MAE 5870 - Aula 3

Estimação

Seja uma série temporal: x_1, x_2, ..., x_n

1. Média:

if a time series is stationary, the mean function (1.21) $\mu_t = \mu$ is constant so that we can estimate it by the sample mean,

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t. \tag{1.32}$$

The standard error of the estimate is the square root of $\text{var}(\bar{x})$, which can be computed using first principles (recall footnote 3 on page 20), and is given by

$$\begin{aligned}\text{var}(\bar{x}) &= \text{var}\left(\frac{1}{n}\sum_{t=1}^{n} x_t\right) = \frac{1}{n^2}\text{cov}\left(\sum_{t=1}^{n} x_t, \sum_{s=1}^{n} x_s\right) \\ &= \frac{1}{n^2}\Big(n\gamma_x(0) + (n-1)\gamma_x(1) + (n-2)\gamma_x(2) + \cdots + \gamma_x(n-1) \\ &\qquad + (n-1)\gamma_x(-1) + (n-2)\gamma_x(-2) + \cdots + \gamma_x(1-n)\Big) \\ &= \frac{1}{n}\sum_{h=-n}^{n}\Big(1 - \frac{|h|}{n}\Big)\gamma_x(h). \end{aligned} \tag{1.33}$$

If the process is white noise, (1.33) reduces to the familiar σ_x^2/n recalling that $\gamma_x(0) = \sigma_x^2$. Note that, in the case of dependence, the standard error of $\bar{x}$ may be smaller or larger than the white noise case depending on the nature of the correlation structure

Definition 1.14 *The* **sample autocovariance function** *is defined as*

$$\widehat{\gamma}(h) = n^{-1}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x}), \tag{1.34}$$

with $\widehat{\gamma}(-h) = \widehat{\gamma}(h)$ *for* $h = 0, 1, \ldots, n-1$.

Definition 1.15 *The* **sample autocorrelation function** *is defined, analogously to (1.23), as*

$$\widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)}. \tag{1.35}$$

Property 1.1 Large-Sample Distribution of the ACF

Under general conditions,[5] *if x_t is white noise, then for n large, the sample ACF, $\widehat{\rho}_x(h)$, for $h = 1, 2, \ldots, H$, where H is fixed but arbitrary, is approximately normally distributed with zero mean and standard deviation given by*

$$\sigma_{\hat{\rho}_x(h)} = \frac{1}{\sqrt{n}}. \tag{1.36}$$

Based on the previous result, we obtain a rough method of assessing whether peaks in $\widehat{\rho}(h)$ are significant by determining whether the observed peak is outside the interval $\pm 2/\sqrt{n}$ (or plus/minus two standard errors); for a white noise sequence, approximately 95% of the sample ACFs should be within these limits. The applications of this property develop because many statistical modeling procedures depend on reducing a time series to a white noise series using various kinds of transformations. After such a procedure is applied, the plotted ACFs of the residuals should then lie roughly within the limits given above.

Definition 1.16 *The estimators for the cross-covariance function, $\gamma_{xy}(h)$, as given in (1.26) and the cross-correlation, $\rho_{xy}(h)$, in (1.27) are given, respectively, by the* **sample cross-covariance function**

$$\widehat{\gamma}_{xy}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(y_t - \bar{y}), \tag{1.37}$$

where $\widehat{\gamma}_{xy}(-h) = \widehat{\gamma}_{yx}(h)$ determines the function for negative lags, and the **sample cross-correlation function**

$$\widehat{\rho}_{xy}(h) = \frac{\widehat{\gamma}_{xy}(h)}{\sqrt{\widehat{\gamma}_x(0)\widehat{\gamma}_y(0)}}. \tag{1.38}$$

Property 1.2 Large-Sample Distribution of Cross-Correlation Under Independence

The large sample distribution of $\widehat{\rho}_{xy}(h)$ is normal with mean zero and

$$\sigma_{\hat{\rho}_{xy}} = \frac{1}{\sqrt{n}} \tag{1.39}$$

if at least one of the processes is independent white noise

Exemplos:

Example 1.23 A Simulated Time Series

To give an example of the procedure for calculating numerically the autocovariance and cross-covariance functions, consider a contrived set of data

generated by tossing a fair coin, letting $x_t = 1$ when a head is obtained and $x_t = -1$ when a tail is obtained. Construct y_t as

$$y_t = 5 + x_t - .7x_{t-1}. \tag{1.40}$$

Table 1.1 shows sample realizations of the appropriate processes with $x_0 = -1$ and $n = 10$.

Table 1.1. Sample Realization of the Contrived Series y_t

t	1	2	3	4	5	6	7	8	9	10
Coin	H	H	T	H	T	T	T	H	T	H
x_t	1	1	−1	1	−1	−1	−1	1	−1	1
y_t	6.7	5.3	3.3	6.7	3.3	4.7	4.7	6.7	3.3	6.7
$y_t - \bar{y}$	1.56	.16	−1.84	1.56	−1.84	−.44	−.44	1.56	−1.84	1.56

The sample autocorrelation for the series y_t can be calculated using (1.34) and (1.35) for $h = 0, 1, 2, \ldots$. It is not necessary to calculate for negative values because of the symmetry. For example, for $h = 3$, the autocorrelation becomes the ratio of

$$\begin{aligned}\widehat{\gamma}_y(3) &= \tfrac{1}{10}\sum_{t=1}^{7}(y_{t+3} - \bar{y})(y_t - \bar{y}) \\ &= \tfrac{1}{10}\Big[(1.56)(1.56) + (-1.84)(.16) + (-.44)(-1.84) + (-.44)(1.56) \\ &\quad + (1.56)(-1.84) + (-1.84)(-.44) + (1.56)(-.44)\Big] = -.048\end{aligned}$$

to

$$\widehat{\gamma}_y(0) = \tfrac{1}{10}[(1.56)^2 + (.16)^2 + \cdots + (1.56)^2] = 2.030$$

so that

$$\widehat{\rho}_y(3) = \frac{-.048}{2.030} = -.024.$$

The theoretical ACF can be obtained from the model (1.40) using the fact that the mean of x_t is zero and the variance of x_t is one. It can be shown that

$$\rho_y(1) = \frac{-.7}{1 + .7^2} = -.47$$

and $\rho_y(h) = 0$ for $|h| > 1$ (Problem 1.24). Table 1.2 compares the theoretical ACF with sample ACFs for a realization where $n = 10$ and another realization where $n = 100$; we note the increased variability in the smaller size sample.

Table 1.2. Theoretical and Sample ACFs for $n = 10$ and $n = 100$

h	$\rho_y(h)$	$n = 10$ $\widehat{\rho}_y(h)$	$n = 100$ $\widehat{\rho}_y(h)$
0	1.00	1.00	1.00
±1	−.47	−.55	−.45
±2	.00	.17	−.12
±3	.00	−.02	.14
±4	.00	.15	.01
±5	.00	−.46	−.01

ACF of a Speech Signal

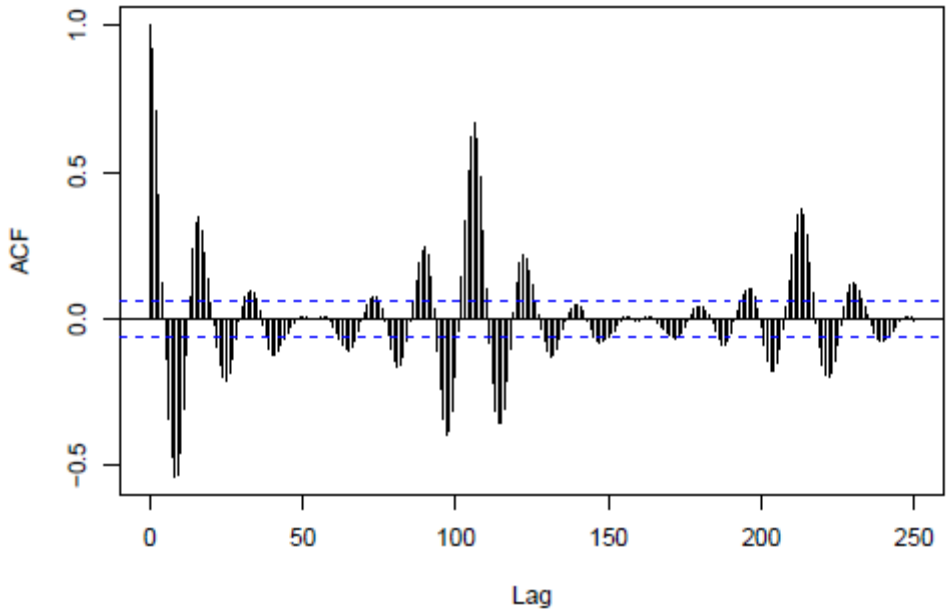

Fig. 1.13. ACF of the speech series.

SOI and Recruitment Correlation Analysis

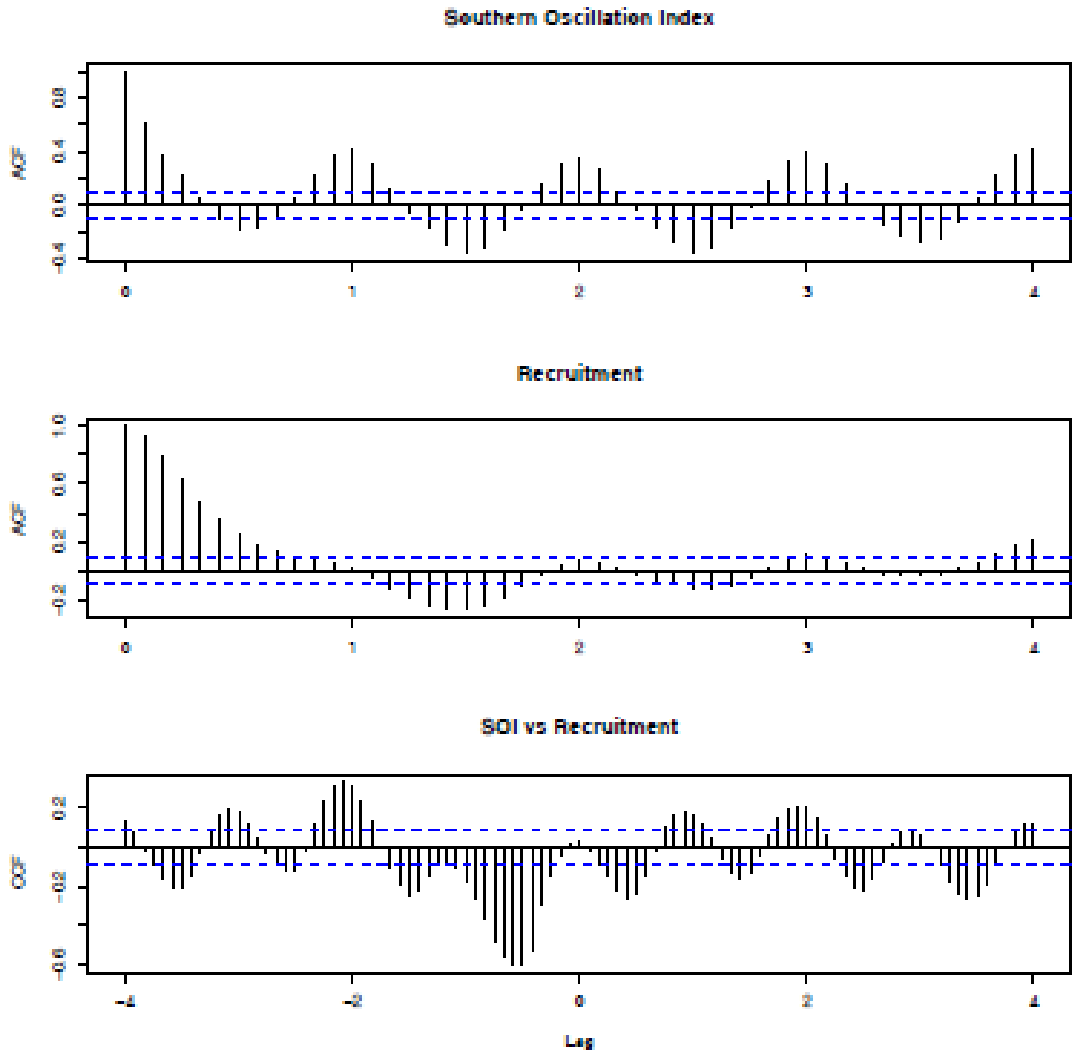

Fig. 1.14. Sample ACFs of the SOI series (top) and of the Recruitment series (middle), and the sample CCF of the two series (bottom); negative lags indicate SOI leads Recruitment. The lag axes are in terms of seasons (12 months).

1.7 Vector-Valued and Multidimensional Series

We frequently encounter situations in which the relationships between a number of jointly measured time series are of interest. For example, in the previous sections, we considered discovering the relationships between the SOI and Recruitment series. Hence, it will be useful to consider the notion of a vector time series $\boldsymbol{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tp})'$, which contains as its components p univariate time series. We denote the $p \times 1$ column vector of the observed series as $\boldsymbol{x}_t$. The row vector $\boldsymbol{x}_t'$ is its transpose. For the stationary case, the $p \times 1$ mean vector

$$\boldsymbol{\mu} = E(\boldsymbol{x}_t) \tag{1.41}$$

of the form $\boldsymbol{\mu} = (\mu_{t1}, \mu_{t2}, \ldots, \mu_{tp})'$ and the $p \times p$ autocovariance matrix

$$\Gamma(h) = E[(\boldsymbol{x}_{t+h} - \boldsymbol{\mu})(\boldsymbol{x}_t - \boldsymbol{\mu})'] \tag{1.42}$$

can be defined, where the elements of the matrix $\Gamma(h)$ are the cross-covariance functions

$$\gamma_{ij}(h) = E[(x_{t+h,i} - \mu_i)(x_{tj} - \mu_j)] \tag{1.43}$$

for $i, j = 1, \ldots, p$. Because $\gamma_{ij}(h) = \gamma_{ji}(-h)$, it follows that

$$\Gamma(-h) = \Gamma'(h). \tag{1.44}$$

Now, the sample autocovariance matrix of the vector series $\boldsymbol{x}_t$ is the $p \times p$ matrix of sample cross-covariances, defined as

$$\widehat{\Gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (\boldsymbol{x}_{t+h} - \bar{\boldsymbol{x}})(\boldsymbol{x}_t - \bar{\boldsymbol{x}})', \tag{1.45}$$

where

$$\bar{\boldsymbol{x}} = n^{-1} \sum_{t=1}^{n} \boldsymbol{x}_t \tag{1.46}$$

denotes the $p \times 1$ sample mean vector. The symmetry property of the theoretical autocovariance (1.44) extends to the sample autocovariance (1.45), which is defined for negative values by taking

$$\widehat{\Gamma}(-h) = \widehat{\Gamma}(h)'. \tag{1.47}$$

In many applied problems, an observed series may be indexed by more than time alone. For example, the position in space of an experimental unit might be described by two coordinates, say, s_1 and s_2. We may proceed in these cases by defining a multidimensional process $x_{\boldsymbol{s}}$ as a function of the $r \times 1$ vector $\boldsymbol{s} = (s_1, s_2, \ldots, s_r)'$, where s_i denotes the coordinate of the ith index.

Example 1.26 Soil Surface Temperatures

As an example, the two-dimensional ($r = 2$) temperature series x_{s_1,s_2} in Figure 1.15 is indexed by a row number s_1 and a column number s_2 that represent positions on a 64×36 spatial grid set out on an agricultural field. The value of the temperature measured at row s_1 and column s_2, is denoted by $x_{\boldsymbol{s}} = x_{s1,s2}$. We can note from the two-dimensional plot that a distinct change occurs in the character of the two-dimensional surface starting at about row 40, where the oscillations along the row axis become fairly stable and periodic. For example, averaging over the 36 columns, we may compute an average value for each s_1 as in Figure 1.16. It is clear that the noise present in the first part of the two-dimensional series is nicely averaged out, and we see a clear and consistent temperature signal.

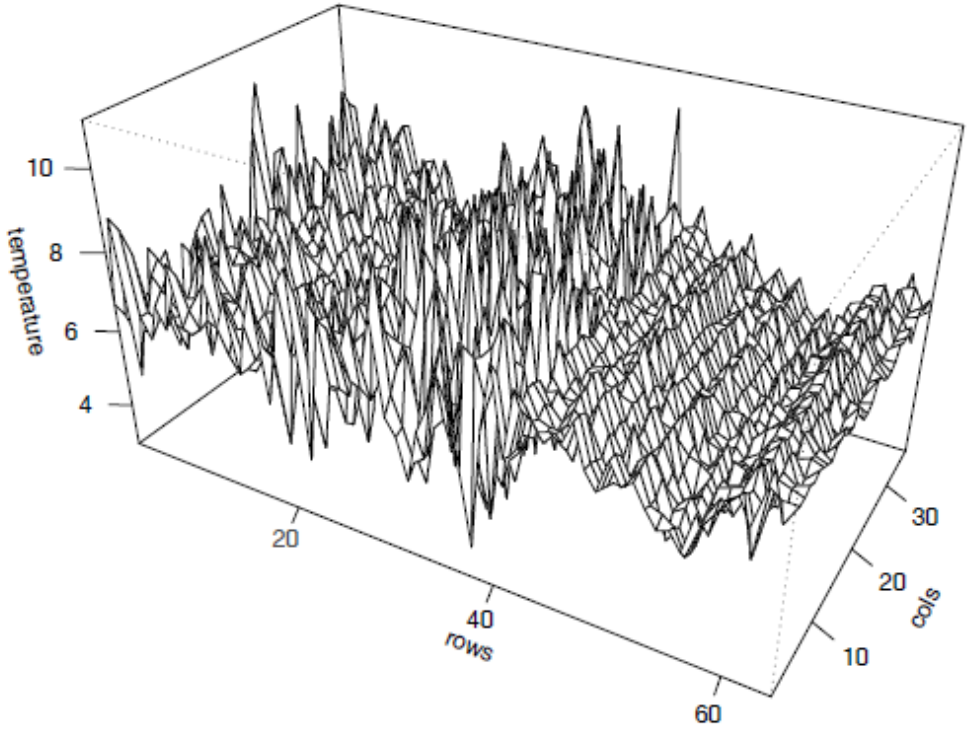

Fig. 1.15. Two-dimensional time series of temperature measurements taken on a rectangular field (64×36 with 17-foot spacing). Data are from Bazza et al. (1988).

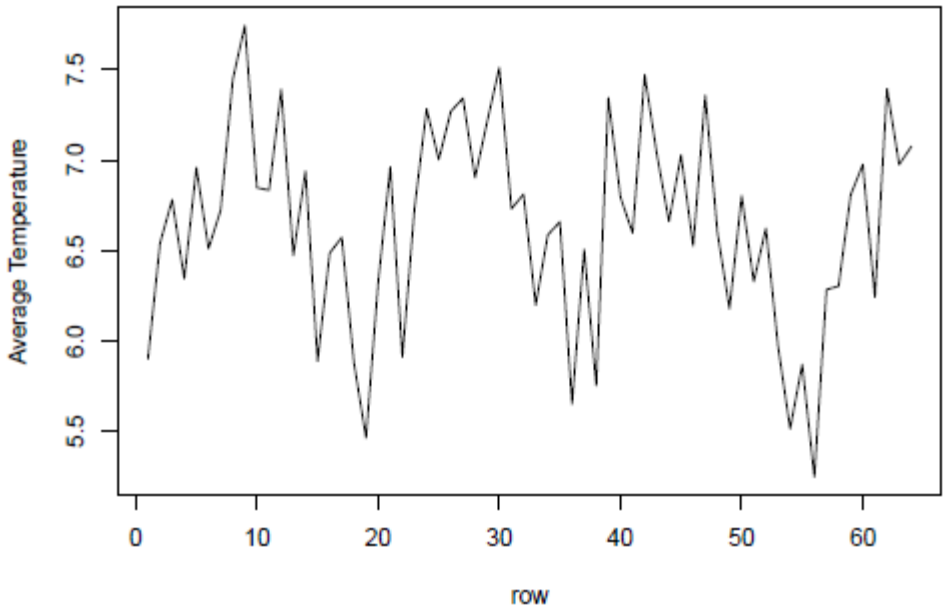

Fig. 1.16. Row averages of the two-dimensional soil temperature profile. $\bar{x}_{s_1} = \sum_{s_2} x_{s_1,s_2}/36$.

The autocovariance function of a stationary multidimensional process, $x_{\boldsymbol{s}}$, can be defined as a function of the multidimensional lag vector, say, $\boldsymbol{h} = (h_1, h_2, \ldots, h_r)'$, as

$$\gamma(\boldsymbol{h}) = E[(x_{\boldsymbol{s}+\boldsymbol{h}} - \mu)(x_{\boldsymbol{s}} - \mu)], \tag{1.48}$$

where

$$\mu = E(x_{\boldsymbol{s}}) \tag{1.49}$$

does not depend on the spatial coordinate $\boldsymbol{s}$. For the two dimensional temperature process, (1.48) becomes

$$\gamma(h_1, h_2) = E[(x_{s_1+h_1,s_2+h_2} - \mu)(x_{s_1,s_2} - \mu)], \tag{1.50}$$

which is a function of lag, both in the row (h_1) and column (h_2) directions.

The multidimensional sample autocovariance function is defined as

$$\widehat{\gamma}(\boldsymbol{h}) = (S_1 S_2 \cdots S_r)^{-1} \sum_{s_1} \sum_{s_2} \cdots \sum_{s_r} (x_{\boldsymbol{s}+\boldsymbol{h}} - \bar{x})(x_{\boldsymbol{s}} - \bar{x}), \tag{1.51}$$

where $\boldsymbol{s} = (s_1, s_2, \ldots, s_r)'$ and the range of summation for each argument is $1 \le s_i \le S_i - h_i$, for $i = 1, \ldots, r$. The mean is computed over the r-dimensional array, that is,

$$\bar{x} = (S_1 S_2 \cdots S_r)^{-1} \sum_{s_1} \sum_{s_2} \cdots \sum_{s_r} x_{s_1,s_2,\cdots,s_r}, \tag{1.52}$$

where the arguments s_i are summed over $1 \le s_i \le S_i$. The multidimensional sample autocorrelation function follows, as usual, by taking the scaled ratio

$$\widehat{\rho}(\boldsymbol{h}) = \frac{\widehat{\gamma}(\boldsymbol{h})}{\widehat{\gamma}(\boldsymbol{0})}. \tag{1.53}$$

Example 1.27 Sample ACF of the Soil Temperature Series

The autocorrelation function of the two-dimensional (2d) temperature process can be written in the form

$$\widehat{\rho}(h_1, h_2) = \frac{\widehat{\gamma}(h_1, h_2)}{\widehat{\gamma}(0, 0)},$$

where

$$\widehat{\gamma}(h_1, h_2) = (S_1 S_2)^{-1} \sum_{s_1} \sum_{s_2} (x_{s_1+h_1, s_2+h_2} - \bar{x})(x_{s_1, s_2} - \bar{x})$$

Figure 1.17 shows the autocorrelation function for the temperature data, and we note the systematic periodic variation that appears along the rows. The autocovariance over columns seems to be strongest for $h_1 = 0$, implying columns may form replicates of some underlying process that has a periodicity over the rows. This idea can be investigated by examining the mean series over columns as shown in Figure 1.16.

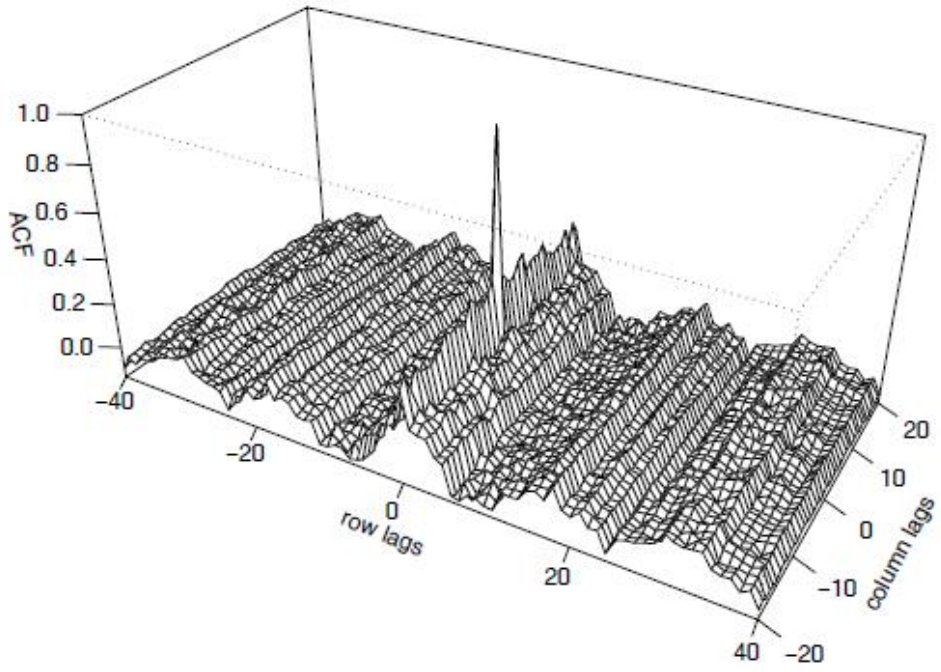

Fig. 1.17. Two-dimensional autocorrelation function for the soil temperature data.