was between 70 and 75 pixels, or 2100 and 2250 m.

Both, Simmons et al. (1992) and Bian and Walsh (1993), used sample transects to calculate the scale of variation from the images. A semi-variogram calculated from digital images of a Douglas-fir forest was employed by Cohen et al. (1990) to distinguish among stands. They found a range of 5–18 m that corresponded to the mean crown diameter of the trees. Curran (1988) used the semi-variogram to determine the minimum spatial resolution required for remotely sensed data. The semi-variogram has also been used to aid in simulating scenes (Woodcock et al., 1988a,b).

In this research, we investigated the scale of variability associated with a remotely sensed and GIS-generated high-resolution data set collected from a semi-arid complex terrain basin and how the response of a distributed energy balance model (Artan, 1996) of the basin is effected by the aggregation level of the input data. The remotely sensed data used in the analysis were images of the normalized difference vegetation index (NDVI) and the radiative surface temperature (RST). The NDVI was selected because it is highly correlated with LAI (Lulla et al., 1987; Artan and Neale, 1991). Latent heat fluxes from a land surface–atmosphere exchange model are sensitive to subgrid leaf area index (LAI) heterogeneity (Band et al., 1991; Bonan et al., 1993). On the other hand, the distribution of RST should be related to the distribution of the sensible heat flux. The GIS-generated data used in the analysis were the terrain slope and aspect angle, two parameters that influence the fraction of solar radiation received by the terrain. The data were analyzed in terms of the semivariance function, as well as the integral of the autocorrelation function (Vanmarcke, 1983).

The objective of this research is to attempt to advance an objective method for determining the optimum grid size for data aggregation, when setting up input data for distributed hydrologic models. In order to determine the optimum grid size when partitioning the watershed to model the hydrologic processes in a distributed manner, the guiding criteria should be: (a) minimize the computation time by reducing the number of grid cells; while (b) at same the maximizing the variation between grids in order to capture the significant patterns in the watershed (Band et al., 1991); and (c) keeping the nonlinear effects of subgrid heterogeneity on the model output to a minimum (Band et al., 1991; Bonan et al., 1993). We will describe how the characteristic length of the remotely sensed input data relates to the optimal modeling grid size specified in the above criteria.

2. Concepts and definitions

The prediction of a spatial process is subject to uncertainty due to the spatial variability and measurement errors of model parameters (Fienerman et al., 1985; Bresler and Dagan, 1988a; Band, 1989). Bresler and Dagan (1988a,b) described the effects of parameter uncertainty on a model of crop yield. In order to expand a point process model to an area, the soil and vegetation parameter (p_i) of the pixels is aggregated to the grid cell level for each location. Each grid cell is characterized by a multivariate density function, $f(p_1,p_2,\ldots,p_n)$ as a result of the aggregation of p_i over the space X (Bresler and Dagan, 1988a; Band, 1989):

$$\mathrm{E}[Y(X)] = \int^{p_1} \int \cdots \int^{p_n} Y(p_1,\ldots,p_n) f(p_1,\ldots,p_n;X)\, \mathrm{d}p_1 \cdots \mathrm{d}p_n \quad (1)$$

where $\mathrm{E}[Y(X)]$ is the expected value of the model response variable. If we assume that p_i are statistically

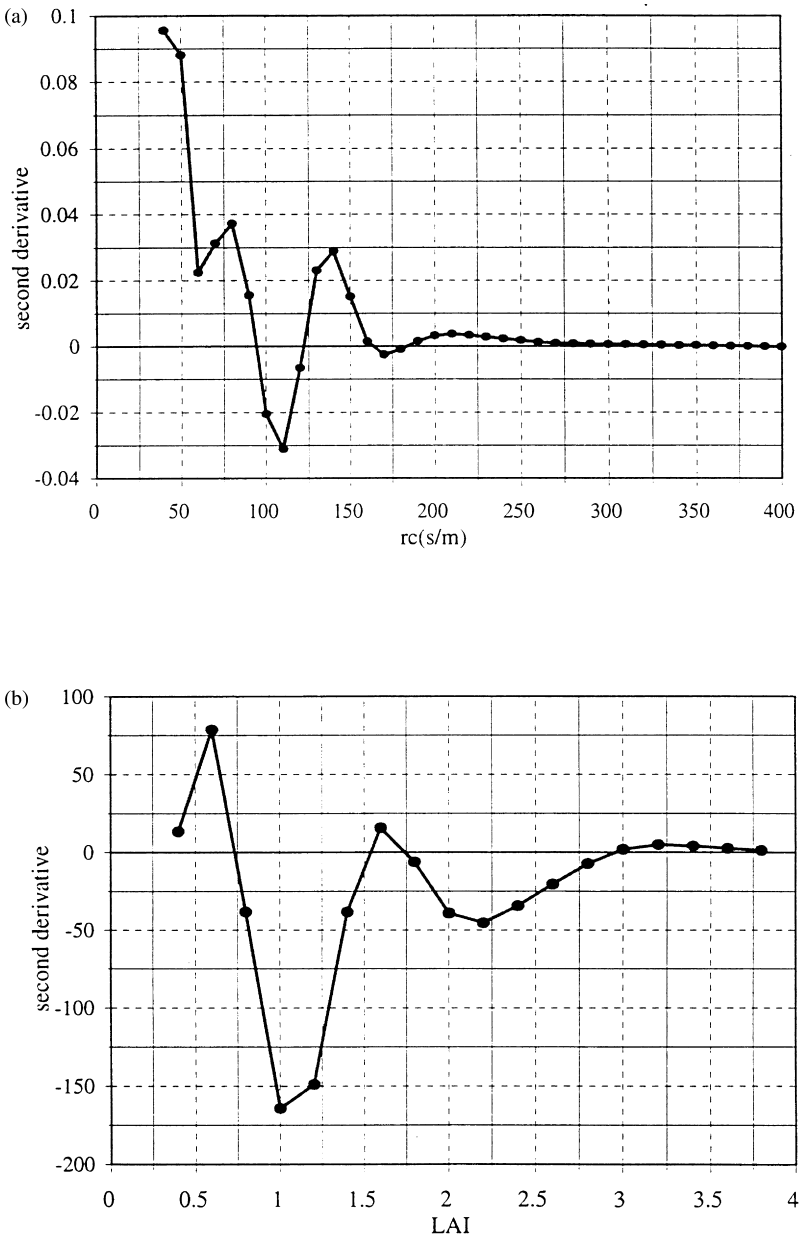

Fig. 1. Second derivative of the model output latent heat flux as a function of: (a) canopy resistance; and (b) LAI.

independent, then:

$$f(p_1, p_2, \ldots, p_n) = f(p_1)f(p_2)\cdots f(p_n) \quad (2)$$

If the variance of the parameters over the unit landscape is small, then we can expand Y in a Taylor series (Fienerman et al., 1985; Bresler and Dagan, 1988a; Band et al., 1991), and Eq. (1) can be simplified to:

$$\mathrm{E}[Y(X)] = y(\bar{p}_1, \ldots, \bar{p}_n) + \frac{1}{2}\sum_{i=1}^{n} \sigma_i^2 \frac{\partial^2 Y(\bar{p}_i, \ldots, \bar{p}_n)}{\sigma \bar{p}_i^2} + \mathrm{O}(\sigma_i^4) \quad (3)$$

where σ_i^2 is the variance of p_i, and $\bar{p}_i$ is the mean value of the parameters p_i. In distributed models at every grid cell the model parameters $\bar{p}_i$ are represented by one value, usually the average. To have $E[Y] \approx Y$ for model output using p_i as the only value to describe p_i over the space X requires that: (1) the model output is linear over the range of p_i; or/and (2) the variance of the parameters is small; or/and (3) a small value of $\partial^2 Y/\partial p_i^2$.

Fig. 1a and b shows the effect of the variation of LAI and canopy resistance (r_c) on the latent heat fluxes (LE) in the context of the model described in Artan (1996). The second derivative ($\partial^2 Y/\partial LAI^2$) values, where the second derivative is equal to the model bias, were obtained from the fit of a cubic spline function to the model response, produced by varying LAI or r_c while keeping all other model parameters constant at their mean value. Both curves tend to zero as LAI and r_c are increased, indicating that at higher values of the two parameters the dependence is linear, and using a mean value LAI and r_c is appropriate. It is apparent that at low LAI values, the error introduced by using only the mean value of LAI can be substantial.

3. Model overview

Since the distributed energy balance model used in this study has been described in great detail by Artan (1996), we will only describe the salient points of the model here. The energy balance model was developed for semi-arid regions. The canopy is assumed to be concentrated on a portion of the modeled area equal to the fraction of vegetation cover. The vegetation is represented as a block of constant density porous material for the radiation transmittance and turbulence transfer. Then the model solves separately for the bare and vegetation-covered part of the grid cell, the heat and moisture balance equations applied to one-canopy and two-soil layers by using coupled sets of five equations. The force-restore method (Deardorff, 1978) is used to solve the soil heat flux.

The resulting sets of equations are then solved for the prognostic variables of canopy temperature, air temperature at canopy source height, water vapor pressure at canopy source height, average temperature of the first soil layer, and average temperature of the second soil layer, by stepping in time from a known initial condition. The various fluxes (latent, sensible, and net radiation) are then calculated from the prognostic variables.

To run the model in distributed fashion, the basin under consideration should be subdivided into grid cells and meteorological, vegetation, and soil data defined for every grid cell, the state variables also should be initialized. The state variables that need initialization are the soil moisture content and temperature for the two soil layers. The model also requires that at each time step the following meteorological variables: air temperature, air vapor pressure, horizontal wind speed, and incoming solar radiation as a forcing input data. The vegetation parameters data expected by the model are the unstressed canopy conductance, leaf area index; average vegetation height, and the leaf reflection coefficients for the solar radiation. Finally, the soil parameters that are assigned a value prior to model run are the soil reflection coefficient for shortwave radiation and the soil bulk density. The distributed model uses a raster data structure. The model solution generally converges between three to five iterations for most of the runs we have conducted.

4. Site and data sources

4.1. The study area

The Upper Sheep Creek is a semiarid, high-altitude sub-basin, with an area of 26 Ha, of the Reynolds Creek Experimental Watershed. The basin is located in the Owyhee Mountains of southwestern Idaho about 80 km west of Boise. The basin is divided by a creek running in a southeast to northwest course. Vegetation communities on the site can be subdivided into three types: low sage/grass (*Artemisia arbuscula/Agropyron spicatum*), mountain big sage/grass (*Artemisia tridentata*/*Festuca idahoensis*), and an area occupied by aspen/willow (*Populus tremuloides/Salix* spp.) (Flerchinger et al., 1994). The vegetation on the southwest (SW)-facing side of the sub-basin was primarily composed of sparse sage/grass. The main ecosystem on the northeast (NE) part of the sub-basin used for this analysis was a mountain big sage/grass community composed of

Table 1
Resolution of remotely sensed data

DOY	NE		SW	
	NDVI (m)	RST (m)	NDVI (m)	RST (m)
156	0.20	0.39	0.20	–
230	0.28	0.55	0.28	0.55

patches of high-density vegetation area interspersed with gaps of low-density vegetation.

4.2. Remotely sensed and GIS-generated data

Airborne multi-spectral digital video imagery was collected twice during the summer of 1993 at different spatial resolutions using the airborne video/radiometer system (Neale, 1991) for the visible and near-infrared (NIR) imagery, and an Inframetrics 760 scanner for thermal infrared imagery. The multi-spectral video system consisted of three Cohu 4810 high-resolution video cameras (525 horizontal lines) with narrow band (10 nm) filters centered in the green (550 nm), red (650 nm), and NIR (850 nm) part of the spectrum. The cameras were set at the factory to override the automatic gain control (agc). The thermal imagery of energy emitted at 10–12 μm was used to obtain the RST. Images used in this analysis were acquired on day of year (DOY) 156 and 230 (June 5 and August 18) and had different nominal pixel size (see Table 1). All imagery was acquired close to solar noon, under cloudless skies.

The images were digitized at an 8-bit radiometric resolution and corrected for lens vignetting and effects of image motion. The individual band images of the visible and NIR bands were registered to a common image base and combined into three band images. The 3-band images were combined along the flight lines. NDVI images were prepared from the 3-band images using the red and NIR bands. Where the thermal imagery was corrected for the vegetation and soil substrate emissivity using a value of 0.96, which was the average measured for three sites at basin using a method described by Hipps (1989).

The data of the NDVI were normally distributed within the low sage/grass and mountain big sage/grass/aspen/willow vegetation communities area due to their diversity, but when the NDVI data of the whole basin were combined and plotted, the data had a bimodal distribution. The data from the two sites were analyzed separately; subset areas were extracted from the NDVI and thermal images to represent the vegetation communities of low sage and mountain big sage. In order to keep the pixel size constant in every image, the subset areas were kept between an elevation of 1860 and 1890 m.

A digital elevation model (DEM) with grid cell size of $2.5 \times 2.5\ m^2$ was used to calculate the aspect and slope angle images. The DEM was produced by digitizing a contour map of the basin with a scale of 1:1200 and contour interval of 0.604 m. From these images a subset of the NE-facing and SW-facing hillside was extracted. The aspect angle varies from 0 to 360°, with the north being both 0 and 360°. To have a data set of remotely sensed and GIS data of comparable support size of about 200×200 points, a bigger area was extracted for the aspect and slope angle images than the images of NDVI and RST used in the analysis.

Semi-variograms and spatial autocorrelation were calculated from the digital brightness of the Upper Sheep Creek imagery. Since the one-dimensional semi-variogram could be described more easily than the two-dimensional semi-variogram (Woodcock et al., 1988b), the one-dimensional was used. The semi-variogram from the images was calculated for all the rows and columns of the images and then averaged by using as a weight the number of observations in each lag (Cohen et al., 1990). The maximum lag used in this study was one-third of the numbers of rows and columns (Webster and Oliver, 1990).

4.3. Model runs

The response of the distributed energy flux model described in Artan (1996) to variation in the input spatial data was examined. A high spatial resolution data set of vegetation, soil, and terrain parameters derived from remote sensing and GIS were used as model input. Since a map of surface temperature was available only on DOY 230, a distributed model run was produced for this day only. The model was initiated at 14:20 MST DOY 229 with an output time of 11:00 MST DOY 230, that corresponded to about 20 h of model spin sufficient to lessen any initialization error effects on model response.

The model forcing meteorological variables (air temperature and humidity, horizontal wind speed, and incoming solar radiation) were from a 7-m tower installed in the NE-facing side of the basin, and two Bowen ratio energy balance systems installed in the NE-facing and SW-facing sides. Four R.M. Young anemometers (Model 12102 Gill 3 cup) and four Vaisala temperature/humidity probes installed on the tower at heights of 1.0, 1.8, 3.3, and 7.0 m from the ground. The incoming shortwave radiation was measured with two LI-COR pyranometers, installed with the Bowen ratio systems with one at each hillside. The horizontal wind speed, air temperature, and air vapor pressure used for model input were measured at 1.8 m above ground level. No variation was done for the horizontal wind speed and air vapor pressure between the grids. In regard to air temperature however, an adjustment was made for LAI of each grid cell using a method described by Running et al. (1987). Finally, the amount of incoming solar radiation on every grid cell was corrected for each grid according to the grid aspect and slope value. Because of the temporal resolution of the meteorological input data, the modeling time step was set to 20 min.

To simulate the changes in cell resolution, the model input data were aggregated to coarse resolution by using a nonmoving window. For example, an image of 4×4 pixels aggregated to the next level becomes a 2×2 pixels image. An aggregation was done by averaging all cells within a unit for parameters with numeric value (e.g. LAI, aspect and slope) and a generalization was implemented by recoding the grid cells with the class value occupying the majority of the cells of that grid (e.g. soil type data and vegetation class type). The model was run using different grid sizes (2.5, 5, 10, 15, 20, and 30 m).

5. Results and discussions

5.1. Video data

June 5 imagery. Due to the pattern of the snow and radiation distribution on the basin, the vegetation on the SW-facing side of the basin at that time of the year was close to peak greenness condition, whereas on the NE-facing side, the grasses, mostly composed of *Festuca idahoensis*, were green and the shrubs were just starting to green-up. The RST imagery of the SW-facing side was dropped from the analysis because of temperature saturation. The SW-facing side had much lower vegetation cover than the NE-facing side.

The semi-variogram calculated from the NDVI and RST imagery for the NE-facing site and the model fitted to it are shown in Fig. 2a and b, respectively, with a line at the top to indicate the variance of the image. The two semi-variograms exhibit similar structure. Both semi-variograms appear to pass through the origin, and to increase rapidly close to the origin. The range of the semi-variogram for the NDVI was estimated as 20 m using a double spherical model (Webster and Oliver, 1990), with an r^2 of 0.99. The range of the RST semi-variogram calculated with a spherical model was 15.8 m with r^2 of 0.97 between simulated and observed data.

The semi-variogram calculated from the NDVI imagery of the SW-facing side was unbounded and did not reach a clear range over the distance it was calculated. One reason for this was that there were large homogeneous areas in the imagery that were longer than the lags for which the semi-variograms were calculated. The square root of the characteristic area calculated from the spatial autocorrelation of each of the three above-mentioned images was compared with the range of the semi-variogram; these results are summarized in Table 2. The numbers were rounded to the nearest integer. The length scales for all variables were close to what was determined from the semi-variogram analysis.

August 18 imagery. On this image acquisition date, the vegetation on the SW-facing site of the basin was in pre-dormancy quiescence; the grasses on the NE-facing site were also in pre-dormancy quiescence, and the shrubs were still in a vegetative stage. The difference in vegetation density on the two sites of the basin is discernable from an LAI image (created from the NDVI image) shown in Fig. 3, with the SW-facing site vegetation having lesser LAI compared to the NE-facing site vegetation due to the sparser and smaller nature of the vegetation on that site. The semi-variogram for the SW-facing site had an unbounded exponential shape that did reach a clear still, due to the large homogeneous area of senescence vegetation present on the site. The semi-variogram of the NE site increased linearly, passing the variance of data due to

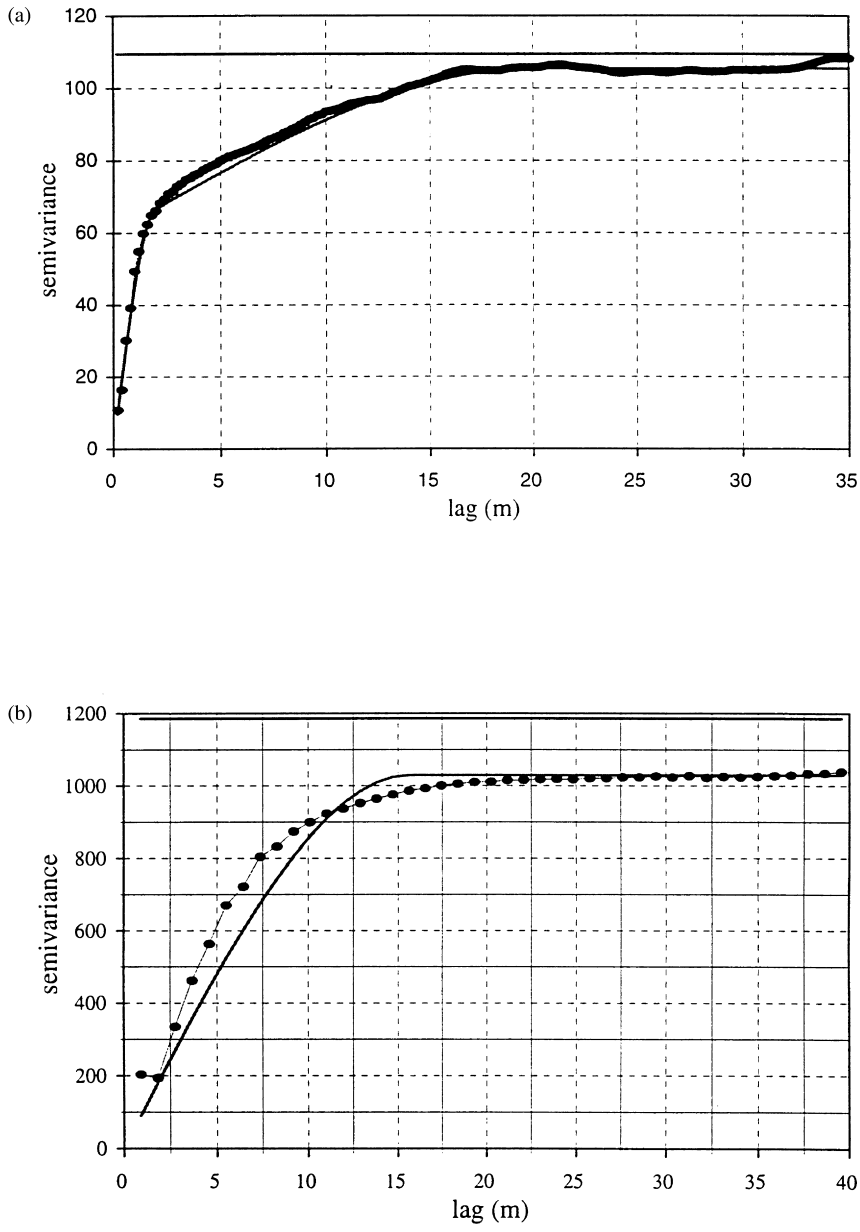

Fig. 2. Semi-variogram, fitted model, and estimated sill of imagery of the NE-facing site collected on June 5, 1993 for: (a) NDVI; and (b) radiometric surface temperature.

the contrast created by the large area dominated by grass or shrubs.

The semi-variograms resulting from the RST variable both exhibited the same pattern as the NDVI semi-variograms of the respective locations, with the semi-variogram of the NE-facing site having a range of about 41 m. The plots of the log–log semi-variogram for the RST variable had a steeper slope than the NDVI log–log semi-variogram plots. Since the evapotranspiration was negligible for both sites

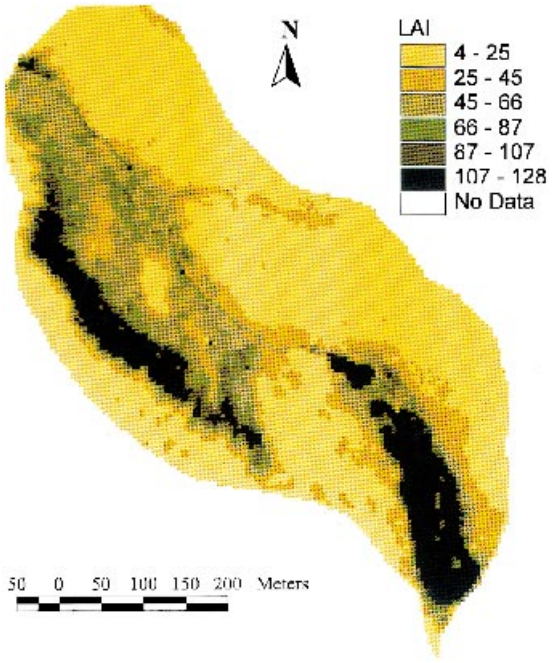

Fig. 3. The NDVI image of the whole Upper Sheep Creek sub-basin, NDVI prepared from the multi-spectral video imagery taken on 18 August, 1993.

(there was an array of three Bowen ratio energy balance systems installed in the basins), the temperature difference between canopy and soil substrata surface was minimal in both sites making the RST images more spatially dependent compared to the NDVI images of the same area.

Table 3 shows the characteristic lengths calculated with the spatial autocorrelation method for the data collected on August 18 (DOY 230). Characteristic length scale was the same on both sides of the basin for the NDVI, while for the RST image it was slightly larger on the SW-facing side. It must be noted, however, that the pixel resolution of images used in this analysis was not apparently influencing the results of the characteristic length scale. The thermal image pixel size is approximately twice the size of the short-wave image pixels. It should be noted that in the semi-variograms there were no visible break points at or near the characteristic length.

Table 2
Characteristic length scale for imagery of DOY 156

Site	NDVI (m)	RST (m)
NE	15	25
SW	22	–

Table 3
Characteristic length scale from imagery of DOY 230

Site	NDVI (m)	RST (m)
NE	24	24
SW	24	27

5.2. GIS data

Since the semi-variograms for slope angle and aspect angle for the two sites did not show any clear range, the break points in the log–log semi-variogram plot were used to infer the range of the processes (Bian and Walsh, 1993). The log–log semi-variogram for slope angle of the NE-facing site exhibits a break point at about 75 m suggesting a strong spatial dependence for distances less than 75 m. For the SW-facing site, there was a weak break point at around 75 m in the semi-variogram and the log–log plots. The break points in the semi-variogram are due to the nature of the basin hillsides, with every hillside having two predominant slopes, more so on the NE-facing site with a steeper drift area and a flatter area close to the stream.

The semi-variogram of the NE-facing site had a peculiar form with points that were close, having a bigger difference than points apart, which is an artifact of the way the aspect angle was calculated. Grid cells facing the south direction had an aspect of 180°, whereas the aspect value of a grid cell facing the north direction could be at the same time 0° (if the cell is approached from the east) or 360° (if approached from the west). The aspect data for the NE-facing site were dropped from the analysis for the semi-variogram. The plots for the SW-facing site did not have a clear range, but had two break

Table 4
Characteristic length scale from imagery of the slope and aspect angle

Site	Slope (m)	Aspect (m)
NE	65	87
SW	76	85

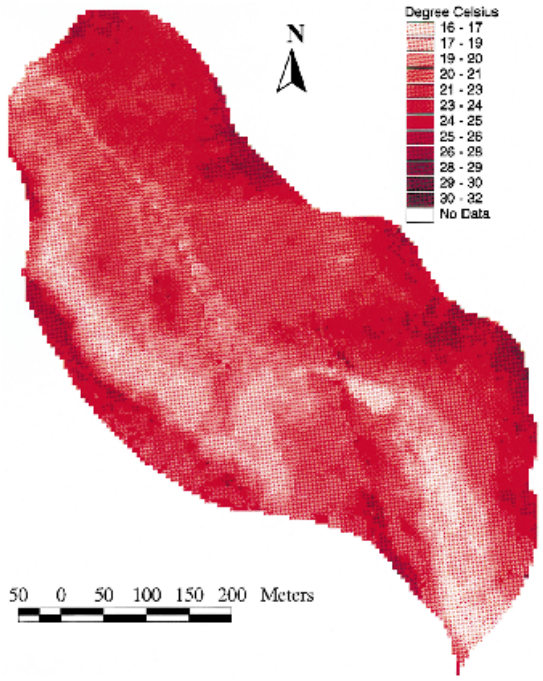

Fig. 4. Basin simulated radiometric surface temperature for grid size of 5×5 m^2.

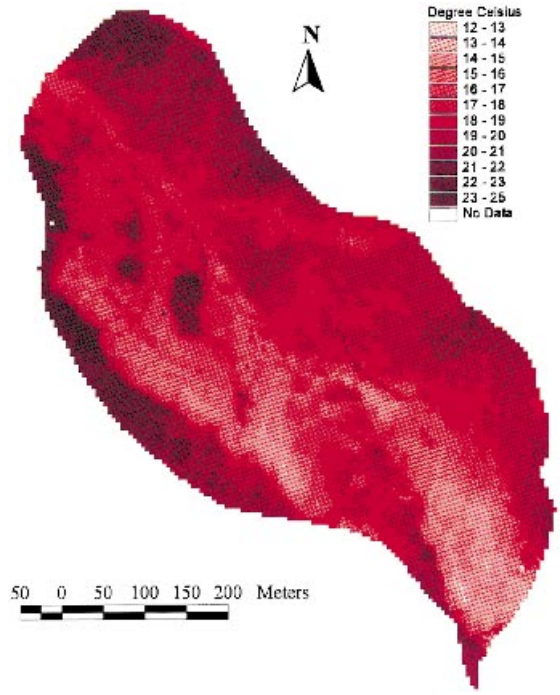

Fig. 5. Basin observed radiometric surface temperature for grid size of 5×5 m^2.

points at about 12 and 110 m. The characteristic length scales calculated with the integral of the autocorrelation for the slope of both sites and aspect of the SW-facing site are shown in Table 4.

5.3. Simulation results

The model presented in Artan (1996) was tested for the sensitivity of input data aggregation and generalization by comparing simulated and observed surface temperature for different levels of data resolution. The model uses a raster data format. An observed surface composite (canopy and soil) temperature image was used to assess the sensitivity of the model for data spatial aggregation resolution. This surface temperature was produced by simulating an instrument with a nadir viewing angle. The observed temperature for the different resolutions was made by aggregating the observed RST image with the pixel resolution of 0.55×0.55 m^2. The simulated and observed temperatures for a modeling grid size of 5 m are shown in Figs. 4 and 5. The simulated surface temperatures captured most of the spatial pattern of the observed surface temperature for this modeling grid size. Worthy of notice is that there is a minor geographical misregistration of about one pixel between the two temperature maps. Due to the complex nature of the topography of the basin, it was difficult to register all the maps to the state plane.

However, Fig. 6 shows that there is less agreement between the simulated and observed values of temperature, as the grid size increases. The decrease of the correlation between simulated and observed is significant between a grid cell size of 10 and 15 m. Beyond the 15 m grid cell, the changes in correlation between the simulated and observed surface temperatures were minimal. The disagreement between the simulated and observed RST, as the grid size was increased, is an indication of the effects of subgrid variability on the RST derived from the distributed model. From the analysis of both the autocorrelation lengths and the semi-variograms of the model input data, the shortest characteristic length observed was 15 m.

From this modeling exercise, it is apparent that the optimal grid size for input data aggregation in the

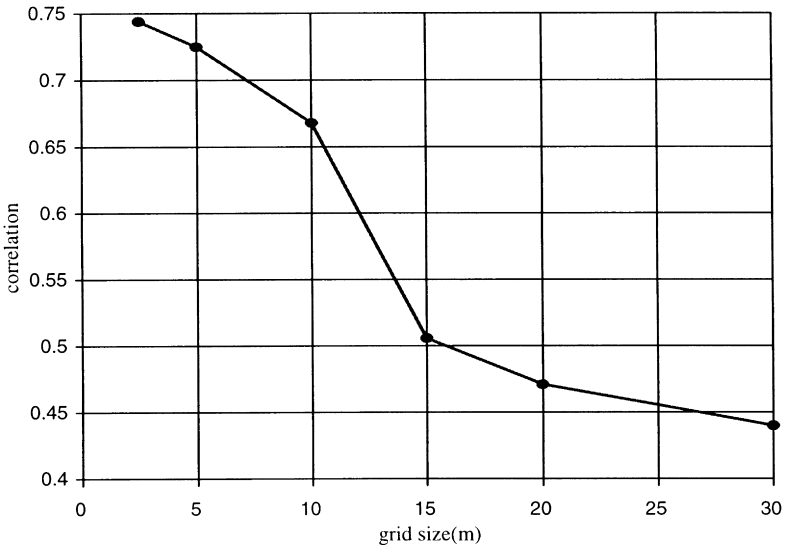

Fig. 6. The r^2s between the simulated and observed radiometric surface temperature.

context of a distributed energy flux model and the data from a complex terrain semi-arid basin was around 10×10 m^2, somewhat smaller than the characteristic length estimated for the input data. This model inadequacy with increased modeling grid size is a manifestation of the effects of increased value of the variance on the model parameters. The model inability to simulate the observed surface temperature was also due in part to the correlation between some model input variables, that were understandably high, since they were derived from the same data set. Table 5 shows the correlation between some of the vegetation, terrain, and soil parameters at a grid size of 5×5 m^2.

6. Summary and conclusions

Most of the sample semi-variograms exhibit the absence of a sill because of spatial trend in the data, making it impossible to predict the range of the processes from the semi-variograms. The semi-variograms of the NDVI and RST differed only slightly. From the break points in the semi-variograms of the aspect and the slope angle it was possible to infer scale of variability much longer than the scales associated with the NDVI and RST. Despite the differences in vegetation cover present at the two sites, the pattern of spatial dependence was the same for both sites. The double-logarithmic plot of the semi-variograms reveals one degree of spatial dependence for scales up to about 2 m, and another for greater lengths.

Both characteristic length scale and semi-variogram range were functions of the vegetation conditions and, to a lesser extent, influenced by the pixel size of the imagery and the remotely sensed

Table 5
Correlation between model input parameters for an aggregation level of 5×5 m^2

Parameter	Aspect angle	g_c^*	Elevation	LAI	Soil albedo
Aspect angle	1				
g_c^*	0.42	1			
Elevation	−0.06	0.15	1		
LAI	−0.48	0.93	0.07	1	
Soil albedo	−0.513	−0.39	0.04	−0.48	1

variable used in the calculation, with the semi-variogram range usually greater than the characteristic length.

The characteristic length associated with the variability of radiometric surface temperature (RST) calculated from the integral of the autocorrelation was greater than 20 m for both SW- and NE-facing sites. Whereas the minimum associated with the variability of the NDVI was 15 m. From the semi-variogram analysis of the NDVI and RST images, the shortest length of variability was determined as 15.8 m for RST and 20 m for NDVI. These results would suggest that for the basin under study, modeling at a 15-m grid cell resolution is sufficient to capture the variability of NDVI and radiometric surface temperature, or their surrogate latent and sensible heat flux.

To see if the characteristic length scale calculated from the remotely sensed and GIS-generated imagery (some of the imagery was used as model input data) corresponded to the optimal modeling scale, according the rules outlined in the introduction, the spatially distributed model was run using input data that were aggregated to a different grid size, and the simulated RST was compared with an observed surface RST field. A model grid size of around $10 \times 10\ m^2$ was the best compromise because: (a) the reduction of computation time and the size of the support data; and (b) a reproduction of the observed RST. In conclusion, the characteristic length scale of input for a distributed model gives an idea of the best modeling scale.

Acknowledgements

This research could not have been completed without the assistance and cooperation of the USDA-ARS Northwest Watershed Research Center field personal. We also thank Padinare V. Unnikrishna and Thomas H. Jackson for their assistance during the data collection period. Thanks are also due to Konstantine P. Georgakakos and two anonymous reviewers for insightful criticism of an earlier draft of this work. This work was funded in part by the USGS Department of the Interior, grant number 14-08-0001-G2110, and US EPA, award number R 824784-01-0. Also the senior author received some support from the Hydrologic Research Centre, San Diego. The view and conclusions are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the US government.

References

Artan, G.A., Neale, C.M.U., 1991. An assessment of video-based vegetation indices for the retrieval of biophysical properties. In: Blazquez, C.H. (Ed.). Proceedings of the 14th Biennial Workshop on Color Aerial Photography and Videography in Plant Sciences, American Society of Photogrammetry and Remote Sensing, Orlando, FL, pp. 135–146.

Artan, G.A., 1996. A spatially distributed energy flux model based on remotely sensed and point-measured data. Thesis, Utah State University, Logan, p. 159.

Band, L.E., 1989. Spatial aggregation of complex terrain. Geogr. Anal. 21, 279–293.

Band, L.E., Peterson, D.L., Running, S.W., Coughlan, J., Lammers, R., Dungan, J., Nemani, R., 1991. Forest ecosystem processes at the watershed scale: basis for distributed simulation. Ecol. Mod. 56, 171–196.

Bathurst, J.C., O'Connell, P.E., 1992. Future of distributed modeling: the systeme hydrologique Europeen. Hydr. Proc. 6, 265–277.

Bian, L., Walsh, S., 1993. Scale dependencies of vegetation and topography in a mountainuous environment in Montana. Prof. Geogr. 45, 1–11.

Bonan, G.B., Pollard, D., Thompson, S.L., 1993. Influence of subgrid-scale heterogeneity in leaf area index, stomatal resistance, and moisture on grid-scale land-atmosphere interactions. J. Clim. 6, 1882–1897.

Bresler, E., Dagan, G., 1998a. Variability of yield of an irrigated crop and its causes 1. Statement of the problem and methodology. Water Resour. Res. 24, 381–387.

Bresler, E., Dagan, G., 1988b. Variability of yield of an irrigated crop and its causes 2. Input data and illustration of results. Water Resour. Res. 24, 389–394.

Brown, D.G., Bian, L., Walsh, S., 1993. Response of a distributed watershed erosion model to variation in input data aggregation levels. Commmun. Geosci. 19, 499–509.

Cohen, W.B., Spies, T.A., Bradshaw, G.A., 1990. Semivariograms of digital imagery for analysis of conifer structure. Rem. Sen. Environ. 34, 167–178.

Curran, P.J., 1988. The semivariogram in remote sensing: an introduction. Rem. Sen. Environ. 24, 493–507.

Deardorff, J.W., 1978. Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation. J. Geophy. Res. 83, 1889–1903.

Fienerman, E., Bresler, E., Dagan, G., 1985. Optimization of a spatially variable resource: an illustration for irrigated crops. Water Resour. Res. 21, 793–800.

Flerchinger, G.N., Cooley, K.R., Seyfried, M.S., Wright, J.R., 1994. A lumped parameter water balance of a semi-arid watershed,

paper presented at the ASEA International Summer Meeting, ASEA, Kansas City, MO, June.

Hipps, L.E., 1989. The infrared emissivities of soil and *Artemisia tridentata* and subsequent temperature corrections in a shrub–steppe ecosystem. Rem. Sen. Environ. 27, 337–342.

Lulla, K., Mausel, P., Skelton, D., Kramber, W., 1987. An evaluation of video-band-based vegetation indices. In: Everitt, J.H. (Ed.). Proceedings of the 11th Biennial Workshop on Color Aerial Photography and Videography in Plant Sciences, American Society of Photogrammetry and Remote Sensing, Weslaco, TX, pp. 270–279.

Neale, C.M.U., 1991. An airborne multi-spectral video/radiometer remote sensing system for agricultural and environmental monitoring, paper presented at the ASAE Symposium on Automated Agriculture for the 21st Century, Chicago, IL.

Running, S.W., Nemani, R.R., Hungerford, R.D., 1987. Extrapolation of synoptic meteorological data in mountainous terrain and its use for simulating forest evapotranspiration and photosynthesis. Canad. J. For. Res. 17, 472–483.

Simmons, M.A., Cullinan, V.I., Thomas, J.M., 1992. Satellite imagery as a tool to evaluate ecological scale. Land. Ecol. 7, 77–85.

Vanmarcke, E., 1983. Random Fields: Analysis and Synthesis, MIT Press, Cambridge, MA, p. 190.

Webster, R., Oliver, M.A., 1990. Statistical Methods in Soil and Land Resource Survey, Oxford University Press, New York, p. 316.

Woodcock, C.E., Strahler, A.H., Jupp, D.L.B., 1988a. The use of variograms in remote sensing: I. Scene models and simulated images. Rem. Sen. Environ. 25, 349–379.

Woodcock, C.E., Strahler, A.H., Jupp, D.L.B., 1988b. The use of variograms in remote sensing: II. Real digital images. Rem. Sen. Environ. 25, 349–379.