Available online at www.sciencedirect.com
ScienceDirect
Advances in Water Resources 30 (2007) 1094–1104
Advances in Water Resources
www.elsevier.com/locate/advwatres

Evolution of the variability of surface temperature and vegetation density in the great plains

Michael H. Cosh [a,*], Jery R. Stedinger [b], Steve C. Ou [c], Kuo-Nan Liou [c], Wilfried Brutsaert [b]

[a] Hydrology and Remote Sensing Laboratory, Rm 104 Bldg 007, BARC-West, USDA-ARS, Beltsville, MD 20705, USA
[b] School of Civil and Environmental Engineering, Cornell University, 220 Hollister Hall, Ithaca, NY 14853-3501, USA
[c] 7127 Math Sciences, Department of Atmospheric Sciences, University of California, Los Angeles, 7127 Math Sciences, Los Angeles, CA 90095, USA

Received 6 March 2006; received in revised form 6 October 2006; accepted 9 October 2006
Available online 5 December 2006

Abstract

This study focuses on how the variability of land surface temperature and vegetation density at the SGP ARM-CART site changes over episodic (day to day) and seasonal time scales using AVHRR satellite data. Four drying periods throughout the year are analyzed. Land surface temperature had an erratic relationship with time exhibiting no deterministic pattern from day-to-day or season-to-season. Furthermore, it did not exhibit spatial pattern persistence. On the other hand, vegetation density had a consistent spatial pattern and temporal decay during average length drying periods (less than 7 days) as well as within a season. However, there were distinct differences in the seasonal pattern of variation between winter and growing seasons. In addition, the paper highlights a methodology to quantify the relationships that exist at the land surface between the primary parameter of interest and the controlling variables.

Keywords: Hydrologic data; Land surface temperature; Hydrometeorology; Vegetation; Statistics

1. Introduction

Land surface temperature and vegetation density are important surface characteristics for estimating hydrologic surface fluxes. The development of remote sensing technologies has enabled intense study of these two characteristics. Several studies have indicated that the scale at which these characteristics are parameterized is too large to be adequately described by the first moment alone [8,9]. Much progress has been made in recent years on methods to incorporate spatial variability in flux formulations [1,7,13]. These parameterization schemes tend to assume a static description of variability; however, in the real world, this variability rarely retains the same characteristics over time. Some aspects of the temporal evolution of spatial variability were analyzed by Chen et al. [4] and Chen and Brutsaert [3]. Further investigation into the spatial pattern persistence of land surface parameters was investigated by Cosh et al. [5]. In the present paper, some of these ideas are reconsidered; accordingly, a systematic methodology is put forth to prioritize the significance of different land surface parameters, specifically land surface temperature and NDVI, to determine which characteristics are most important in modeling surface flux spatial variability.

The main objective of this paper is to show the evolution of variability of vegetation density and land surface temperature at the time scales of a drying period (day-to-day) and across seasons. The study uses standard statistical tools, including moment analysis, regression analysis, aggregation analysis, and correlation analysis to determine whether the variability of the surface parameters is consistent and stable, and statistically predictable. Also, the predictability of a variable can be assessed

* Corresponding author. Tel.: +1 301 5046461; fax: +1 301 5048931.
E-mail address: mcosh@hydrolab.arsusda.gov (M.H. Cosh).

0309-1708/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advwatres.2006.10.003

Nomenclature

The following symbols are used in this paper

C_k	coefficient of kurtosis
C_s	coefficient of skewness
e_{ijtL}	residual
i	land cover index
j	site index
L	latitude
n	number of observations
R^2	coefficient of determination
r_k	correlation coefficient
s	standard deviation of data
t	time index
x_i	data value at point i
$\bar{x}$	average of data
Y_{ijtL}	value of variable (*LST* or *NDVI*) at land cover i, site j, for time i, for latitude L
Y_{ijt}	value of variable at land cover i, site j, for time i
$Y_{ij(t+k)}$	value of variable at land cover i, site j for time $t+k$
$\overline{Y}_{\bullet\bullet t}$	average of variable for all land covers, for all sites, and for time t
$\overline{Y}_{\bullet\bullet(t+k)}$	average value of variable for all land cover types for all sites, for time $t+k$
α	coefficient for time t
β	coefficient for latitude L
δ_{ij}	categorical variable for land cover i and site j
ϕ	a constant
γ_i	categorical variable for land cover type i
η_{it}	interaction term between categorical time t and land cover type i
ρ_{NIR}	reflectance in the near infrared band (0.7–1.1 μm)
ρ_{VIS}	reflectance in the visible band (0.4–0.7 μm)
κ_t	categorical variable for time t
υ	coefficient for interaction term for time t and latitude L
ψ_i	coefficient of interaction term for latitude L and land cover i and
ζ_i	coefficient for interaction between time t and land cover i.

versus several land surface components, and their correlation can be measured. This will determine how a land surface variable could be modeled by component parameters. The area of study was the Southern Great Plains of the United States, which has a wealth of surface data available; this region has been the focus of intense experimental investigations in recent years and the benefit of this information should aid in the development of future land surface databases and modeling efforts. Estimates of both variables of interest, land surface temperature and vegetation density, are available from the National Oceanic and Atmospheric Administration (NOAA) Satellite Program.

2. Variables and experimental data

2.1. Land surface temperature and vegetation density

Remote sensing can be used to estimate land surface temperature (LST) and vegetation density measured by the normalized difference vegetation index (NDVI). Two techniques are commonly used to carry out the atmospheric correction in the estimation of LST from satellites. The split-window technique is an empirical approach that uses the relationship between emissivities from two satellite channels to estimate the land surface temperature [2]. Split-window equations are based on numerous satellite scenes, which are calibrated with surface measurements. This technique is efficient, but does not consider the specific atmospheric attenuation of the signal on a given day. The second technique is physically based and it accounts for this attenuation by incorporating atmospheric soundings to characterize the moisture and absorption properties of the air between the satellite sensor and the surface. The attenuation is then estimated by a radiative transfer model, such as MODTRAN (MODerate spectral resolution atmospheric TRANsmittance algorithm and computer model). The signal at the satellite is corrected to accurately show the temperature at the surface. For this study, Ou et al. [12] demonstrated that a split-window technique is roughly equivalent to a radiative transfer inversion technique for the Southern Great Plains, so this technique is used to generate LST maps for this study because of its ease of use.

Several indices have been developed over the years, which measure the 'greenness' of the land surface and could be used to approximate vegetation density or biomass. A popular index is the normalized difference vegetation index (NDVI) [14], which is defined as

$$\mathrm{NDVI} = \frac{\rho_{NIR} - \rho_{VIS}}{\rho_{NIR} + \rho_{VIS}} \quad (1)$$

Here ρ_{NIR} is the reflectance in the near infrared band (AVHRR infrared band 2 is 0.725–1.10 μm) and ρ_{VIS} is the reflectance in the visible band (AVHRR visible band 1 is 0.58–0.68 μm). A higher value indicates a more 'green' surface. Conversely, urbanization and waterways yield low and sometimes negative values.

2.2. Description of the experimental area

Satellite data were collected over the SGP ARM-CART area. As described by Stokes and Schwartz [15], this testbed is a 500 km by 350 km experimental area located in Kansas and Oklahoma in the Southern Great Plains of the United States. This study region is heavily monitored for a variety of radiation and soil variables to characterize the land atmosphere interface. Remote sensing measurements taken over the region permit analysis of many different surface variables, including land surface temperature, surface soil moisture, and vegetation. The region has gently rolling topography and a semi-humid climate. The dominant land cover types are: Cropland – 12%, Grassland – 38%, and Savanna – 27%. For the purposes of this study, only vegetated surfaces were analyzed. Less than 5% of the area was urbanized so that ignoring the non-vegetated pixels should not be significant. This data is available from the SGP data archive and is derived from the Oklahoma and Kansas Geological Surveys.

2.3. Description of satellite data

Thirty-five daytime scenes were retrieved for dates from within the period 1997 to 2000. AVHRR on board the NOAA-14 satellite retrieved the necessary information for deriving the surface variables in this study. AVHRR scans the entire earth twice a day with a resolution of 1 km^2 across five spectral bands. For this study, only the ~20:00 GMT (Greenwich Mean Time) overpass is considered, which for Oklahoma and Kansas (GMT–5:00) is approximately 15:00 local time. The surface temperature is derived from two thermal bands, channel 4 (10–11.5 μm) and channel 5 (11.5–12.5 μm). Ou et al. [12] demonstrated the ability of the split window method to estimate LST in this region with sufficient accuracy for the purpose of this analysis. Cloud masking techniques [11] were used to identify pixels containing clouds. These techniques included a combination of taking histograms of various channels and ratios of channels and masking pixels, which meet various criteria. These criteria are established from visual inspection per scene. These were removed for the analysis of both land surface temperature and vegetation density. Drying periods were defined as consecutive clear days of no precipitation.

A spatial analysis of this data set is presented in Cosh et al. [5]. The spatial distribution of vegetation density was anisotropic and statistically heterogeneous. However, as a drying period progressed the heterogeneity decreased and the characteristic length scale shortened. When the analysis was confined to similar land cover, the vegetation density was homogeneous with a characteristic length scale of up to 40 km. Conversely, land surface temperature was anisotropic and heterogeneous with no dependence on land cover. The spatial structure of land surface temperature was highly variable and changed dramatically from day to day.

3. Methods of analysis

The following four sets of analyses were used in this study to explore and define the significant parameters, which affect the surface variables under consideration, namely LST and NDVI.

3.1. Moment calculations

An investigation into the character of surface variability often begins with an analysis of the moments of the variable. These moments are capable of adequately describing a probability distribution function for a variable [16]. The mean, variance, and coefficient of variation were computed with the standard estimators. The coefficient of skewness was computed by

$$C_s = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^3, \tag{2}$$

and the coefficient of excess kurtosis as

$$C_k = \frac{1}{ns^4} \sum_{i=1}^{n} (x_i - \bar{x})^4 - 3. \tag{3}$$

Confidence intervals can be estimated for each moment estimate by means of bootstrapping [6]. Given a set of n observations, m sets of n samples are randomly selected with replacement. The desired statistic is then calculated for each m subset and sorted. For m equals 1000, the 25th and 975th (sorted) statistics provide an approximate 95% confidence interval for the population value.

3.2. Regression analysis

Regression analysis was used to study the relationship of the variability of surface temperature and vegetation density with three surface characteristics, namely vegetated land cover, time, and latitude. Latitude was included to determine if the scale of the study region influences the surface variables. In the Southern Great Plains, precipitation and soil moisture decrease with distance from the Gulf of Mexico in general. If a study region is significantly large enough to be affected by this spatial trend, this will be apparent in the regression analysis. Time measures the evolution of the surface variables from day to day as well as over a yearly cycle. In this case, time is quantified as Day of Year. Land cover was chosen as a categorical variable to study if the type of vegetated land cover affects the surface variables of vegetation density and land surface temperature. Land cover was represented by the three dominant types (cropland, grassland, and savanna), which account for approximately 75% of the total area. The variation in Y_{ijtL} represents the variable of interest (in the present study LST or NDVI) and was determined by

$$Y_{ijtL} = \phi + \gamma_i + (\alpha + \zeta_i) * t + (\beta + \psi_i) * L + \upsilon * t * L + e_{ijtL} \tag{4}$$

Specifically, Y_{ijtL} is the variable of interest for the jth site of the ith land cover type with the covariates of time, t, and latitude, L. The categorical variable for land cover type, γ_i, has levels $i = 1, 2, 3$, (cropland, grassland, and savanna). The covariates, time and latitude, are represented by t and L with coefficients of α and β. The terms, $\zeta_i{*}t$, $\psi_i{*}L$, and $\upsilon{*}t{*}L$, represent the interactions between the categorical variable and the covariates. The coefficients for the interaction terms, ζ_i and ψ_i, each have three levels. The residual is represented by e_{ijtL}, where j refers to the site. ϕ is a constant. The model is over-parameterized so two constraints are added: $\sum \gamma_i = 0$ (zero) and $\sum \beta_i = 0$ (zero) [10]. Precipitation amount (or antecedent precipitation index) is not included in the current study, but could be analyzed in the future.

This model can be fit to the remote sensing data retrieved for this study. The coefficient of determination, R^2, describes the variability that this model explains [10]. It is possible to represent the amount of variability due to a single factor by dividing the sum of squared errors for that factor by the total sum of squares. In this analysis, an incremental R^2 is calculated, because a sequential sum of squares is used. This method of analysis assumes a hierarchy of importance in predicting the variable in question. These R^2 values can be used to study the influence of a factor on the overall variability of a variable.

3.3. Correlation analysis

Correlation analysis is a useful tool to study variability pattern persistence over a drying period. Persistence is the tendency of a variable at a site to remain unchanged relative to the adjacent locations over time. The correlation coefficient, r_k, between two days, $t = i$ and $t = i + k$, in a data matrix, Y_{ijt}, is given by

$$r_k = \frac{\sum_i \sum_j \sum_t (Y_{ijt} - \overline{Y}_{\bullet\bullet t})(Y_{ij(t+k)} - \overline{Y}_{\bullet\bullet(t+k)})}{\sqrt{\left[\sum_i \sum_j \sum_t (Y_{ijt} - \overline{Y}_{\bullet\bullet t})^2\right] * \left[\sum_i \sum_j \sum_t (Y_{ij(t+k)} - \overline{Y}_{\bullet\bullet(t+k)})^2\right]}} \quad (5)$$

where $\overline{Y}_{\bullet\bullet t}$ and $\overline{Y}_{\bullet\bullet(t+k)}$ are the means of LST and NDVI for all soil types and all sites for times t and $t + k$. In the present study, (5) was applied to the residuals of a regression model with a term for land cover type. The residuals are defined as the difference between a regression model estimate and the observed values. This accounts for the dependence of the variable on land cover type, which can be substantial. Preliminary results indicated that land cover was the most practical to consider for each of the land surface characteristics in question.

3.4. Aggregation analysis

Lastly, aggregation analysis is a method of investigating how the statistical structure of a surface variable is affected by scaling [13]. In this analysis, the variance of an image is plotted versus the area of an aggregated pixel to create a variance cascade, as the resolution is linearly aggregated to more coarse resolutions. The results of this analysis can assist in making decisions regarding needed sampling density and remote sensing strategies. Modelers will often aggregate satellite data to a more coarse resolution, therefore it is necessary to determine how the aggregation technique affects the scaling behavior of a variable.

4. Results of analysis

4.1. Land surface temperature

4.1.1. Moment calculations

Four drying periods were identified within the AVHRR data available for the study. The study days during each drying period followed a precipitation event and were consecutive days without precipitation. In some instances only two of these will be presented, but the results were similar for all four. Fig. 1 contains the plots of the standard deviations (a), coefficients of skewness (b), and coefficients of excess kurtosis (c) for a drying period extending from 12-28-99 to 1-2-00 (December/January). Fig. 2 illustrates those same moments for a drying period from 7-4-00 to 7-8-00 (July) with 95% confidence intervals calculated by bootstrap. The coefficients of skewness and excess kurtosis vary considerably throughout a drying period. No pattern or trend is apparent in any of the calculated moments for either drying period.

The moment analysis was extended to an annual scale to determine if there was a seasonal pattern. Moments were calculated for 35 days across 1997–2000 and plotted against day of the year. Shown in Fig. 3a, the variability

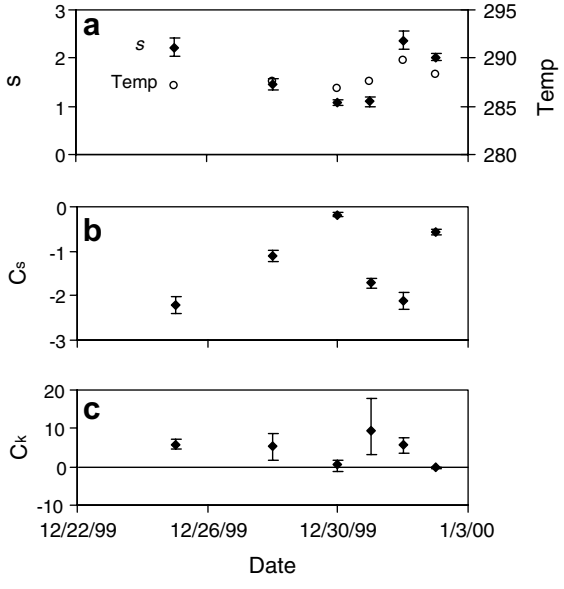

Fig. 1. Moments of LST (in Kelvin) for a December/January drying period with 95% confidence intervals. The symbol s refers to the standard deviation, C_s to the coefficient of skewness, and C_k to the coefficient of excess kurtosis.

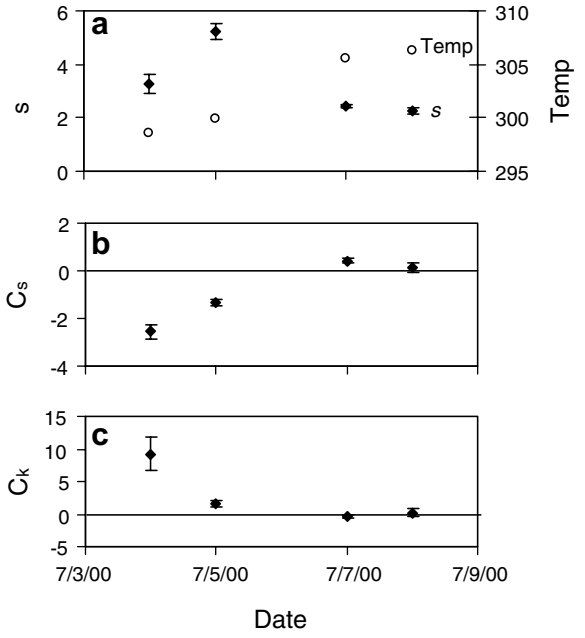

Fig. 2. Moments of LST (in Kelvin) for a July drying period with 95% confidence intervals.

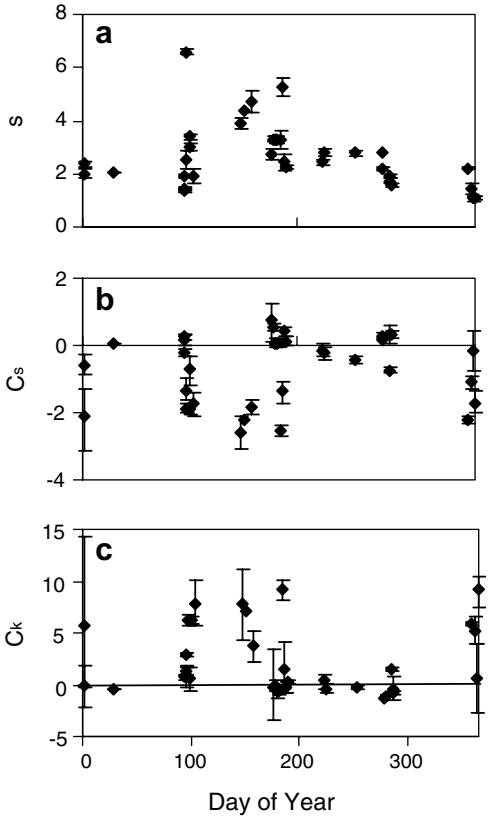

Fig. 3. (a) The standard deviations (s) for LST (in Kelvin) for the AVHRR data with 95% confidence intervals plotted versus day of the year, (b) the coefficients of skewness, C_s, for LST for the AVHRR data with 95% confidence intervals plotted versus day of the year and (c) the coefficients of excess kurtosis, C_k, for LST for the AVHRR data with 95% confidence intervals plotted versus day of the year.

of land surface temperature, as represented by the standard deviations, did not display a seasonal pattern or trend. Values of s ranged between 1.0 and 7.0 with most values between 1.0 and 4.0. These values vary significantly and further analysis would be necessary to determine whether this variation would affect modeling. The coefficients of skewness (Fig. 3b) and excess kurtosis (Fig. 3c) did not exhibit any annual pattern. The magnitudes of these moments varied over a significant range. The coefficient of skewness varied between −2.5 and 1.0, while the coefficient of excess kurtosis varies between −1.0 and 9.0. There was no clear pattern from day to day or season to season.

4.1.2. Regression analysis

Regression analysis was used to investigate which land surface characteristics play a role in the variability of LST during a drying period. Table 1 summarizes the R^2 values for each of the factors in a model. The model considers land cover first and then time. As will now be shown, these results reveal much about the possible differences between drying periods in different seasons. During most of these drying periods, on the whole those factors were responsible for between 10% and 57% of the variability of land surface temperature. Only for the July drying period did the R^2 value exceed 50%, with most of the increase being attributed to time, which is most likely anomalous. This will be addressed below. For all of the drying periods, the variability due to each factor or interaction was generally small. No single factor appeared to account for a major portion of the variability with any consistency. The strength of the relationship of LST variability with land use, time, and latitude changed markedly between the July period and the other three periods considered. Furthermore, though most factors and interactions are statistically significant in these models, the level of practical importance is generally minimal. Specifically, the R^2 values of the interactions between the three factors studied, land cover, time, and latitude contribute less than 4% of the variability of LST, as shown in Table 1. Other factors need to be examined to determine whether there are any other land surface variables that contribute consistently to LST variability or

Table 1
Incremental R^2 values for LST for four drying periods in the AVHRR data set

Factor	Dec./Jan.	April	May	July
Land cover, γ_i	0.8	0.5	2.2	4.8
Day, $\alpha*t$	1.2	3.2	0.2	49.6
Latitude, $\beta*L$	31.5	14.5	2.8	0.1
Land cover*Day, ζ_i*t	1.0	0.4	0.1	0.1
Land cover*latitude, ψ_i*L	0.0	0.3	4.8	2.4
Day*latitude, $\upsilon*t*L$	0.2	4.0	0.1	0.1
Total	34.7	22.9	10.2	57.0

R^2 describes the amount of variability of land surface temperature which can be explained by each factor. All factors were determined to be significant at 5% level with the exceptions underlined.

that land surface is not a valuable factor in LST variability parameterization.

A similar regression analysis can also be conducted to determine exactly how much variability can be accounted for by time. To accomplish this, a categorical time factor is substituted into the model to determine the effect of time beyond a linear relationship. This factor has the disadvantage of not being useful for future modeling, but can aid in the understanding of how much variability can be attributed to time and temporal effects. This new regression model is

$$Y_{ijtL} = \phi + \gamma_i + (\beta + \psi_i) * L + (\alpha + \zeta_i * t) + \upsilon * t * L + \kappa_t + \eta_{it} + e_{ijtL}. \quad (6)$$

Here κ_t is the new categorical time factor with 8 levels, and η_{it} is an interaction term between categorical time and land cover type. Additional constraints of $\sum_t \kappa_t = 0$ (zero) and $\sum_t \eta_{it} = 0$ (zero) are added. Table 2 provides a summary of the results of an analysis of variance (ANOVA) for each of the drying periods. The variables were reordered from the original sequence so that in the new regression analysis all non-time terms are introduced first. Linear time and its interactions are introduced second, and finally categorical time and its interaction with land cover is added. Categorical time in this instance accounts for all non-permanent land surface characteristics and could include such large scale temporally dynamic conditions as wind speed, solar radiance, and humidity. Including each of these temporal dynamic parameters can be computationally exhaustive and may yield invalid results, but they can be accounted for by the categorical time variable. The R^2 values are calculated for each step of this model so that a clear understanding of the effects of time can be discerned. The amount of variability that could be explained by the categorical variable of time varied between the periods. Its contribution to the overall variable was as small as 2.7% for July and 2.8% for December/January and as large as 45.9% for April and 31.6% for May. Time played a significant role in the variability of LST when considered as both a categorical and linear variable. Time was the largest contributing factor for three of the four periods studied. The linear model appeared to work only during the July drying period and further study of other drying periods would be necessary to determine whether this was truly anomalous.

Table 2
Incremental R^2 values for LST for four drying periods with an additional time factor, which is treated as a categorical variable, *daycat*

Factor	December/January	April	May	July
Model w/o day terms	32.3	15.2	0.8	7.3
Day, $(\alpha + \xi_i)*t$	2.4	0.8	<u>0.4</u>	49.7
Daycat, κ_t	2.8	45.9	31.6	2.7
Daycat*land, ξ_{it}	1.5	<u>0.2</u>	<u>0.1</u>	1.0
Total	39.1	69.0	41.9	60.7

The table is based on a sequential sum of squares and all factors are significant at the 5% level, with exceptions underlined.

4.1.3. Correlation analysis

Correlation analysis was conducted on the dates included in this study and two drying periods are examined in more detail below. Two hundred sites (pixels) were randomly sampled and regression analysis conducted, while maintaining the relative proportions of land cover types in the sample. The residuals from this regression were used to study the correlation structure from day to day. This was done to ensure that the persistence measured is not a result of underlying patterns of land surface temperature due to any of the patterns of land cover, such as the distribution of forest and pastures or agricultural fields. Tables 3 and 4 contain the correlation coefficients between days for the December/January and July drying periods. The results demonstrate that there is no clear spatial pattern of land surface temperature during a drying period. Confidence intervals for the values of r_k are approximately ± 0.15.

It is possible to calculate the amount of variability attributable to persistence by revisiting the regression model and introducing a new factor. If a categorical variable is added to the model (Eq. (6)) to account for site-to-site differences in LST, the R^2 for that factor would give an additional (beside correlation analysis) estimate of the amount of pattern persistence. This new model is defined as

$$Y_{ijtL} = \phi + \gamma_i + (\alpha + \zeta_i) * t + (\beta + \psi_i) * L + \upsilon * t * L + \delta_{ij} + e_{ijtL} \quad (7)$$

where δ_{ij} is the categorical variable for the jth site with the ith land cover. The constraint, $\Sigma \delta_{ij} = 0$ (zero), is added to the analysis also. Table 5 is a summary of the results of a regression of the four drying periods. The amount of variability attributable to this site factor is between 10% and 20%, which is not a large amount when considering the variability of LST from day to day. In fact, for the December/January and April drying periods, the p-values reveal that the site factor is not significant at the 5% level. It can be concluded that there is little persistence of LST during a drying period.

This obvious lack of a consistent structure during a drying period would lead one to conclude that there is no persistent structure in the LST pattern throughout the year. It appears that the structure of LST from day to day is affected by factors which vary on a daily scale and at a smaller scale than the study region, the most obvious being weather conditions for a region of this scale, 350 km by

Table 3
Correlation coefficients among residuals of the regression analysis for LST for a December/January drying period

	12/28/99	12/30/99	12/31/99	1/1/00	1/2/00
12/28/99	1.000	0.222	0.233	0.063	0.113
12/30/99		1.000	0.359	−0.034	0.204
12/31/99			1.000	−0.068	0.191
1/1/00				1.000	−0.210
1/2/00					1.000

A 95% confidence interval for each coefficient is approximately ± 0.15.

Table 4
Correlation coefficients among residuals of a regression analysis for LST for a July drying period

	7/4/00	7/5/00	7/7/00	7/8/00
7/4/00	1.000	−0.004	0.248	0.191
7/5/00		1.000	0.452	0.427
7/7/00			1.000	0.612
7/8/00				1.000

A 95% confidence interval for each coefficient is approximately ±0.15.

Table 5
Summary table of the incremental R^2 values for the regression model of LST containing the site factors (Eq. (5))

Factor	December/January	April	May	July
Model w/o sites	34.7	22.9	10.2	57.0
Sites, δ_{ij}	13.7	10.8	22.3	17.4
Model	39.1	69.0	41.9	60.7

All factors are significant at a 5% level with exceptions underlined.

500 km with a 1 km resolution. This may not be true for imagery with a smaller resolution or of smaller scale.

4.1.4. Aggregation analysis

The satellite data used to estimate LST are based on radiance values that are related to LST by the Stefan–Boltzmann equation, where temperature is to the fourth power. Aggregation analysis was used to determine if the non-linear calculation of LST affected the variance estimation significantly. Two variance cascades were calculated; one based on a derived LST image, the other based on the original radiance scene used to calculate LST. The radiance scene is linearly aggregated to more coarse resolutions and then after taking the one-fourth power, the LST variance is calculated. This analysis demonstrated that the non-linear relationship between the variable, land surface temperature, and the observed variable, radiance, does not affect the variance cascade. As shown in Fig. 4, the variance cascade of the temperature maps derived from the temperature aggregation was virtually the same as that derived from the radiance (temperature brightness) aggregation.

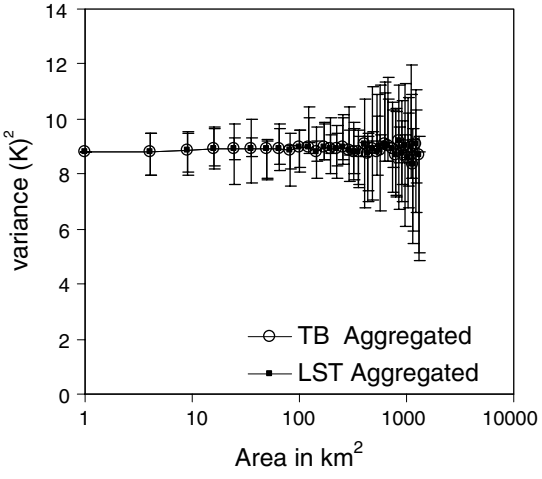

Fig. 4. Variance cascade for LST on 1/28/97 with 95% confidence intervals. The variance cascades are coincident and the confidence intervals overlap.

4.2. Vegetation density

4.2.1. Moment calculations

Four sets of drying period data were available for an analysis of vegetation density. Those four drying periods were analyzed at the scale of the SGP ARM-CART area. The moments of NDVI are presented in Fig. 5 for the December/January drying period and in Fig. 6 for the July drying period. One thousand sites (pixels) were randomly sampled and moment estimates (with bootstrap confidence intervals) were calculated for each day. The coefficients of variation for the July drying period vary between 0.18 and 0.24, which may be represented as a constant without introducing significant error. It can be seen that NDVI presented a much more stable variable throughout each of the four drying periods so that it can be assumed that estimates for a given day are reasonable estimates for that entire drying period. This assumption can also be justified from observations of vegetation and its slow response to drying conditions. Usually, it will take several days of drying before the vegetation begins to be affected.

At the annual scale, the higher moments of NDVI exhibited a definite pattern. This pattern showed that the December/January period was more variable than the other seasons. Fig. 7a shows the evolution of the standard deviation of NDVI in the region. The increase in variability

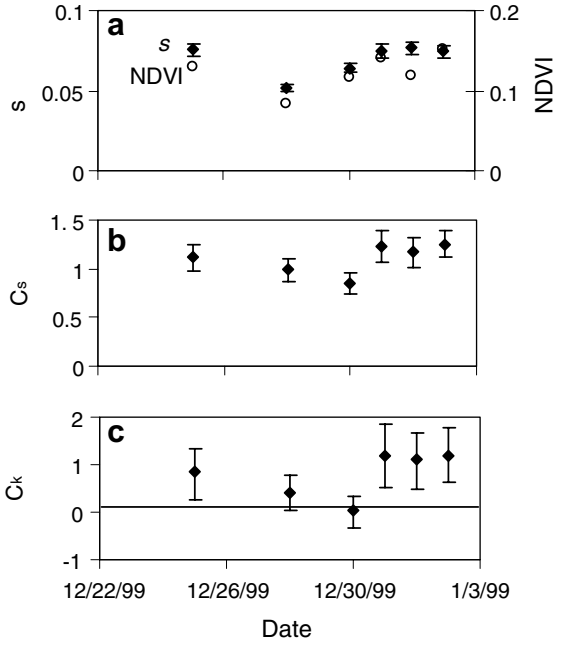

Fig. 5. Dimensionless moments of NDVI for a December/January drying period beginning 12/28/99 to 1/2/00 with 95% confidence intervals.

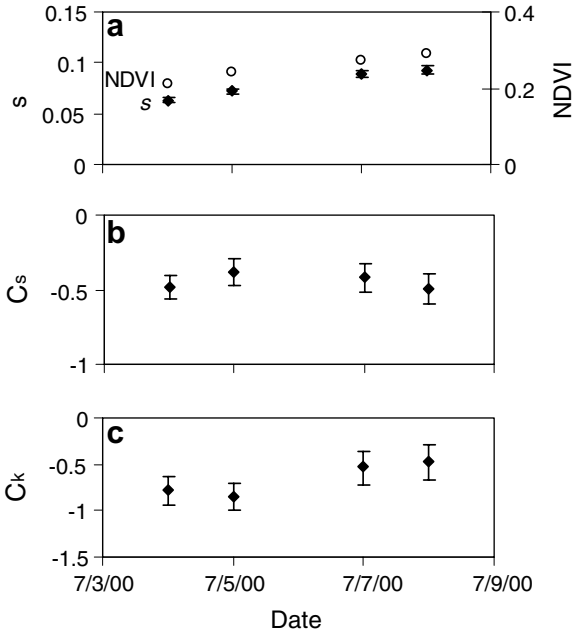

Fig. 6. Dimensionless moments of NDVI for a July drying period beginning 7/4/00 to 7/8/00 with 95% confidence intervals.

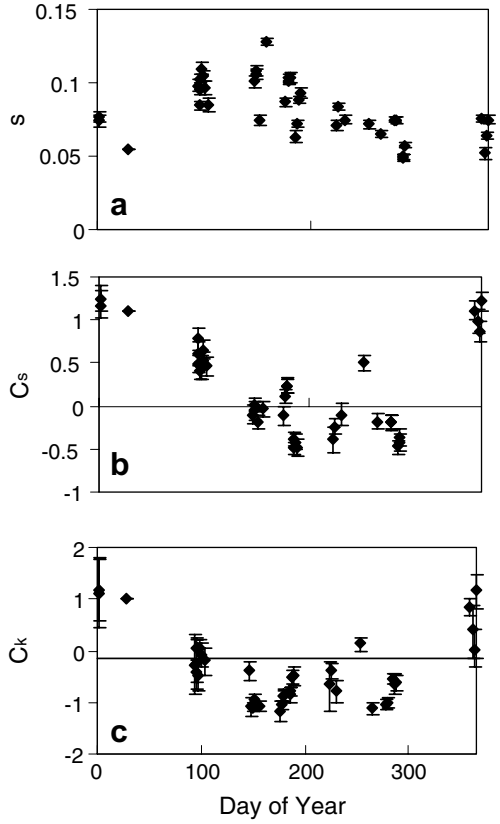

Fig. 7. (a) The standard deviations for NDVI for the AVHRR data with 95% confidence intervals plotted versus day of the year, (b) the coefficients of skewness for NDVI for the AVHRR data with 95% confidence intervals plotted versus day of the year and (c) the coefficients of excess kurtosis for NDVI for the AVHRR data with 95% confidence intervals plotted versus day of the year.

during the winter can be traced to the diverse nature of land cover and land use in the region. Savanna was dormant in the winter and green during the growing season, which when combined with the annual cycle of the various crops contributed to an increase in variability during the winter. Fig. 7b shows the coefficient of skewness went through a similar cycle.

During the winter months, the skewness was approximately equal to 1.0, while, during the growing months, it was closer to 0.0. The winter wheat and forest/wooded regions still have a high NDVI, while most of the fallow and bare soil will have low NDVI, resulting in a very different pattern from the growing season. During these growing/harvesting months, the greenness of the surface remained relatively constant because the cropland, grassland, and savanna were all relatively green.

Finally, the coefficient of excess kurtosis followed a pattern of positive values for winter, and negative values for the growing seasons, as shown in Fig. 7c. A thick-tailed distribution has a positive excess kurtosis, while negative values result from a thin-tailed distribution. This seasonality can be explained by the preponderance of green vegetation and therefore less variability during the growing seasons. The confidence intervals for the winter C_k are larger than those for the summer dates. This increase in the confidence intervals can be attributed to the wider distribution of vegetation 'greenness' during the winter months. In the summer, most of the land surface is green. However in the winter, much of the land is in fallow, but there are still sizeable areas of green vegetation in the form of evergreen tree stands and winter wheat.

A simple analysis of means can reveal how the vegetation density varies over the year. Table 6 contains the means of NDVI for each of the four drying periods in the study. Cropland goes through an evolution during the year as the crops are planted, grown, and harvested. Grassland and savanna go through a similar evolution, but the range of vegetation is more extreme. For example, cropland has a similar NDVI for December/January as for July (post-harvest), while savanna goes from 0.09 for December/January to 0.33 for May, height of the growing season.

4.2.2. Regression analysis

Regression analysis was conducted to determine how the model (Eq. (4)) predicts the vegetation density. The primary land surface characteristic used in the analysis was

Table 6
Average values of NDVI for each land cover type

	Cropland	Grassland	Savanna
December/January	0.17 ± 0.01	0.11 ± 0.01	0.09 ± 0.01
April	0.32 ± 0.02	0.26 ± 0.01	0.24 ± 0.01
May	0.24 ± 0.02	0.28 ± 0.01	0.33 ± 0.01
July	0.16 ± 0.01	0.23 ± 0.01	0.29 ± 0.01

95% confidence intervals derived from standard error are also listed.

Table 7
Incremental R^2 values for NDVI for the four drying periods in this study

Factor	December/January	April	May	July
Land cover, γ_i	16.2	4.2	6.0	23.4
Day, $\alpha*t$	6.4	1.2	9.0	9.3
Latitude, $\beta*L$	2.7	12.2	1.8	0.0
Land cover*day, ζ_i*t	0.3	0.4	0.1	0.6
Land cover*latitude, ψ_i*L	2.2	1.4	3.4	2.9
Day*latitude, $\upsilon*t*L$	0.0	0.0	0.5	0.0
Total	27.9	19.4	20.9	36.3

Those factors that have a p-value less than 0.05 are underlined.

land cover type with two covariates, time since precipitation and latitude, also included. Table 7 contains a summary of the incremental R^2 values for each of the factors and interactions included in the model. None of the models achieved a very high R^2 value. All were smaller than 0.40. It should be noted that NDVI is obviously related to land cover, but the ANOVA tables reveal that this relationship is weak when considering the variability of NDVI throughout the entire region.

Table 8 reports an analysis of variance was conducted that included a categorical factor for time as well as its interaction with land cover. The categorical variable for time displays small R^2 values with the largest being 12.9% for the May drying period. The categorical interaction term for time and land cover also show very small R^2 values and for each of the drying periods the factor was not significant at a 5% level. In all cases, time did not appear to be a significant factor in contributing to vegetation density variability during a drying period.

4.2.3. Correlation analysis

The serial correlation matrices for the residuals of the regression analysis for NDVI for two drying periods are contained in Tables 9 and 10. There is a much stronger correlation from day to day for NDVI than LST. However, the persistence of this pattern degrades over time, as illustrated by how the values away from diagonal decreases as time separation increases. Correlation coefficients tend to be greater than 0.40, which is indicative of a moderate to strong spatial structure.

Table 8
Incremental R^2 values for simple linear regression of NDVI for four drying periods with an additional time factor that is treated as a categorical variable, *daycat*

Factor	December/January	April	May	July
Model w/o day terms	21.2	22.8	11.3	26.3
Day terms	6.7	2.0	9.6	9.9
Daycat, κ_t	4.2	0.8	12.9	1.2
Daycat*land, ζ_{it}	0.1	0.1	0.9	0.2
Model	32.2	25.7	34.7	37.6

The results are based on a sequential sum of squares. Those factors which have a p-value less than 0.05 are underlined.

Table 9
Correlation coefficients among the residuals from the regression analysis for the December/January drying period for NDVI

	12/28/99	12/30/99	12/31/99	1/1/00	1/2/00
12/28/99	1.000	0.697	0.454	0.395	0.389
12/30/99		1.000	0.569	0.459	0.496
12/31/99			1.000	0.695	0.738
1/1/00				1.000	0.690
1/2/00					1.000

A 95% confidence interval for each coefficient is approximately ±0.10.

Table 10
Correlation coefficients among the residuals of the regression analysis for the July drying period for NDVI

	7/4/00	7/5/00	7/6/00	7/7/00
7/4/00	1.000	0.636	0.385	0.491
7/5/00		1.000	0.521	0.473
7/6/00			1.000	0.465
7/7/00				1.000

A 95% confidence interval for each coefficient is approximately ±0.10.

As before, regression analysis can also be used to study the persistence in the NDVI data. In the manner described for LST, a categorical variable for each site was added to the regression model. Table 11 contains a summary table of all the drying periods. The R^2 values for the site factor are between 37% and 55%, which would indicate that persistence plays a significant role in NDVI variability. When considering the influence of variables such as time since precipitation, land cover, and latitude, on NDVI variability, it would appear that persistence for a specific site is more important to the overall characterization of variability than any of these three factors.

The final step in the correlation analysis of variability is the detection of pattern persistence at the seasonal scale. NDVI demonstrates strong pattern persistence within both the dormant ($r_k = 0.50+/-0.03$) and growing seasons ($r_k = 0.47+/-0.03$), but those patterns are uncorrelated ($r_k = 0.26+/-0.01$) with each other.

4.2.4. Aggregation analysis

An aggregation analysis, which was developed in a manner similar to that for LST, is presented in Fig. 8. NDVI can be aggregated in two different ways. The first is based on taking an NDVI image and aggregating the pixels

Table 11
Summary table of incremental R^2 values for the regression model containing the site factors

Factor	December/January	April	May	July
Model w/o sites	27.9	24.8	20.9	36.2
Sites, δ_{ij}	47.8	55.0	43.5	37.3
Model	75.7	79.8	64.4	73.5

Factors with a p-value less than 0.05 are underlined.

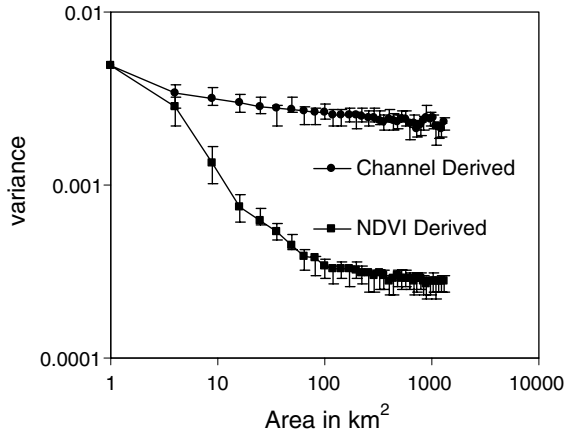

Fig. 8. Variance cascade with bootstrap confidence intervals for NDVI on 1-28-97. Higher plot is derived from original channel readings while lower plot is based on original NDVI.

together to a lower resolution. The second method takes the original AVHRR channels and aggregates them prior to computing NDVI. These results illustrate how the non-linearity of the NDVI calculation affects the variance cascade. It is intuitive that the appropriate method of aggregation is to use the original satellite channel readings in any map generation. In Fig. 8, it is shown that, if original channel readings are not used, the variance is underestimated in the case of NDVI. This will also hold true for variables derived from NDVI as well as other non-linear formulations. As scaling issues becoming more important with the variety of satellite sensors available at different resolutions, it is important to recognize that derived data products such as NDVI, may not be scalable in their calculated state.

5. Conclusions

This study documents how surface variability evolves with time. Specifically, it examines the key aspects of land surface temperature and vegetation density as represented by NDVI at drought-episodic and seasonal time scales. It is apparent from the moment analysis that the distribution of LST is highly variable with respect to both time scales. The range of variability of LST is small when considering temperature on the grid scale. However, higher order moments display erratic behavior, which could make parameterization of the distribution difficult.

Using several factors, such as vegetated land cover, latitude, and time since precipitation, a regression model for LST was constructed. The fraction of variability, R^2, explained by different variables was almost always relatively small, and changed in an inconsistent manner from period to period. This study shows that with regards to land cover and latitude, the region is indiscriminate and statistically unpredictable. The dependence of LST on time since precipitation varied dramatically between the drying periods. Time as a linear variable accounted for nearly 50% of the variability during the July drying period, but for the other periods it was practically insignificant. Similar results were obtained when time was treated as a categorical variable. Several drying periods can have a greatly improved R^2 if time is included as a categorical variable, while others showed only small increases of no practical importance. This indicates that the variability of LST showed no pattern with time and is no doubt the result of factors that operate on time scales shorter than a day. Residual correlation analysis demonstrates that LST did not maintain a consistent spatial pattern from one day to the next and less than 20% of the variability could be attributed to temporal persistence. The results of this study show that the temporal change of LST subgrid scale variability is more complex than generally assumed in the SGP-ARM-CART region at a resolution of 1 km. Because of the dynamic nature of the variability, it would be difficult if not impossible to correctly parameterize it in a model.

In contrast, vegetation density has a persistent pattern at the two time scales of interest. NDVI showed no differences in variability over a drying period of five days or less. This is reasonable considering the resilience of most plants to such drying periods. The moment analysis indicated that the coefficients of skewness and excess kurtosis were positive during the growing season and negative during the dormant season. Savanna, land covered by small trees and low shrubs, covers a significant portion of the study area and exhibits a strong seasonal pattern. Table 6 reveals that during the summer months, it is very green with an average NDVI higher than cropland and grassland. However, during the winter months, savanna goes dormant and brown, while the cropland maintains some greenness. Cropland undergoes a different type of evolution during the year. During the winter months, there is often a cover of winter wheat making the land more green than surrounding dormant grassland and savanna. During the summer months, cropland varies in greenness because of the growing/harvesting cycle. The proportions of these land cover types combined with their annual patterns may result in the annual patterns in the moments.

Another observation made by this study is that NDVI variability estimation is dependent on the source data. Variability estimates were affected by the method of pixel aggregation. It was demonstrated that NDVI maintains a spatial structure during at least the first stage of drying in the Southern Great Plains. Correlation coefficients for the residuals (observed minus average for a specific day) at the same location but separated in time were usually greater than 0.35 for up to five days. It was determined that the temporal persistence can account for between 37% and 55% of the variability of vegetation density. Furthermore, the year could be divided into periods which exhibited different patterns. Within the winter months considered, there was a moderate to strong positive temporal correlation. Similarly, within the July data from the growing season there was a strong pattern persistence. Vegetation density

variability does change over the course of the year; therefore, it is important to consider this change when working on a large scale, but this study shows that for normal drying periods, an assumption of constant vegetation greenness is not inappropriate.

In summary, the methodology presented in this paper demonstrates a simple and quick way to quantify the relationships between a surface parameter and other variables that have an impact on its variability. This should enable the prioritization of land surface characteristics for the purpose of understanding the subgrid scale variability.

Acknowledgements

This research was conducted with the cooperation of the US Department of Energy as part of the Atmospheric Radiation Measurement Program. In addition, this work has been supported by grants ATM-9708622 from the National Science Foundation and NAG8-1518 from the National Aeronautics and Space Administration.

References

[1] Avissar R, Pielke RA. A parameterization of heterogeneous land surfaces for atmospheric numerical models and its impact on regional meteorology. Mon Weather Rev 1989;117(10):2113–36.
[2] Becker F, Li Z. Toward a local split-window method over land surface. Int J Remote Sens 1990;11(3):269–393.
[3] Chen D, Brutsaert W. Satellite-sensed distribution and spatial patterns of vegetation parameters over a tallgrass prairie. J Atmos Sci 1998;55(7):1225–38.
[4] Chen D, Engman ET, Brutsaert W. Spatial distribution and pattern persistence of surface soil moisture and temperature over prairie from remote sensing. Remote Sens Environ 1997;61(3):347–60.
[5] Cosh MH, Stedinger J, Brutsaert W. Time changes in spatial structure of surface variability in the Southern Great Plains. Adv Water Resour 2003;26(4):407–15.
[6] Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman Hall; 1993.
[7] Entekhabi D, Eagleson PS. Land surface hydrology parameterization for atmospheric general circulation models including subgrid scale spatial variability. J Climate 1989;2(8):816–31.
[8] Li B, Avissar R. The impact of spatial variability of land-surface characteristics on land-surface heat fluxes. J Climate 1994;7(4):527–37.
[9] Lynn BH, Rind D, Avissar R. The importance of mesoscale circulations generated by subgrid-scale landscape heterogeneities in general circulation models. J Climate 1995;8(2):191–205.
[10] Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. Applied linear statistical models. Chicago, Illinois: Irwin Press; 1996.
[11] Ou SC, Liou KN, Baum BA. Detection of multilayer cirrus cloud systems using AVHRR data: verification based on FIRE II IFO composite measurements. J Appl Met 1996;35(2):178–91.
[12] Ou SC, Liou KN, Chen Y, Cosh MH, Brutsaert W. Satellite remote sensing of land surface temperatures: application of the atmospheric correction method and split-window technique to data of ARM-SGP site. Int J Remote Sens 2002;23(24):5177–92.
[13] Rodriguez-Iturbe I, Vogel GK, Rigon R, Entekhabi D, Castelli F, Rinaldo A. On the spatial organization of soil moisture fields. Geophys Res Lett 1995;22(20):2757–60.
[14] Schowengerdt RA. Remote sensing: models and methods for image processing. San Diego, CA: Academic Press; 1997.
[15] Stokes GM, Schwartz SE. The atmospheric radiation measurement (ARM) program: programmatic background and design of the cloud and radiation test bed. Bull Amer Met Soc 1994;75(7):1201–21.
[16] Yaglom AM. An introduction to the theory of stationary random functions. Englewood Cliffs, New Jersey: Prentice-Hall; 1962.