GODDARD GRANT
IN-43-CR
108743
14 P.

Semi-Annual Report

The Correspondence of Surface Climate Parameters with Satellite and Terrain Data

Principal Investigator
Jeff Dozier

Co-Investigator
Frank Davis

Department of Geography
University of Califonia, Santa Barbara
Santa Barbara, California 93106

(NASA-CR-181515) THE CORRESPONDENCE OF SURFACE CLIMATE PARAMETERS WITH SATELLITE AND TERRAIN DATA Semiannual Report (California Univ.) 14 p Avail: NTIS HC A03/MF A01 CSCL 08B G3/43

N88-12185

Unclas 0108743

This document outlines work in progress for the NASA grant "The Correspondence of Surface Climate Parameters with Terrain and Satellite data." Our research to date has been concentrated in the following areas:

1) Appropriate sampling designs for FIFE
2) Field location of sample sites
3) Database development
4) Spatial analysis

1. FIFE Sampling Locations

One of the goals of our research is to develop a ground sampling strategy for calibrating remotely sensed measurements of surface climate parameters. Our initial sampling strategy involved a stratification of the terrain based on important ancillary surface variables such as slope, exposure, insolation, geology, drainage, fire history, etc. For a spatially heterogeneous population, sampling error is reduced and efficiency increased by a) stratification of the landscape into more homogeneous sub-areas, and b) employing periodic random spacing of samples.

We applied these concepts in the initial stratification of the FIFE study site for the purpose of locating and allocating instrumentation. This stratification is important since it constrains subsequent areal integrations of point ground measures and the accuracy of these integrations.

The FIFE site was divided into eight strata based on topography and fire history. The strata were chosen to represent four topographic regimes -- plateau, valley bottom, moderate slope (3 - 8°), steep slope ($> 8^\circ$) -- and burned and unburned. Digital elevation data for the Swede Creek Quadrangle were used to estimate the frequency distribution of the topgraphic regimes, from which appropriate classes were chosen. Figure 1 shows a map of categorized slope classes. The allocation and location of instrumentation over the site is given in Table 1.

2. Field Location of Sample Sites

For sampling error to be minimized, it is important that strata be sampled proportionately to stratum size and variability. However, for small samples, maximal information is provided by periodic random spacing of sample locations such that samples are spread evenly over the site. The distribution given in Table 1 was the initial allocation of sample site locations for FIFE. It is a compromise between sampling the FIFE site evenly and sampling proportional to stratum size. Several investigators subsequently relocated their sites, moving them in some cases completely out of the the terrain class they were allocated to. This necessitated relocating each site so that a) a new frequency distribution estimate could be made and, b) the experimental ground samples could be tied to a precise geographic location for subsequent spatial analysis of the data.

Davis completed an exhaustive survey of each sampling location during IFC1. Each site was surveyed as accurately as possible onto 1:24000 map sheets. Information on slope, exposure, fire history, vegetation type and cover, grazing history were also collected at this time. The sites were then digitized from the map sheets and UTM easting and northings obtained. Table 2 summarizes this information, and as far as we know is the most accurate location data available.

3. Database Development

One of the crucial questions we are addressing in ISLSCP is that of error propagation thru several layers of data. For example, in assesing a satellite's ability to measure soil moisture, remotely sensed data must be compared with "actual" ground data. However, this ground data in must be integrated from point measurements into a map with accompanying cartographic and interpolative errors. If intermediate layers of data are used, each with their own set of spatial errors, error propagation can become serious. The errors at the finest scales must be known in order to predict errors at coarser scales employed for some analyses.

In addition, for certain types of calculations on terrain grids the errors in the grid are compounded by the actual computation. In the case of slopes, the usual method of computation produces a slope map with a spatial accuracy much worse than the original grid.

To help with this problem a 10 m DEM has been produced at our suggestion. The USGS 30 m Swede Creek DEM was plagued with errors and simply did not have the required spatial resolution. It is our belief that the spatial processes being studied in FIFE vary significantly at scales below 30 m. The 10 m DEM will allow us to study these processes and the errors at a scale comensurate with the highest resolution remotely sensed data. This in turn will greatly enhance our ability to accurately measure the association of terrain, satellite, map and other ancillary data with DEM derived data.

We are also concerned with the spatial accuracy of the multidate SPOT imagery since we intend to map fire history and vegetation from it. We chose to process the raw SPOT data at UCSB to minimize the errors in the rectification and registration of these images.

In rectifying the SPOT images we chose to forego any type of bulk geometric processing to correct for skew and instead relied solely on ground control points. We also ignored look-angle effects judging these to be insignificant given the width of the FIFE site and the maximum relief evident. The images were rectified using VICAR and net of 30 control points.

4. Fire History and Vegetation Mapping

We believe there will be a strong association of certain surface climate parameters (soil moisture, for one) with vegetation and fire history. We are therefore in the process of creating maps of fire history and vegetation.

We are mapping fire history using field notes and SPOT imagery. Unfortunately, only two dates are of use in mapping fire history: 6-Mar-87 and 27-Jun-87. Other dates have only partial coverage due to clouds and/or satellite position or are temporally redundant. A principal component analysis of the data will highlight vegetation differences based on burned and unburned areas.

Our vegetation maps will be based on SPOT imagery and air photos and will be field checked.

5. Spatial Analysis

Any efforts to produce areal estimates of surface climate parameters over large, heterogeneous areas from ground point measurements must exploit the scale-dependent systematic association of surface climate parameters with terrain and other mapped data. These associations, once discovered can be used to stratify a landscape into more homogenous units facilitating sampling and stratification. However, these associations *among* variables are difficult to find. We are exploring such techniques as mutual information analysis and regression trees for finding associations.

The first step in our attempt to understand these associations however, is to identify the *scale* and *pattern* of variation for any single variable. Certain variables may vary quite slowly across the landscape; others may change value in relatively short distance. In either case knowledge of this spatial variation is critical.

We have developed a theoretical framework for understanding these variations which will be presented at a later date. On the experimental level, two exploratory techniques have been used: spectral analysis and geostatistics.

Geostatistics

Any variable which is distributed in space is said to be "regionalized". This definition is purely descriptive. Generally, regionalized variable theory is associated with a model of spatial variation based on the semi-variance and semi-variogram and is often referred to as *geostatistics*. This is distinct from the more usual measures of spatial variation used in Geography which are based on the spatial covariance, the spatial autocorrelation and the correlogram. In either case, we are interested in describing the spatial variation associated with a random spatial function.

We can briefly describe the *semi-variance* as follows. Consider a random spatial function $Z(\mathbf{x})$, where $\mathbf{x}$ is a position vector. Assuming second order stationarity we have for all $\mathbf{x}$:

$$\mathrm{var}[Z(\mathbf{x})] = E\{[Z(\mathbf{x}) - m]^2\} = C(0)$$

$$\gamma(\mathbf{h}) = \frac{1}{2}E\{[Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})]^2\} = C(0) - C(\mathbf{h})$$

The quantity, $\mathbf{h}$, is the spatial lag between the value of the random function at $Z(\mathbf{x}_1)$ and $Z(\mathbf{x}_2)$. $C(0)$, the covariance at zero lag, is simply the variance of the spatial process and is called the *a priori* variance. The semi-variance, $\gamma(\mathbf{h})$, is equal to the variance minus the covariance at lag $\mathbf{h}$.

A plot of $\gamma(\mathbf{h})$ versus $\mathbf{h}$ shows how the variance of a function changes with distance. The distance at which the semi-variance stops increasing marks the effective range of the autocorrelation of that function. Values of the function are not related to neighboring values beyond this distance.

We have created variograms of the Swede Creek DEM for elevation, slope and exposure. These are shown in Figures 2, 3, and 4.

Autocorrelation

An alternative method of characterizing a spatial process is thru the autocorrelation function $\rho(\mathbf{h})$.

The autocorrelation at lag $\mathbf{h}$ is defined as the covariance at lag $\mathbf{h}$ divided by the *a priori* variance. It is the proportion of the total spatial variation explained by the variation of the lagged process.

$$\rho(\mathbf{h}) = \frac{C(\mathbf{h})}{C(0)} \tag{5}$$

The plot of $\rho(\mathbf{h})$ versus $\mathbf{h}$ is often called the *correlogram* . The autocorrelation can also be defined in terms of the variogram function as:

$$\rho(\mathbf{h}) = \frac{C(\mathbf{h})}{C(0)} = 1 - \frac{\gamma(\mathbf{h})}{C(0)} \tag{6}$$

The Fourier transform of the covariance function (actually the auto-covariance) is the perhaps more familiar *power spectrum.* The power spectrum shows the amount of variation explained at a given spatial frequency. We have calculated the power spectrum of the Swede Creek DEM for elevation, slope and exposure (Figures 5, 6, 7).

Under the hypothesis of second-order stationarity the covariance, correlogram, and variogram are entirely equivalant methods for characterizing the auto-correlation at a lag of $\mathbf{h}$.

The hypothesis of second-order stationarity assumes the existence of the covariance, which in turn implies that the *a priori* variance is stationary and hence finite, that is, var $Z(\mathbf{x}) = C(0)$. In fact many terrain processes are markedly not stationary-- they show an infinite capacity to vary as the area of interest increases. They have neither an *a priori* variance nor a covariance and hence the autocorrelation cannot be defined. In these cases the variogram can be used. However, preliminary analysis has not shown this to be the case for any of the variables so far studied.

6. Continuing Research

The completion of the 10-meter DEM will enable us to proceed with the spatial analysis portion of the research. It is likely that there is significant variation in certain variables (radiation, for example) at scales below the 30-m model presently used. The techniques outlined above will determine if such variation is present.

The integration of ground measurements will proceed based on the observed variance of the data and terrain stratifications. The semi-variogram can be used to *krige* the scatterd data into map form. We will also explore techniques which utilize conventional interpolations and also those which exploit the association of the data with other ancillary data.

Finally, we will evaluate the preliminary terrain stratifications in light of the data obtained. We can throw away data and produce new integrated maps and compare these with maps obtained from intensively sampled areas. All of this will lead to the refinement of a regional sampling strategy applicable in other experimental sites.

Categorized Slopes

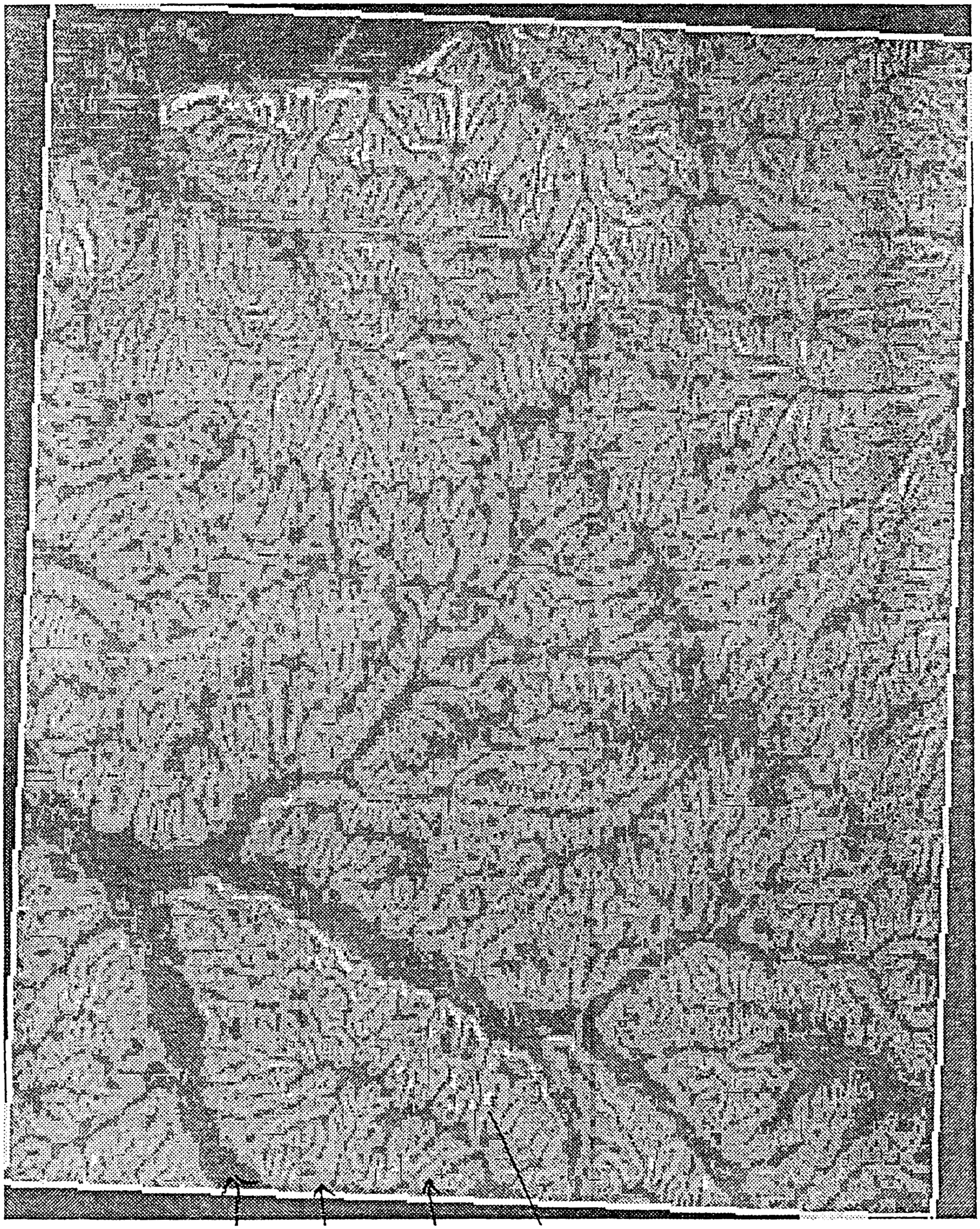

Figure 1. Map of categorized slopes derived from the Swede Creek DEM. Grey levels correspond to categories. Numbers are in degrees.

Proposed Location of Instrumentation
over FIFE Study Site *

Stratum		Proportion of Site	Relative Variability	No. of BR Stations		No. of AMS stations	
				In Konza	Out	In Konza	Out
Burned	Plateau	.07	Low	0[1]	3[3]	1	1(3)
	Valley Bottom	.11	Low	0	1	0	0
	Moderate Slope	.41	Moderate	1	2[2]	0	2
	Steep slope (n)	.07	High	0	1	0	0
	Steep slope (s)	.07	High	0	1	0	1
Unburned	Plateau	.02	Low	0	0	(1)	0
	Valley Bottom	.04	Low	1	0	0	1
	Moderate Slope	.14	Moderate	3	3	2	1
	Steep slope (n,w)	.05	High	1	0	1	1
	Steep slope (s)	.05	High	0	0	0	0
Agric. valley bottom				0	0	0	1
Sub-total				6[1]	11[5]	4(1)	8(3)
Total				17[6]		12(4)	

* — Numbers in brackets are stations for eddy correlation measurements, numbers in parentheses are "super-AMS" stations

Table 1. Original allocation of sampling sites for FIFE.

Sample Site Locations				
Site #	N. W. Latitude	Longitude	UTM Northing	UTM Easting
01	39 05 00	-96 33 32	4328669.00	711149.69
02	39 05 54	-96 35 32	4330256.00	708223.56
03	39 05 20	-96 33 56	4329291.00	710571.13
04	39 04 57	-96 33 33	4328602.50	711139.25
05	39 05 41	-96 34 38	4329887.50	709522.44
06	39 05 35	-96 33 24	4329772.50	711309.13
07	39 04 37	-96 34 55	4327924.00	709169.13
08	39 04 28	-96 33 53	4327680.00	710676.31
09	39 03 36	-96 34 53	4326050.50	709267.19
10	39 04 17	-96 35 51	4327278.50	707841.44
11	39 03 05	-96 32 30	4325165.00	712734.06
12	39 04 46	-96 35 42	4328174.00	708041.69
13	39 00 39	-96 33 04	4320640.00	712042.00
14	39 05 13	-96 35 42	4328997.50	708029.81
15	39 01 29	-96 34 16	4322158.00	710278.13
20	39 01 07	-96 32 24	4321554.00	712977.19
22	39 02 59	-96 36 42	4324816.50	706681.69
23	39 00 52	-96 28 26	4321246.00	718718.00
24	39 00 25	-96 36 20	4320095.00	707341.25
26	38 58 30	-96 32 33	4316702.00	712898.69
27	39 06 11	-96 28 58	4331042.00	717687.19
28	39 00 19	-96 32 01	4320061.50	713573.81
29	39 06 56	-96 31 10	4332359.00	714459.69
30	39 03 13	-96 28 26	4325598.00	718604.94
31	39 05 36	-96 32 22	4329851.00	712797.44
34	39 03 59	-96 26 48	4327081.50	720917.50
36	39 04 58	-96 30 07	4328767.50	716069.50
38	39 06 13	-96 26 54	4331196.50	720648.63
40	39 06 34	-96 31 21	4331670.50	714225.31
42	39 06 17	-96 31 26	4331122.50	714109.25
44	39 05 40	-96 31 48	4329981.00	713616.06

Table 2. Sample Site locations field checked during IFC 1. Co-located sites are coded as follows:

Stations 11,16, and 18 together are coded as **11**.
Stations 17 and 22 together are coded as **22**.
Stations 19 and 24 together are coded as **24**.
Stations 21 and 26 together are coded as **26**.
Stations 25, 30 and 32 together are coded as **30**.

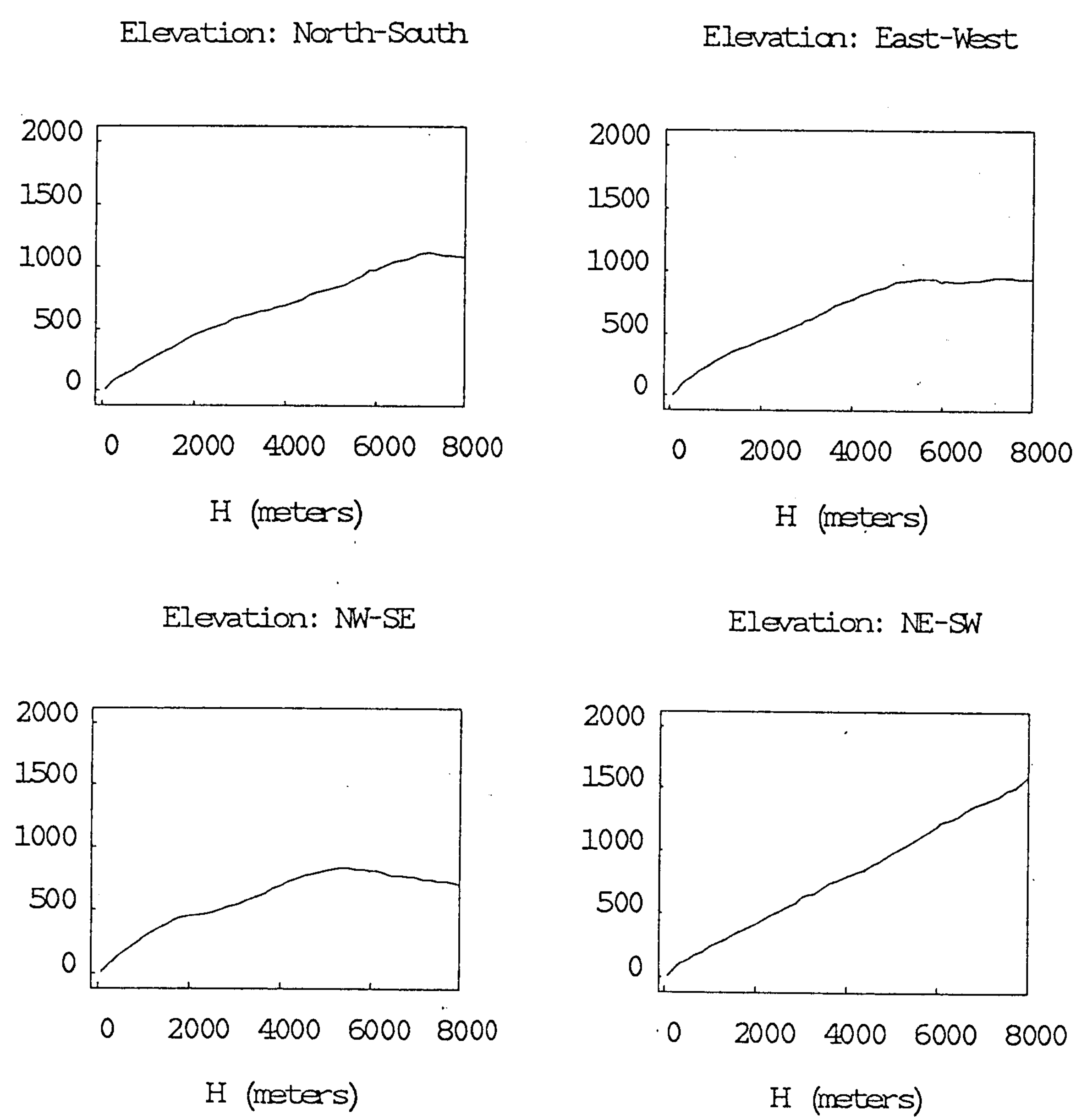

Figure 2. Semi-variograms of Swede Creek DEM Elevation Data. Variograms are calculated for each of four directions shown. Y-axis values are the semi-variances.

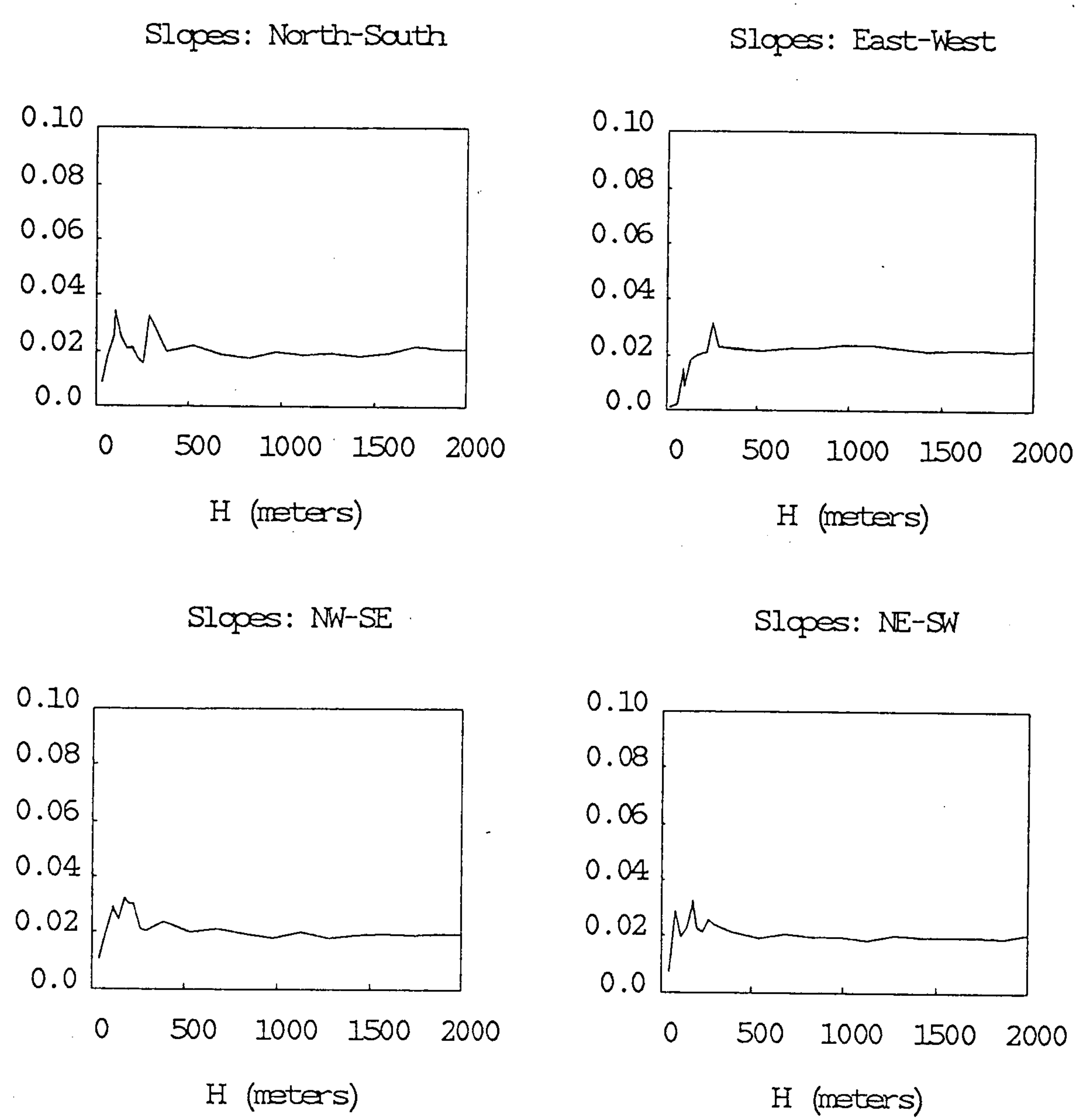

Figure 3. Semi-variograms of Swede Creek DEM Slope Data. Variograms are calculated for each of four directions shown. Y-axis values are the semi-variances.

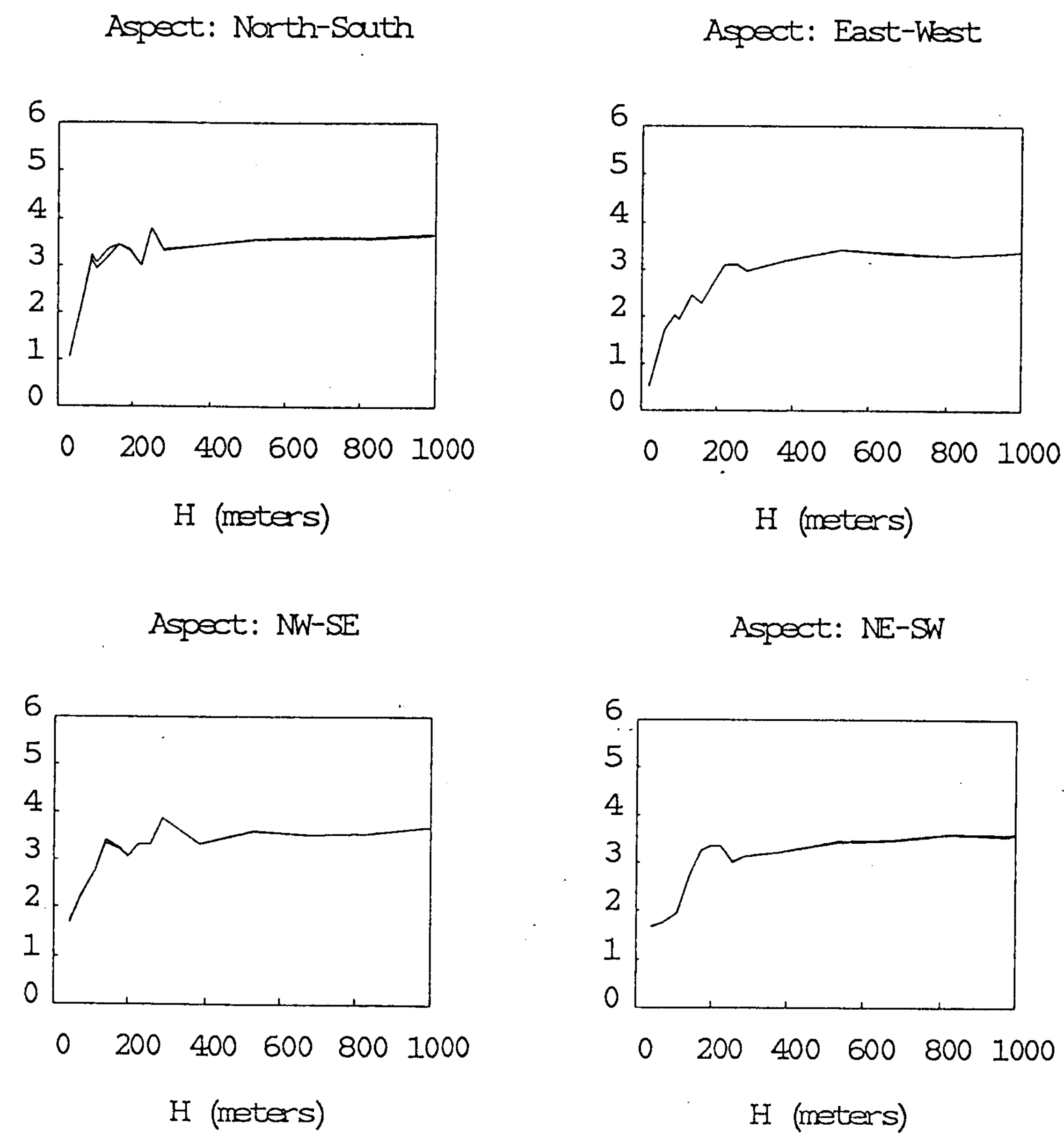

Figure 4. Semi-variograms of Swede Creek DEM Aspect Data. Variograms are calculated for each of four directions shown. Y-axis values are the semi-variances.

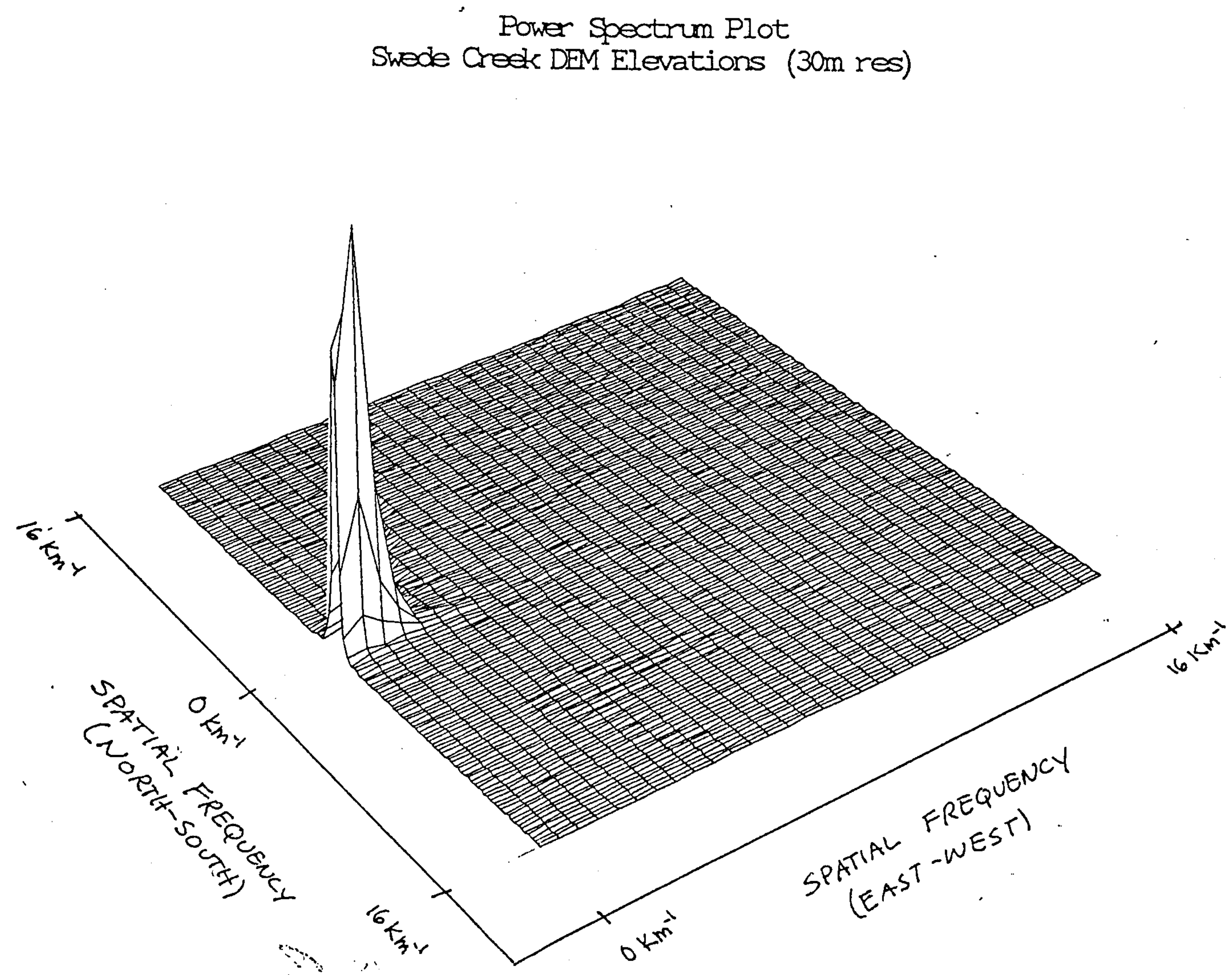

Figure 5. Power spectrum plot. Y-axis value is the power (variance explained) at any given spatial frequency. The inverse of the spatial frequency gives the spatial wavelength or lag.

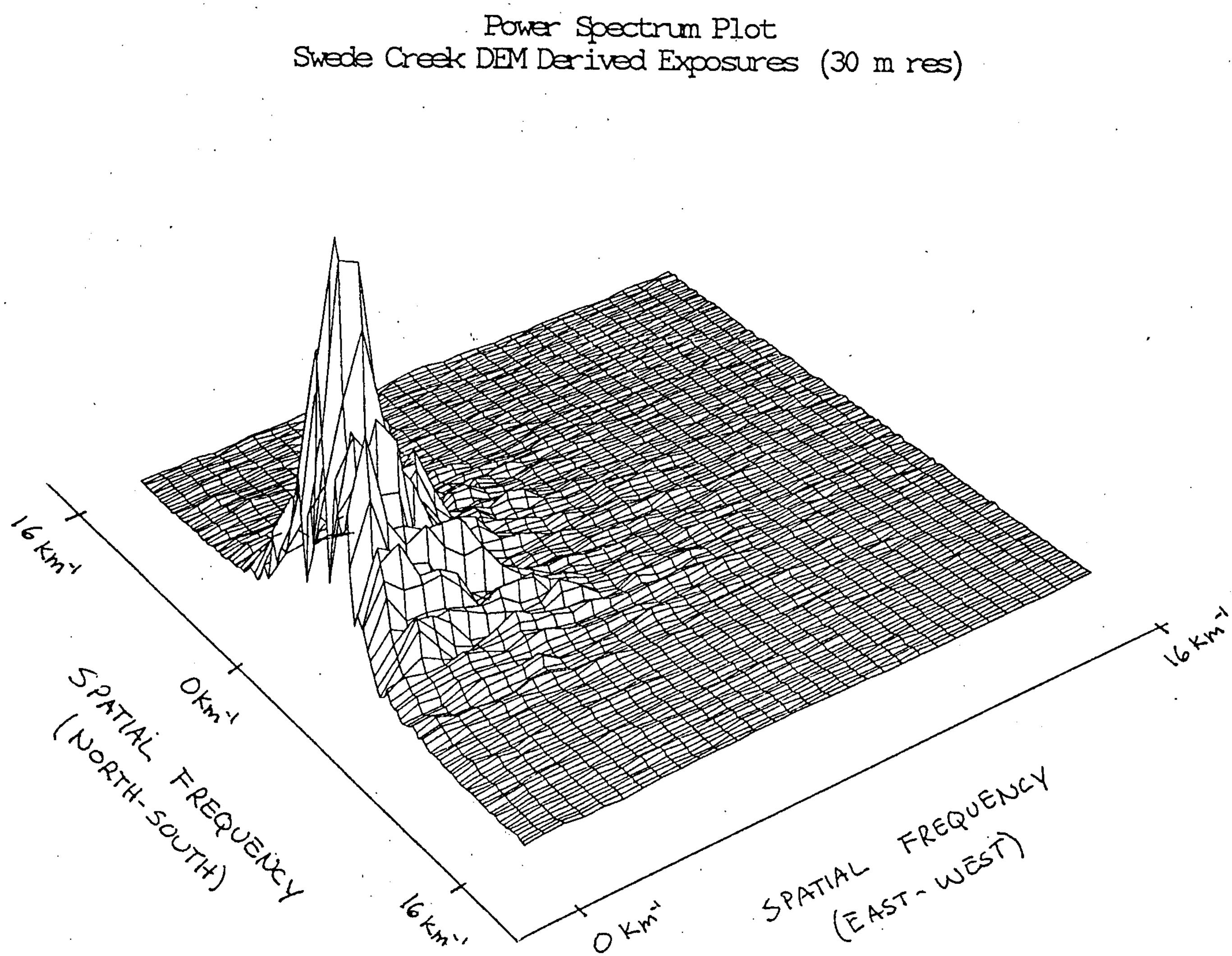

Figure 6. Power spectrum plot. Y-axis value is the power (variance explained) at any given spatial frequency. The inverse of the spatial frequency gives the spatial wavelength or lag.

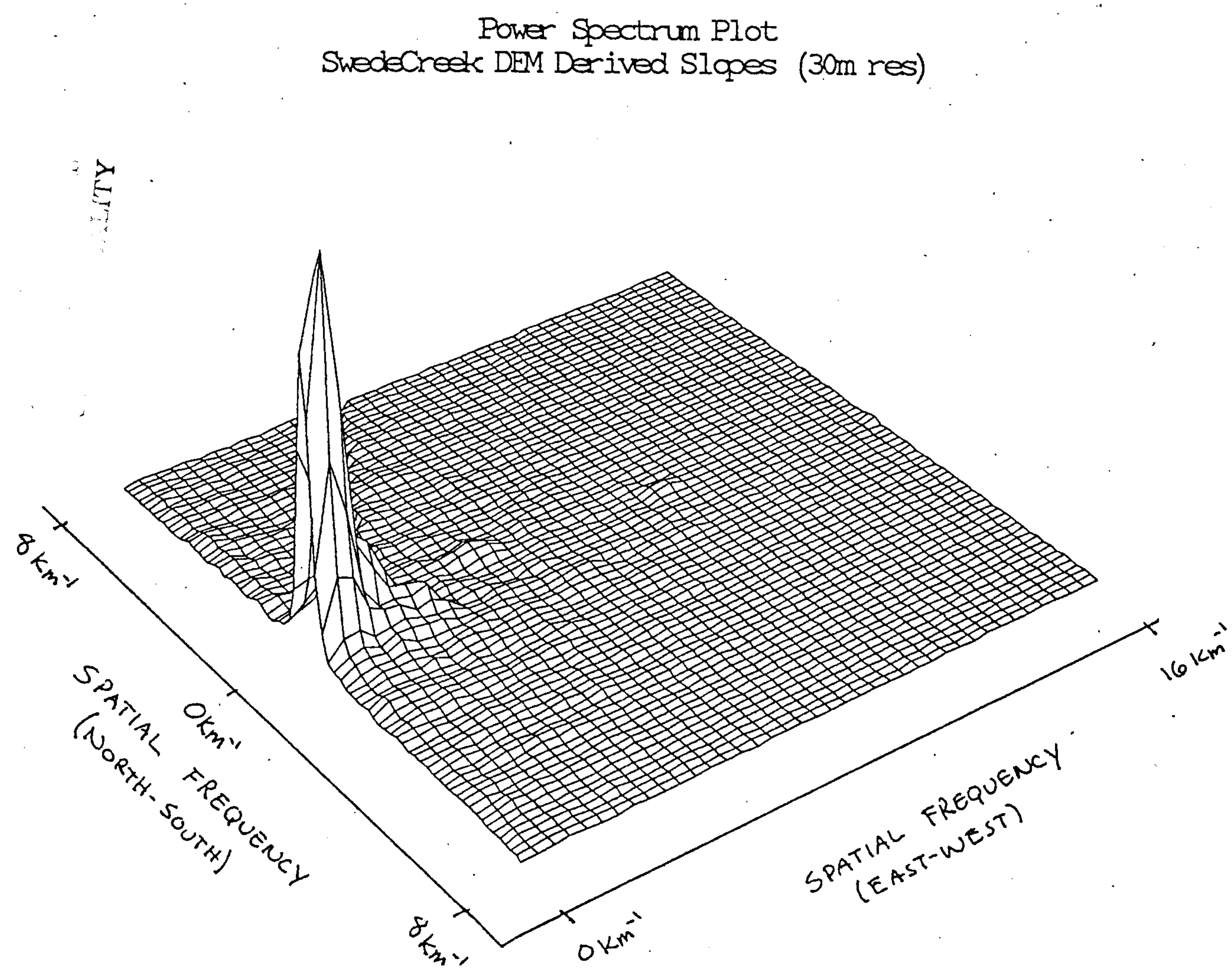

Figure 7. Power spectrum plot. Y-axis value is the power (variance explained) at any given spatial frequency. The inverse of the spatial frequency gives the spatial wavelength or lag.