Spatial Variability of Soil Physical Properties

College on Soil Physics – 25/02 to 01/03/2013

30th Anniversary (1983-2013)

*Luís Carlos Timm – Federal University of Pelotas, Brazil;
Regular Associate of the ICTP; lctimm@ufpel.edu.br
*Klaus Reichardt – Soil Physics Lab., CENA, USP, Brazil
*Adilson Bamberg – Researcher at Embrapa CPACT, Brazil

Classical Statistics...

- **INDEPENDENCE** of observations among themselves;

- **SAMPLING LOCATIONS** in the field are **IGNORED**, **DISREGARDING** the potential **SPATIAL DEPENDENCE** of observations within a field;

- **INADEQUATE** experimental design as a result.

Taking account the spatial dependence of observations.....

-**Geostatistics Analysis**: semivariograms, kriging, cross-semivariograms, co-kriging, etc

-**Time/Spatial Series Analysis**

Time/Spatial Series

Series definition:

-**Physical phenomena that, when observed and numerically quantified, result in a sequence of data distributed along time (or space).**

<u>*“Time”*</u> *Series (time data sequences) examples:*

a) ***<u>weekly</u>*** **average values of soil water storage at a given location;**

b) ***<u>yearly</u>*** **sugarcane crop yield data for a given field; etc.**

Spatial series (space data sequences) examples:

a) soil organic carbon content measured across a field;

b) soil water content values measured across a sugarcane field on the same day; etc.

“Basic objectives” to analyze a Time/Space Series (Tukey,1980):

-modeling of the process under consideration;

-obtaining conclusions in statistical terms; and

-evaluating model´s ability in terms of forecast.

"Two" ways to analyze a temporal (spatial) series:

1st) "frequency" domain models: presence of a <u>periodic</u> phenomenum.

<u>Examples</u> are:

-<u>Spectral</u> and <u>Cospectral</u> analyses: many applications in the soil-plant-atmosphere system.

2nd) "time (space)" ***domain models*****: to identify the stationary components (aleatory or random variables) and the nonstationary components which define the mean function of the process.**

Examples are:

AR models; ARIMA models;

"STATE-SPACE" models

Analysis of a series in the "time (or space)" domain:

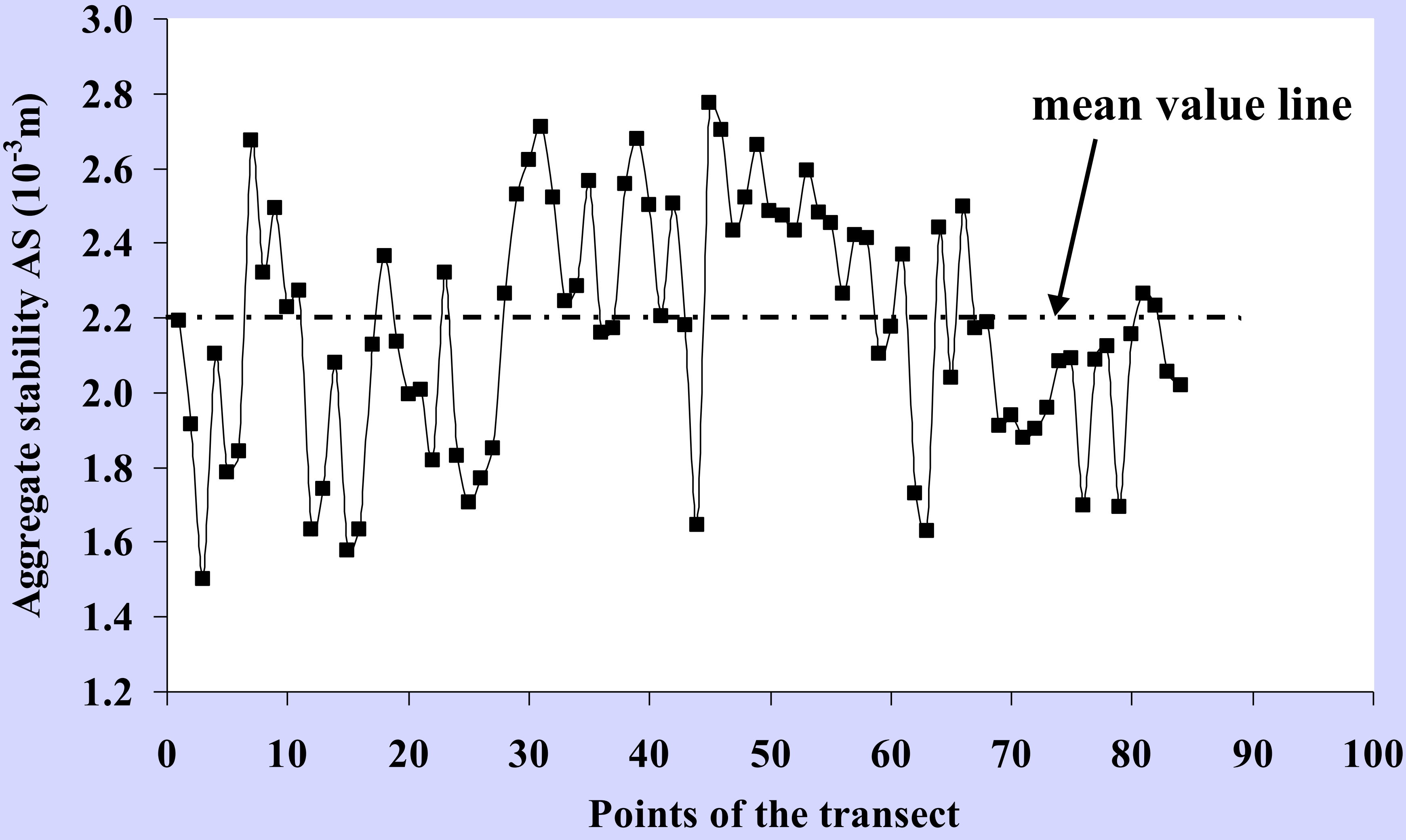
Frequent assumption:

***series is* "stationary"**

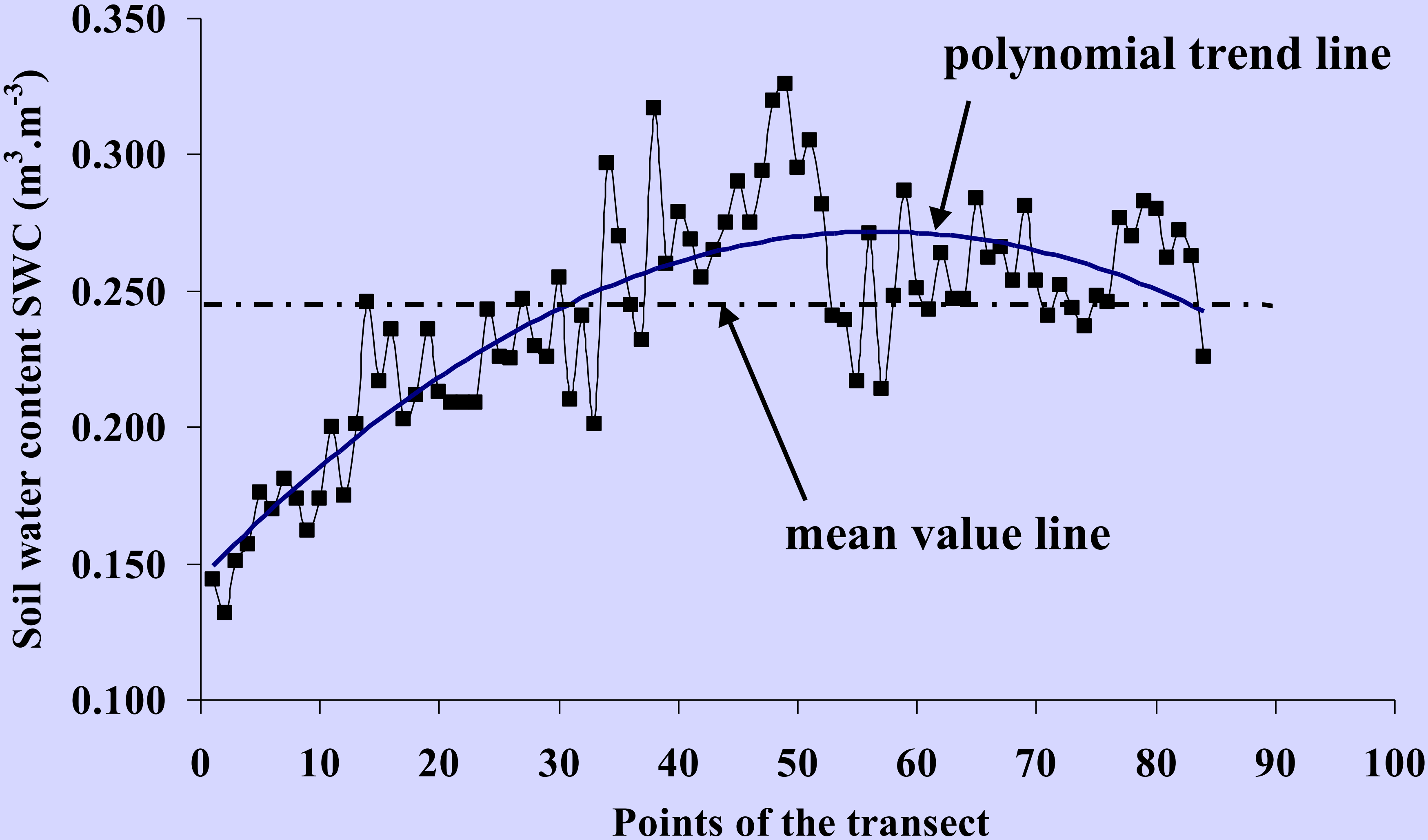
Stationarity: What does it mean ?

-series develops in a "random way" in time (or space) reflecting some sort of a "stable equilibrium" (no trend line).

"Stationary" data series: example



"Nonstationary" data series:example



Statistical tools for

analyzing and characterizing

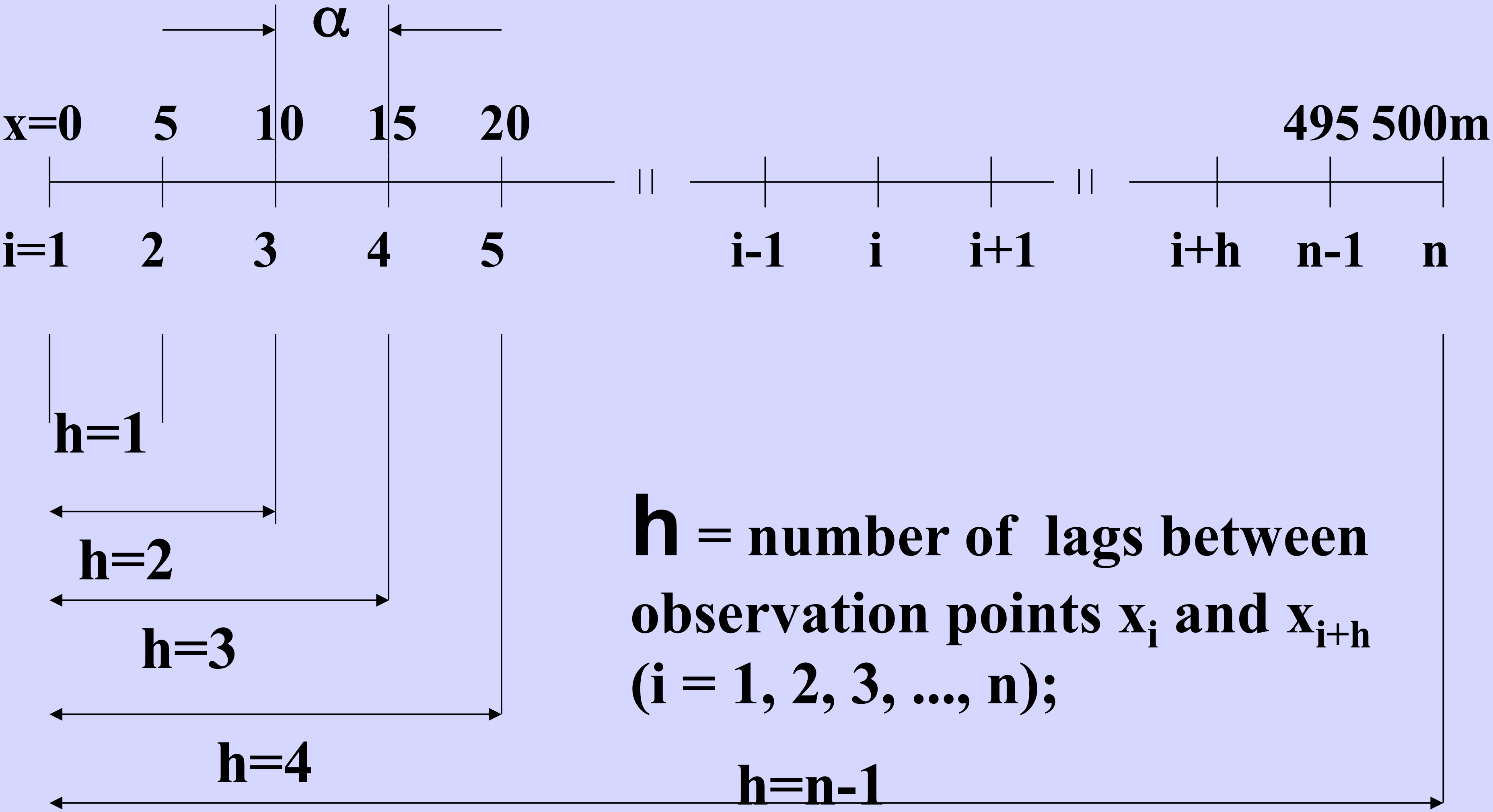
"time (or spatial)"

variability of data sets

a) **Autocorrelation function: to indicate the distance of "auto-dependence" between "adjacent observations" of a variable.**

"Space series" are collected along <u>transects</u> at spacings of "<u>α</u>" in cm, m, km, etc.

α = is the distance between observation points of the variable Y. Example: Y = soil temperature; $\alpha = 5$ m

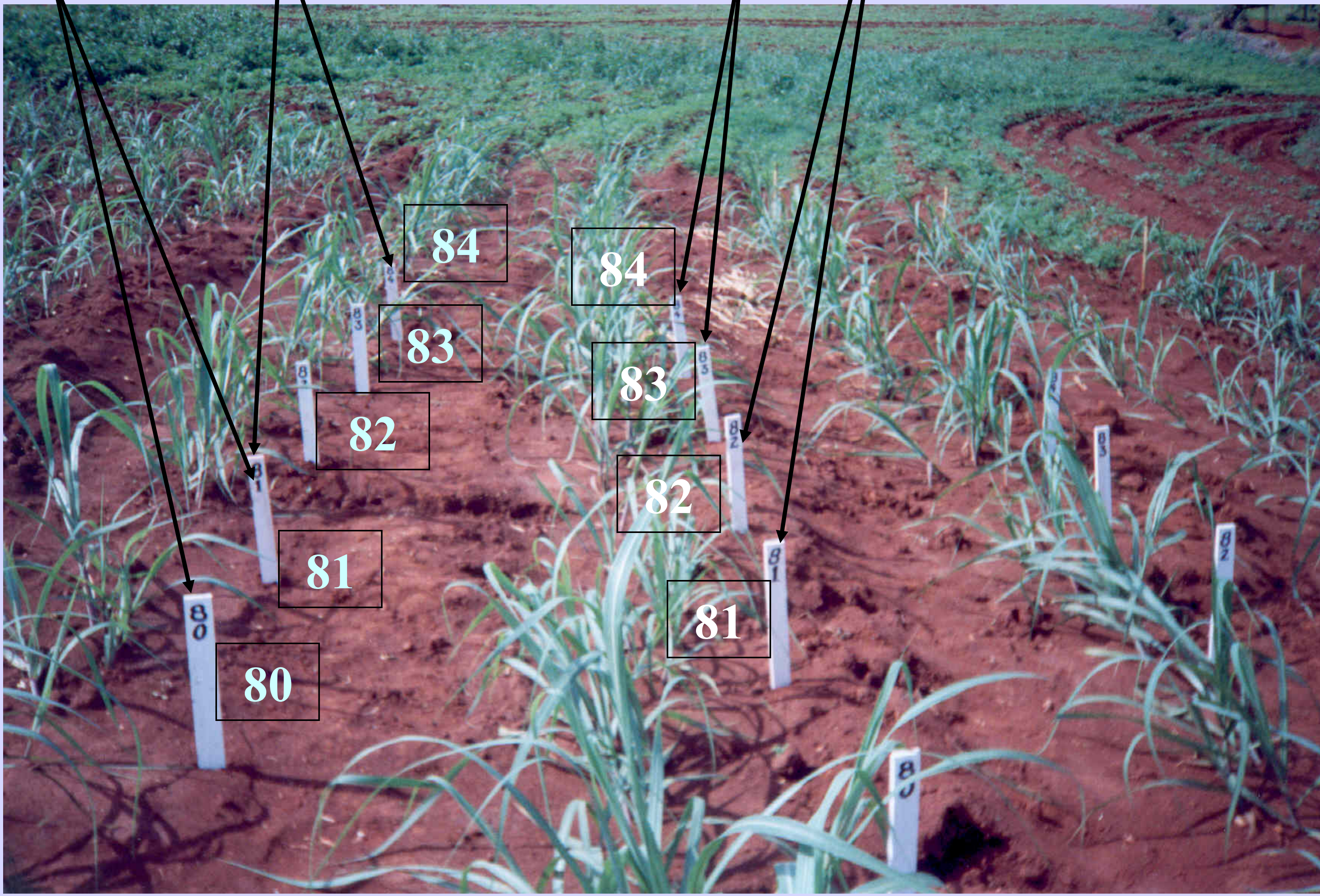


h = number of lags between observation points x_i and x_{i+h} (i = 1, 2, 3, ..., n);

h = 1 lag

h = 3 lags

α = 1 m



"Autocorrelation" *coefficient* r(h):

$$r(h) = \frac{C(h)}{s^2}$$

r(h) = **autocorrelation coefficient as a function of the number of lags h between observation points;**

C(h) = **covariance function between adjacent observation points;**

s^2 = **variance of the sample.**

Covariance C(h):

$$C(h) = \frac{1}{n-h}\sum_{i=1}^{n-h}\left[Y(x_{i+h}) - \overline{Y}\right]\left[Y(x_i) - \overline{Y}\right]$$

$Y(x_i)$ and $Y(x_{i+h})$ = variable values under consideration at position x_i and x_{i+h}, respectively;

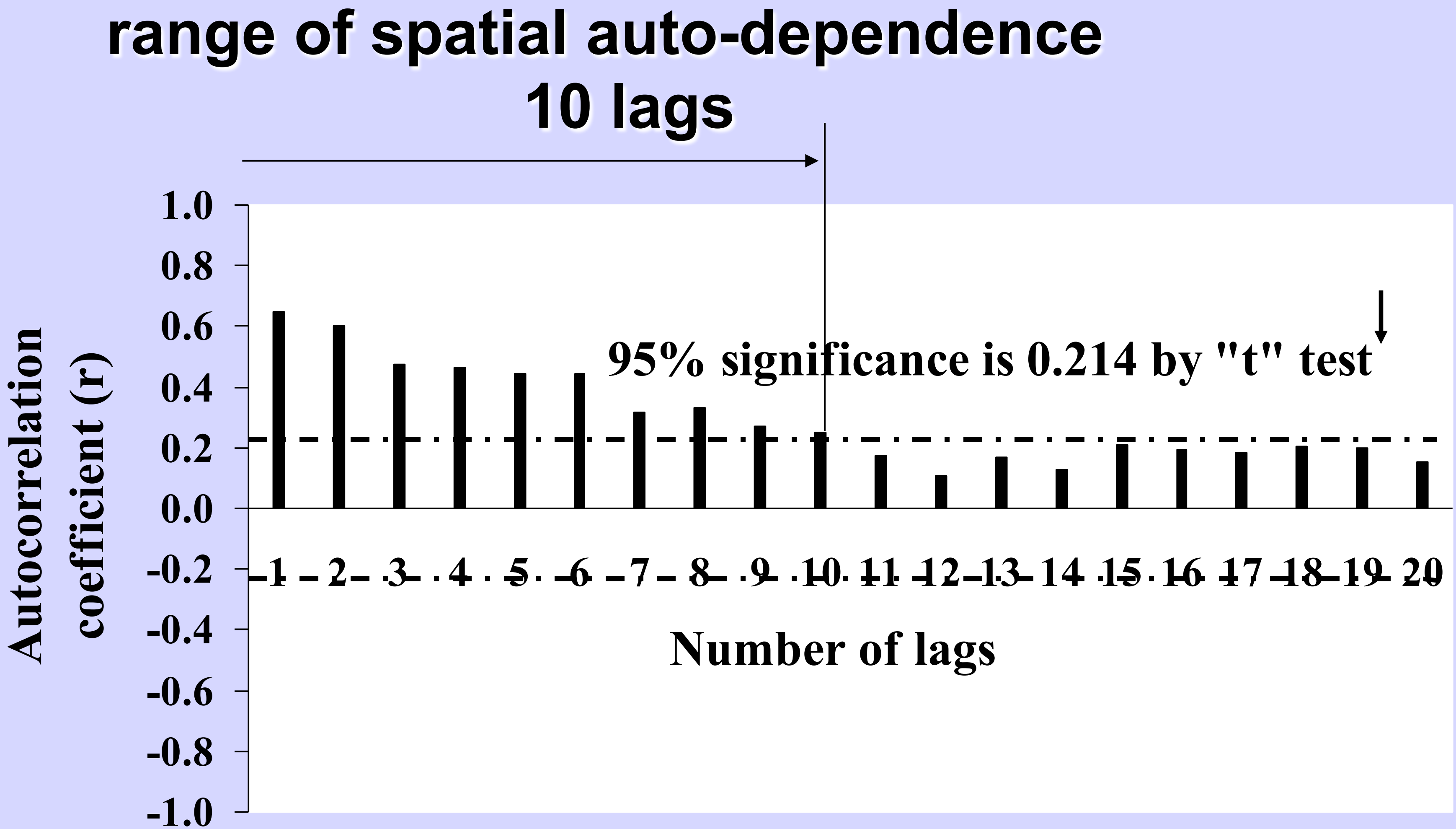
n = total number of observation points;

$\overline{Y}$ = is the mean value of the variable Y, for all n observations.

Autocorrelogram plot

autocorrelation coefficient (r) is plotted as a function of the number of lags (h) between observation points x_i and x_{i+h}.

Autocorrelogram *plot: an example*



REMARKS:

-observations Y have to be collected at regularly spaced intervals α;

-range of variation of r(h) coefficient: between +1 and -1;

REMARKS: *cont..*

-for h =0, i.e., the correlation between Y(xi) and Y(xi) is obviously equal to 1;

-Plotting <u>r</u> as a function of <u>h</u>

Autocorrelogram of the variable Y.

b) Crosscorrelation function (CCF):

-to indicate the spatial correlation between

two sets of variables: $Y(x_i)$ and $W(x_i)$

observed at the same locations x_i.

Example:

Y = soil temperature; W = soil water content

Crosscorrelation coefficient

$[r_c(h)]$:

$$r_c(h) = \frac{\mathrm{cov}_{YW}(h)}{s_Y \times s_W}$$

Covariance function : two variables

$$\text{cov}_{YW}(h) = \frac{1}{n-h}\sum_{i=1}^{n-h}\left[Y(x_i) - \overline{Y}\right]\left[W(x_{i+h}) - \overline{W}\right]$$

REMARKS:

-observations Y and W have to be collected at regularly spaced intervals α;

REMARKS *(cont.)*:

- range of variation of $r_c(h)$ coefficient: between +1 and -1;

- for h=0, i.e., the correlation between $Y(x_i)$ and $W(x_i)$ is the linear regression coefficient obtained by classical statistics;

- the same procedure is used for more neighbors, obtaining values of $r_c(h)$ and $r_c(-h)$.

REMARKS *(cont.)*

- the **crosscorrelation coefficient** $r_c(h)$ **is obtained for h=1 and h=-1, i.e., we have two pairs [- (Y_i, W_{i+1}) ; + (Y_i, W_{i-1})];**

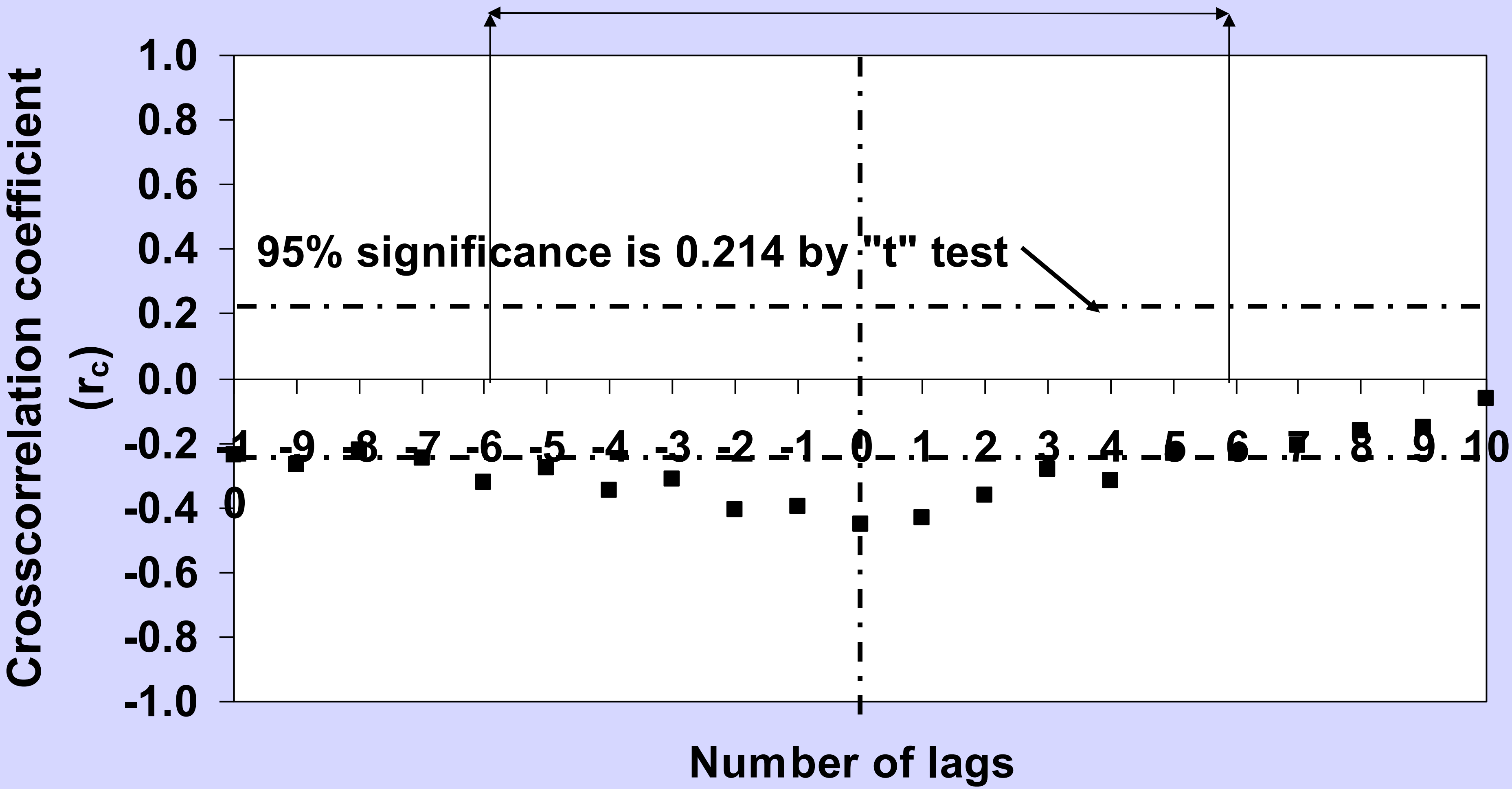
-Plotting r_c as a function of h

↓

Crosscorrelogram between $Y(x_i)$ and $W(x_i)$

Spatial dependence: two variables

6 lags in both directions



“State-Space” approach:

-the “**State-Space**” model of a stochastic process involving j data sets Y_j, all collected at the same locations x_i, is based on the property of Markovian systems, i.e., the independence of the future of the process in relation to its past, once given the present state;

-It is a combination of **two** systems of equations:

The Abdus Salam International Centre for Theoretical Physics

Universidade Federal de Pelotas RS - Brasil

CNPq Conselho Nacional de Desenvolvimento Científico e Tecnológico

Fapergs

1st: "Observation" equation

$$Y_j(x_i) = M_{jj}(x_i) Z_j(x_i) + v_{Y_j}(x_i) \qquad (1)$$

Y_j = **observation vector** of the process at **location** x_i;

M_{jj} = **observation matrix** at **position** x_i;

Z_j = **non observed state vector** of the process at **location** x_i;

v_{Yj} = **observation error vector** at position x_i.

The "matrix M_{jj}" comes from a set of j linear observation equations (all at position i):

$$Y_1(x_i) = m_{11}Z_1(x_i) + m_{12}Z_2(x_i) + \dots\dots\dots + m_{1j}Z_j(x_i) + v_{Y_1}(x_i)$$

$$Y_2(x_i) = m_{21}Z_1(x_i) + m_{22}Z_2(x_i) + \dots\dots\dots + m_{2j}Z_j(x_i) + v_{Y_2}(x_i)$$

$$Y_j(x_i) = m_{j1}Z_1(x_i) + m_{j2}Z_2(x_i) + \dots\dots\dots + m_{jj}Z_j(x_i) + v_{Y_j}(x_i)$$

Which can be written in the matrix form:

$$
\begin{bmatrix} Y_1(x_i) \\ Y_2(x_i) \\ \\ Y_j(x_i) \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} \ldots\ldots\ldots m_{1j} \\ m_{21} & m_{22} \ldots\ldots\ldots m_{2j} \\ \\ m_{j1} & m_{j2} \ldots\ldots\ldots m_{jj} \end{bmatrix} \times \begin{bmatrix} Z_1(x_i) \\ Z_2(x_i) \\ \\ Z_j(x_i) \end{bmatrix} + \begin{bmatrix} v_{Y_1}(x_i) \\ v_{Y_2}(x_i) \\ \\ v_{Y_j}(x_i) \end{bmatrix}
$$

↓ observation vector

↓ observation matrix

↓ non observed state vector

↓ observation error vector

2nd: "State" equation

$$Z_j(x_i) = \phi_{jj}(x_i) Z_j(x_{i-1}) + u_{Z_j}(x_i) \qquad (2)$$

ϕ_{jj}= state coefficient matrix (transition matrix) at location x_i;

$Z_j(x_{i-1})$ = non observed state vector of the process at location x_{i-1};

u_{Zj} = is an error vector associated to the state at position x_i;

The <u>matrix</u> ϕ_{jj} comes from a set of j linear state equations (relating position i to position i-1):

$$Z_1(x_i) = \phi_{11} Z_1(x_{i-1}) + \phi_{12} Z_2(x_{i-1}) + \ldots\ldots\ldots\ldots + \phi_{1j} Z_j(x_{i-1}) + u_{Z_1}(x_i)$$
$$Z_2(x_i) = \phi_{21} Z_1(x_{i-1}) + \phi_{22} Z_2(x_{i-1}) + \ldots\ldots\ldots\ldots + \phi_{2j} Z_j(x_{i-1}) + u_{Z_2}(x_i)$$

$$Z_j(x_i) = \phi_{j1} Z_1(x_{i-1}) + \phi_{j2} Z_2(x_{i-1}) + \ldots\ldots\ldots\ldots + \phi_{jj} Z_j(x_{i-1}) + u_{Z_j}(x_i)$$

Or in the <u>matrix</u> form:

$$\begin{bmatrix} Z_1(x_i) \\ Z_2(x_i) \\ \\ Z_j(x_i) \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \cdots\cdots\cdots \phi_{1j} \\ \phi_{21} & \phi_{22} \cdots\cdots\cdots \phi_{2j} \\ \\ \phi_{j1} & \phi_{j2} \cdots\cdots\cdots \phi_{jj} \end{bmatrix} \times \begin{bmatrix} Z_1(x_{i-1}) \\ Z_2(x_{i-1}) \\ \\ Z_j(x_{i-1}) \end{bmatrix} + \begin{bmatrix} u_{Z_1}(x_i) \\ u_{Z_2}(x_i) \\ \\ u_{Z_j}(x_i) \end{bmatrix}$$

non observed state vector at position x_i

state coefficient matrix

non observed state vector at position x_{i-1}

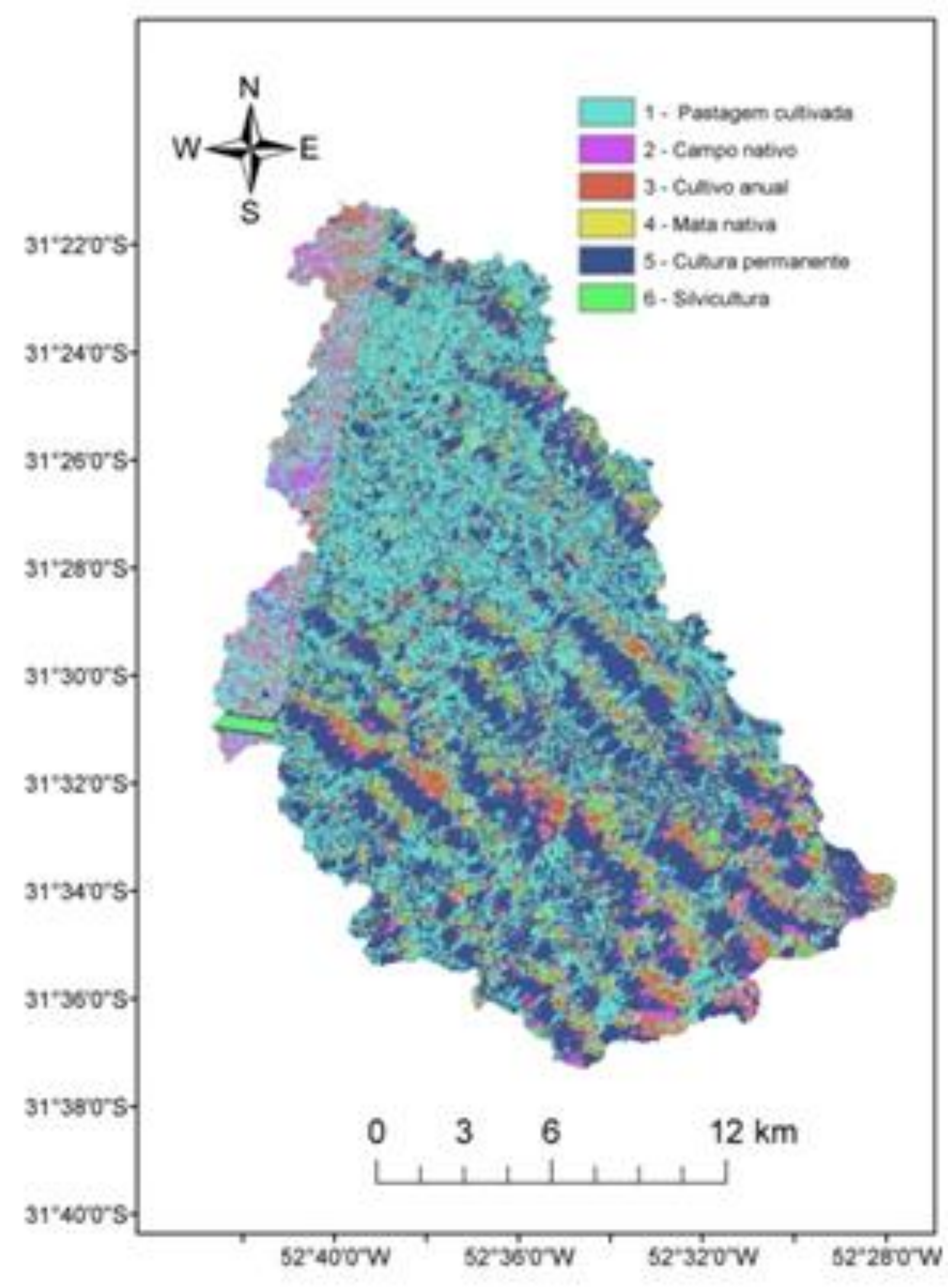
state error vector

Applications of the "State-space" approach:

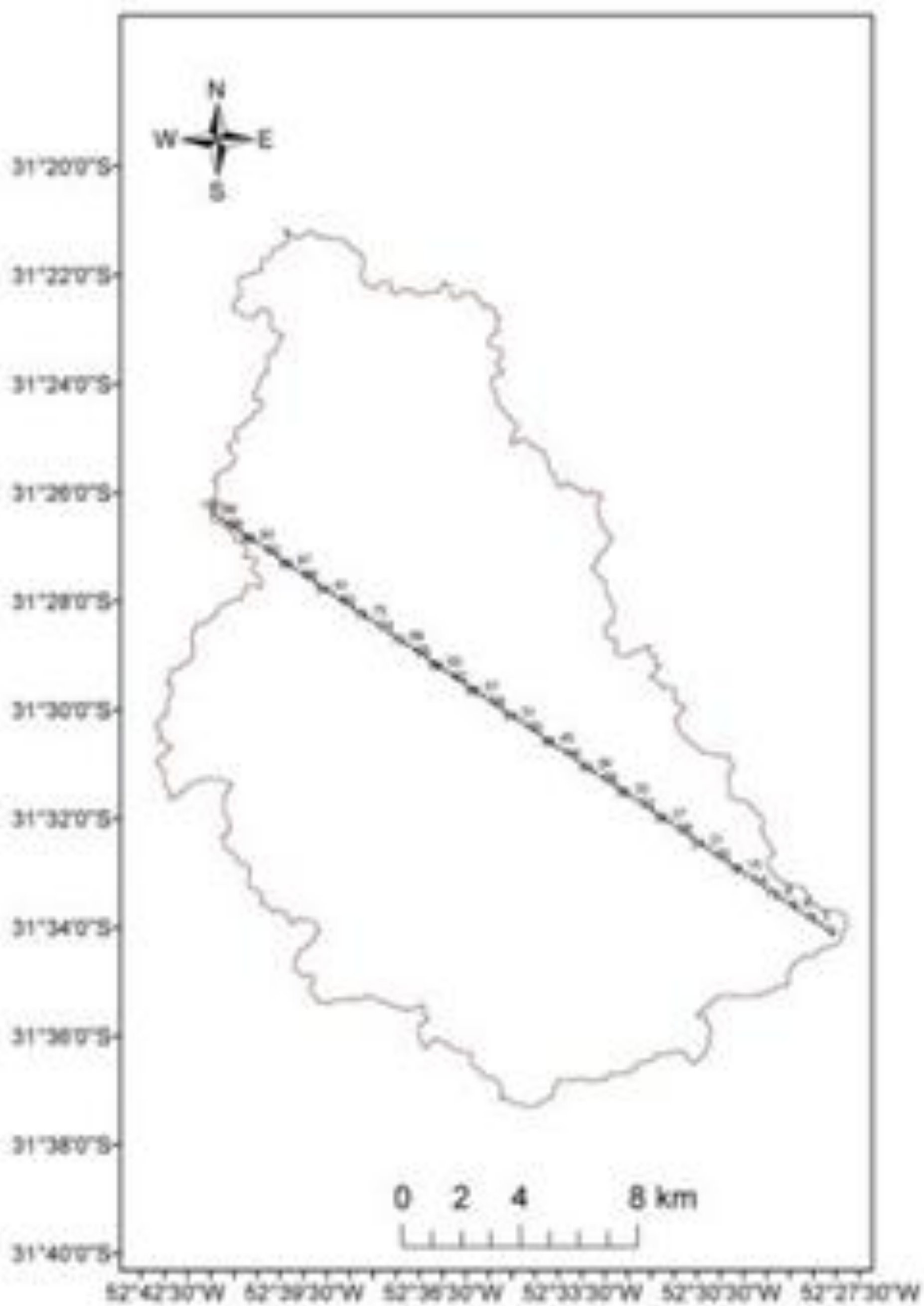
"<u>watershed</u>" scale

Arroio Pelotas watershed

drainage area: 362 km^2

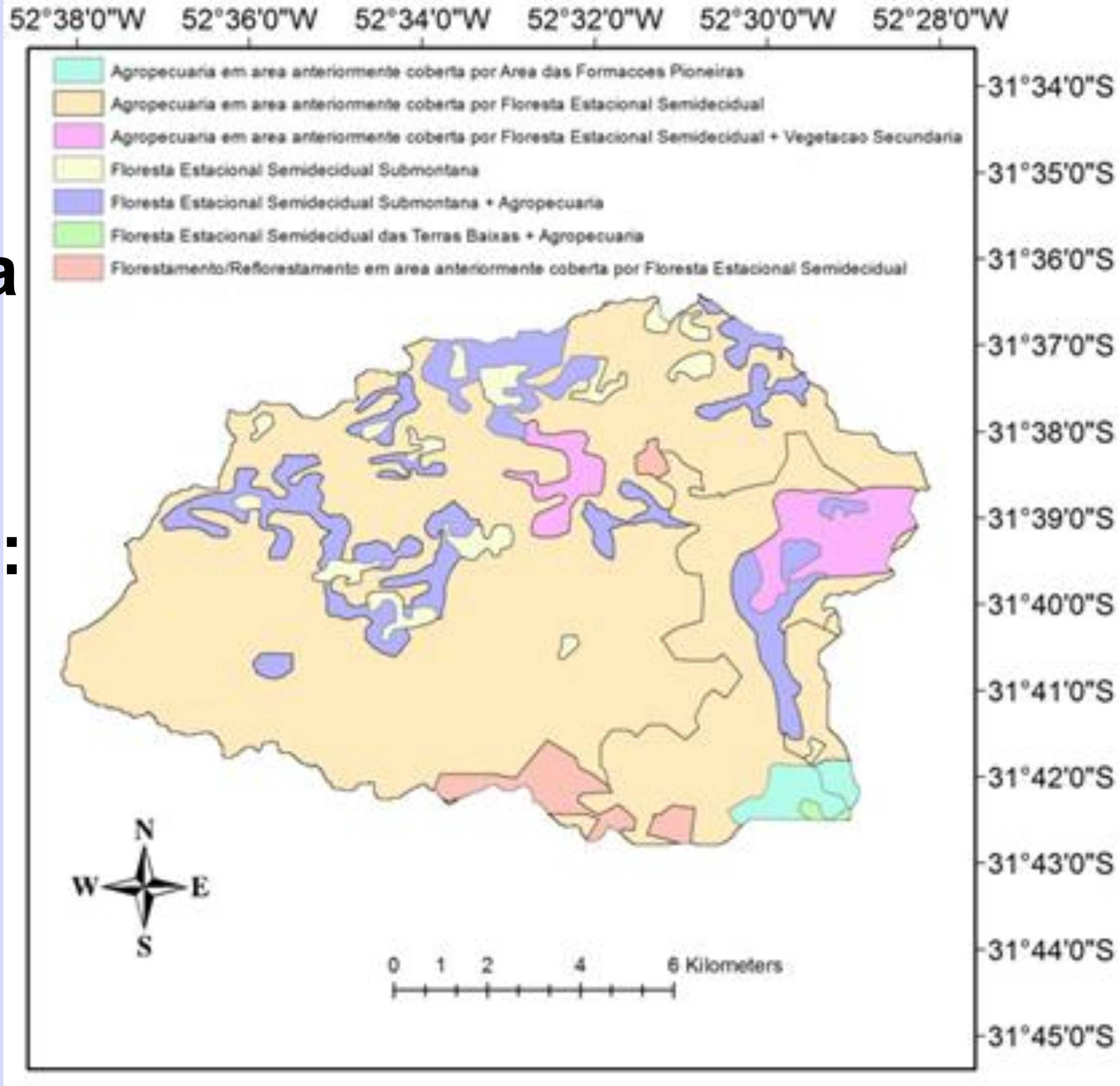


25 km spatial transect

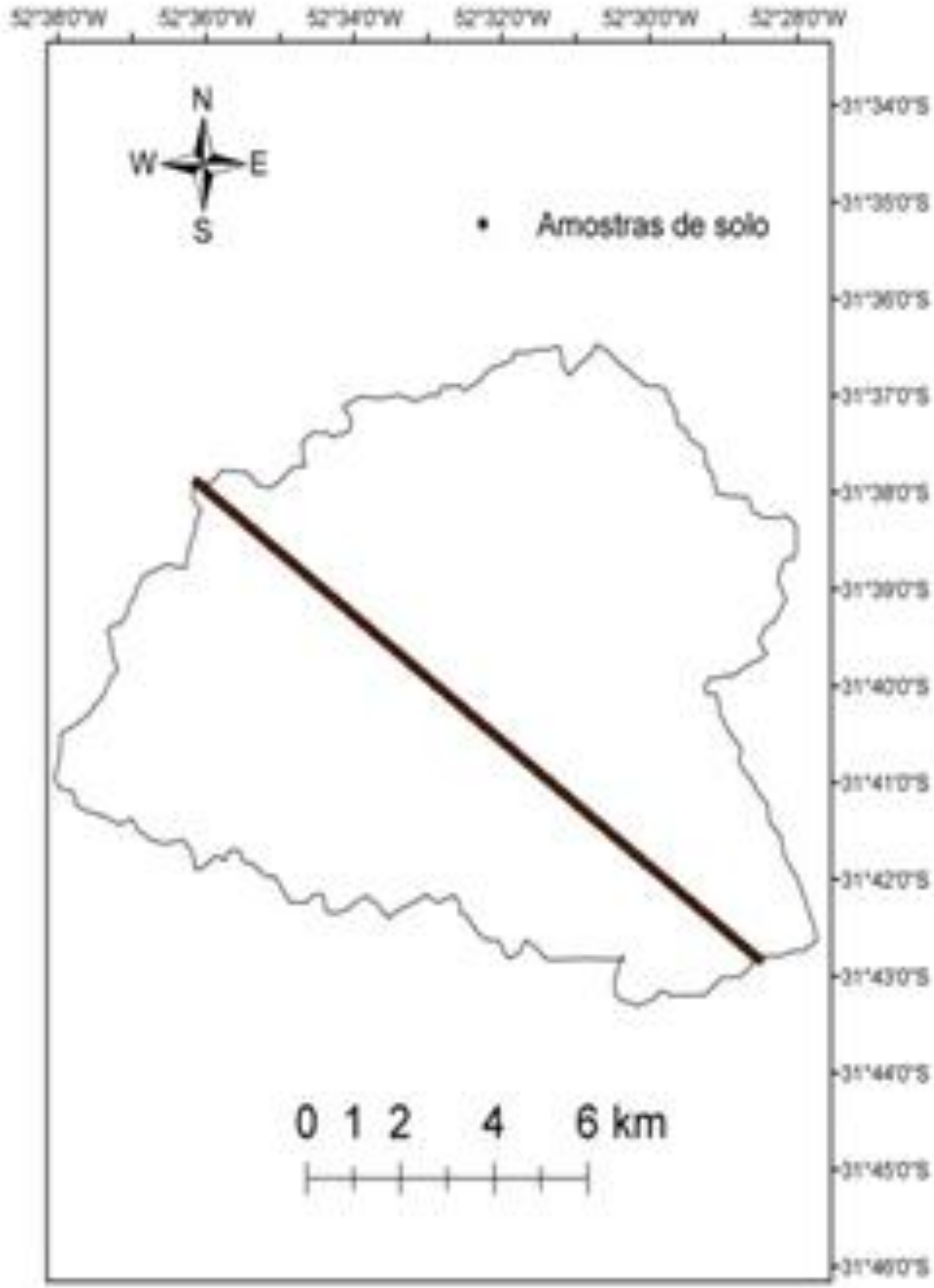


Arroio Fragata watershed

drainage area: 133 km^2



15 km spatial transect



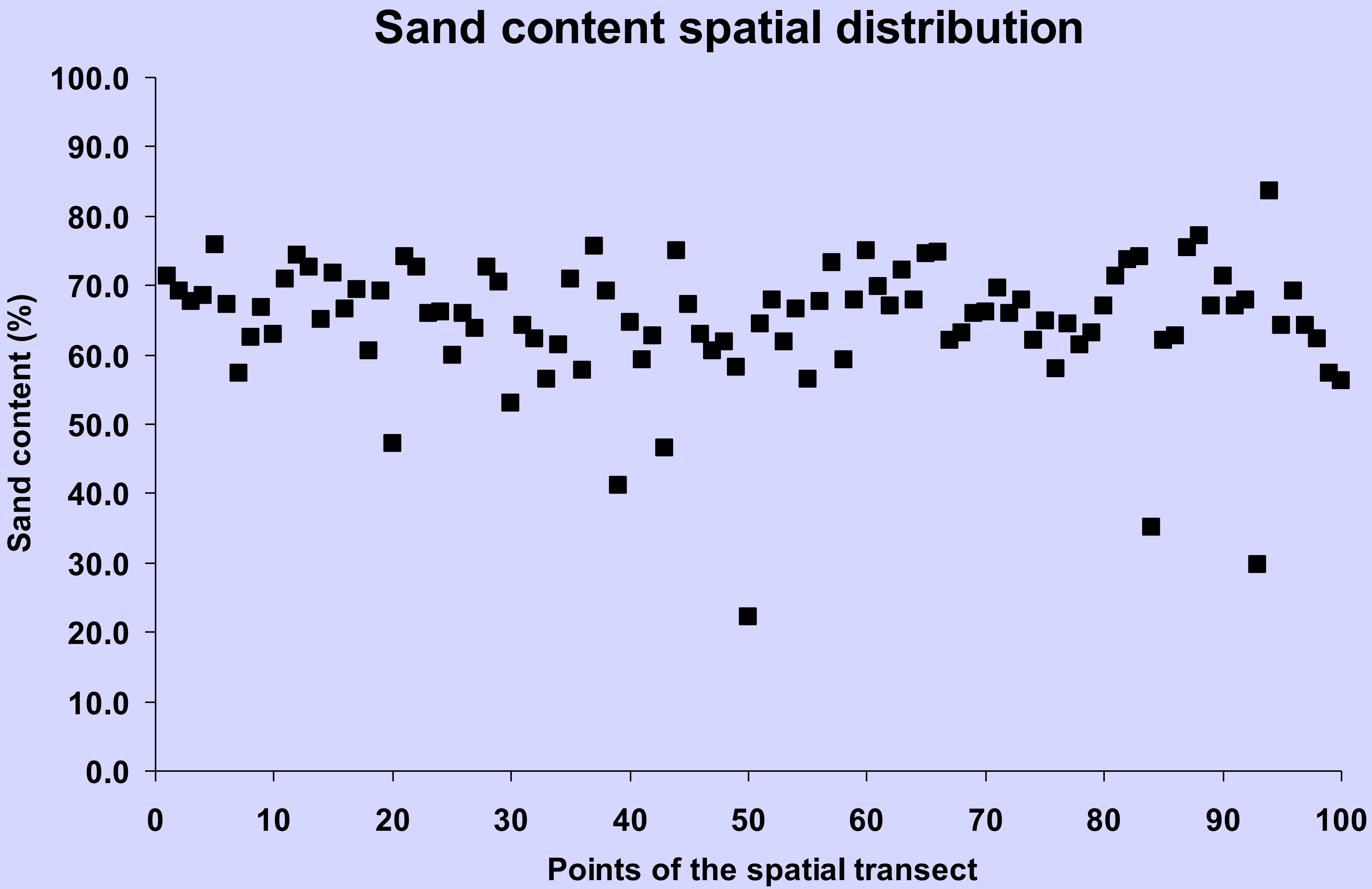
Preliminary Results

“Arroio Fragata” watershed

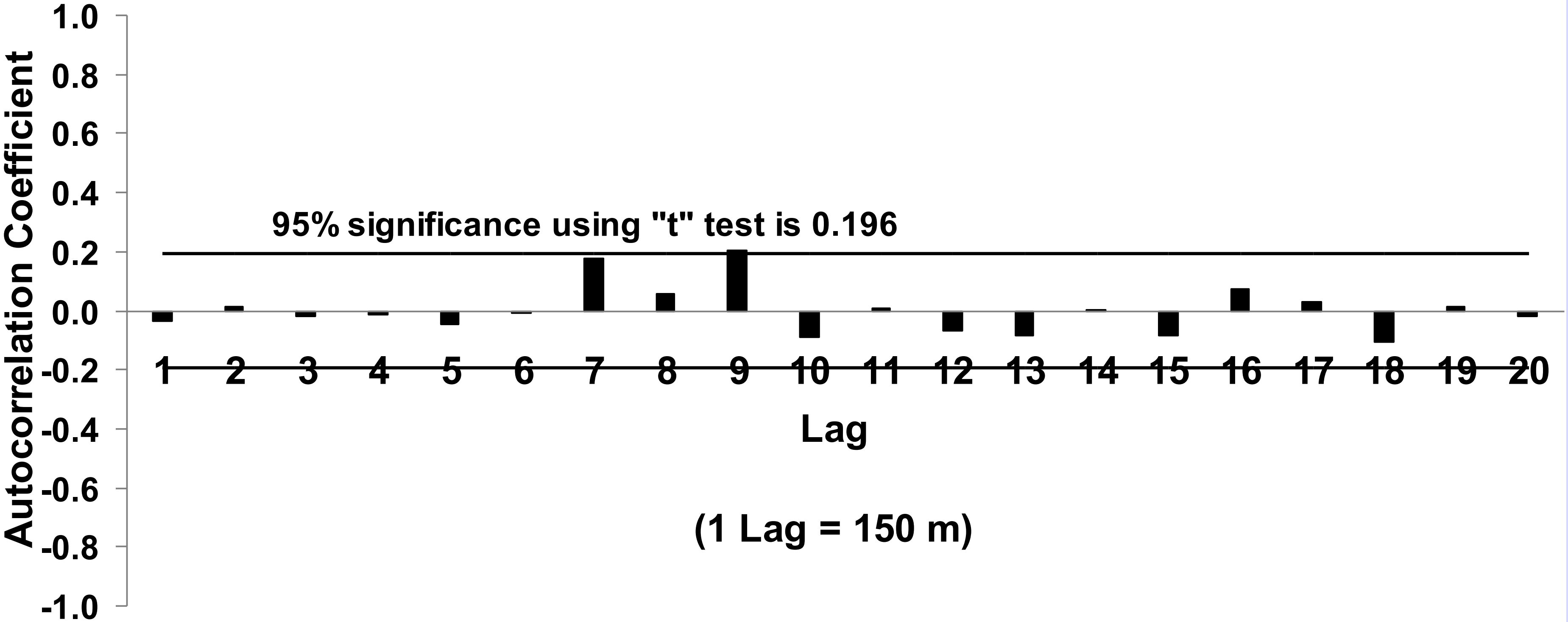
-**Soil samples**: spaced 150 m from each other totalizing 100 samples;

- Evaluated soil layer: 0- 0.20 m depth

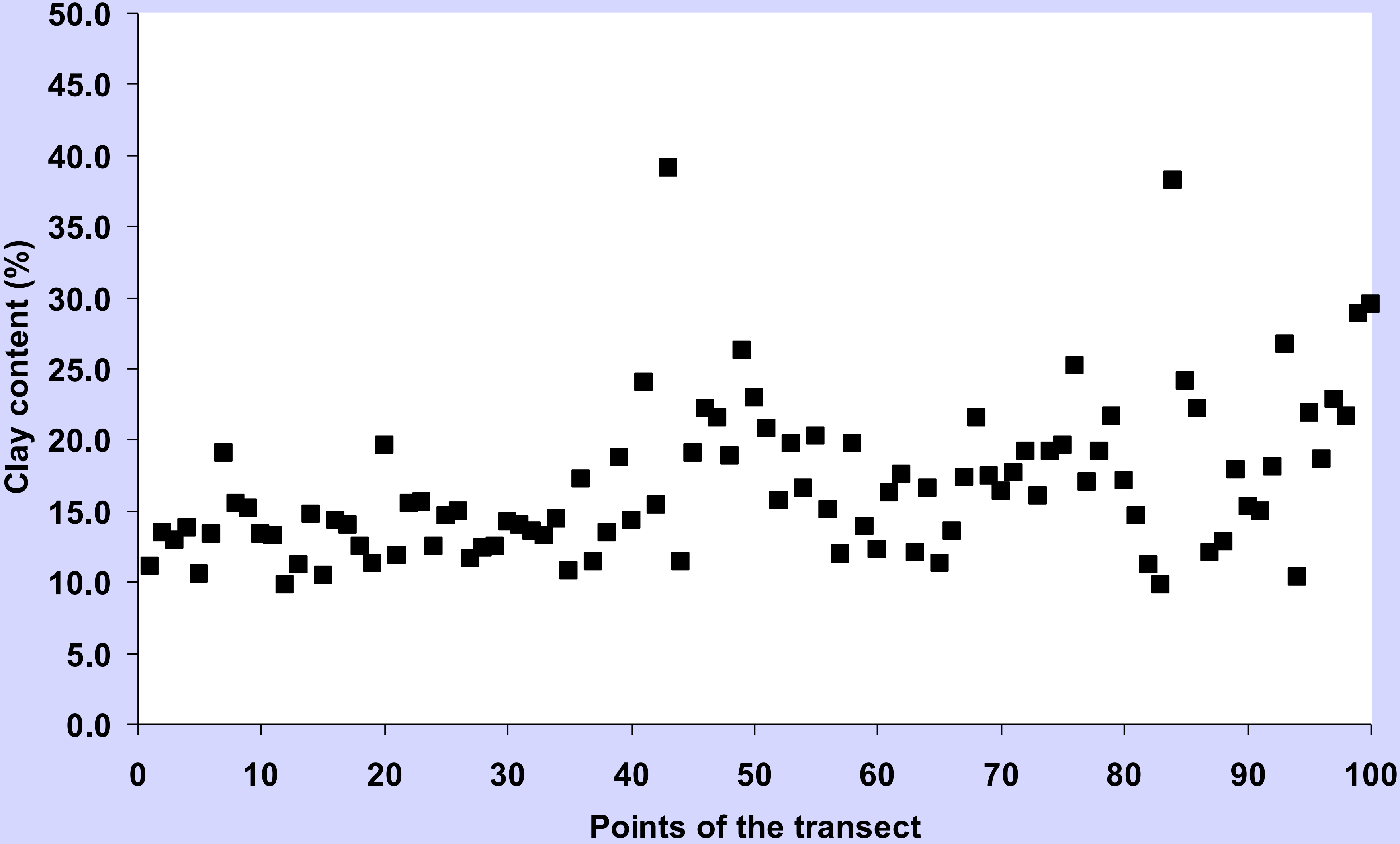
-**Determined soil physical and hydraulic properties**: soil texture, soil bulk density, soil macroporosity, SWRC, ...


Sand content spatial distribution
Sand content (%)
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
0
10
20
30
40
50
60
70
80
90
100
Points of the spatial transect

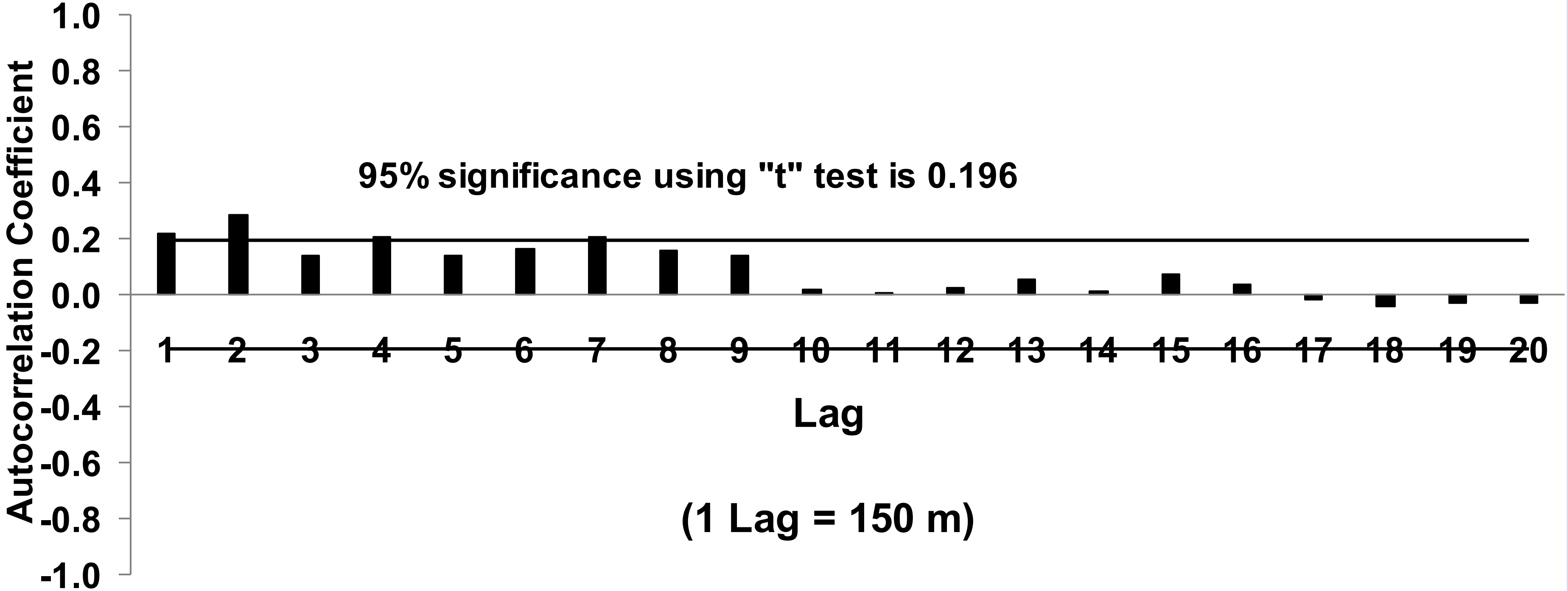
Autocorrelogram for Sand content



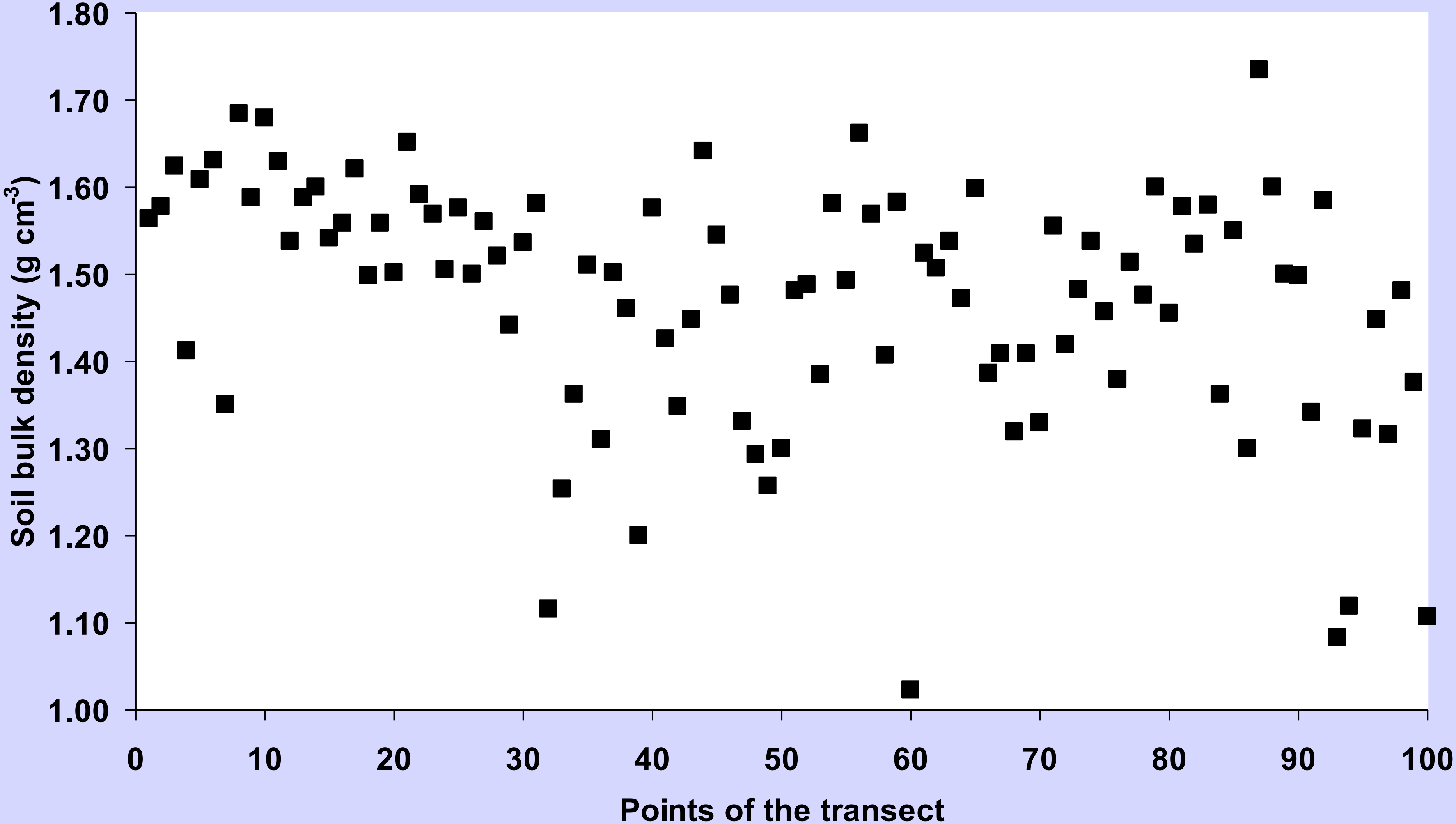
Clay content spatial distribution



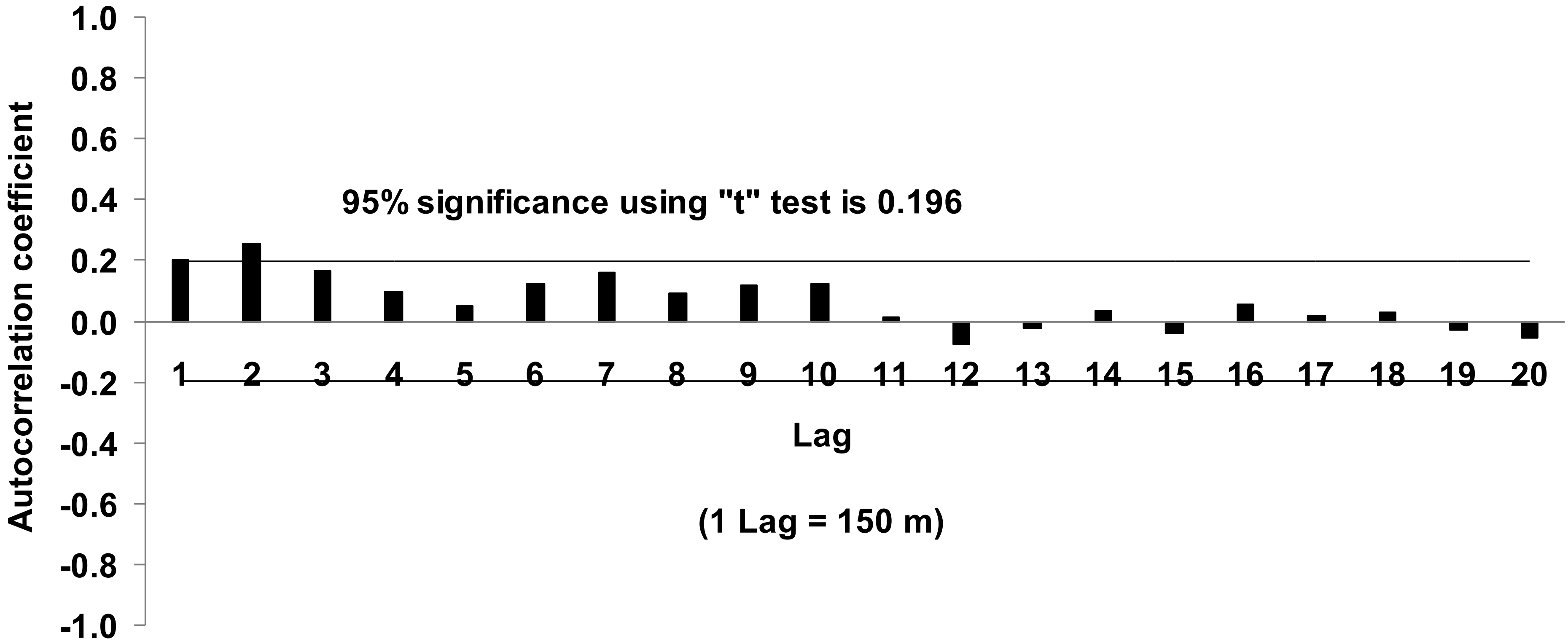
Autocorrelogram for Clay content



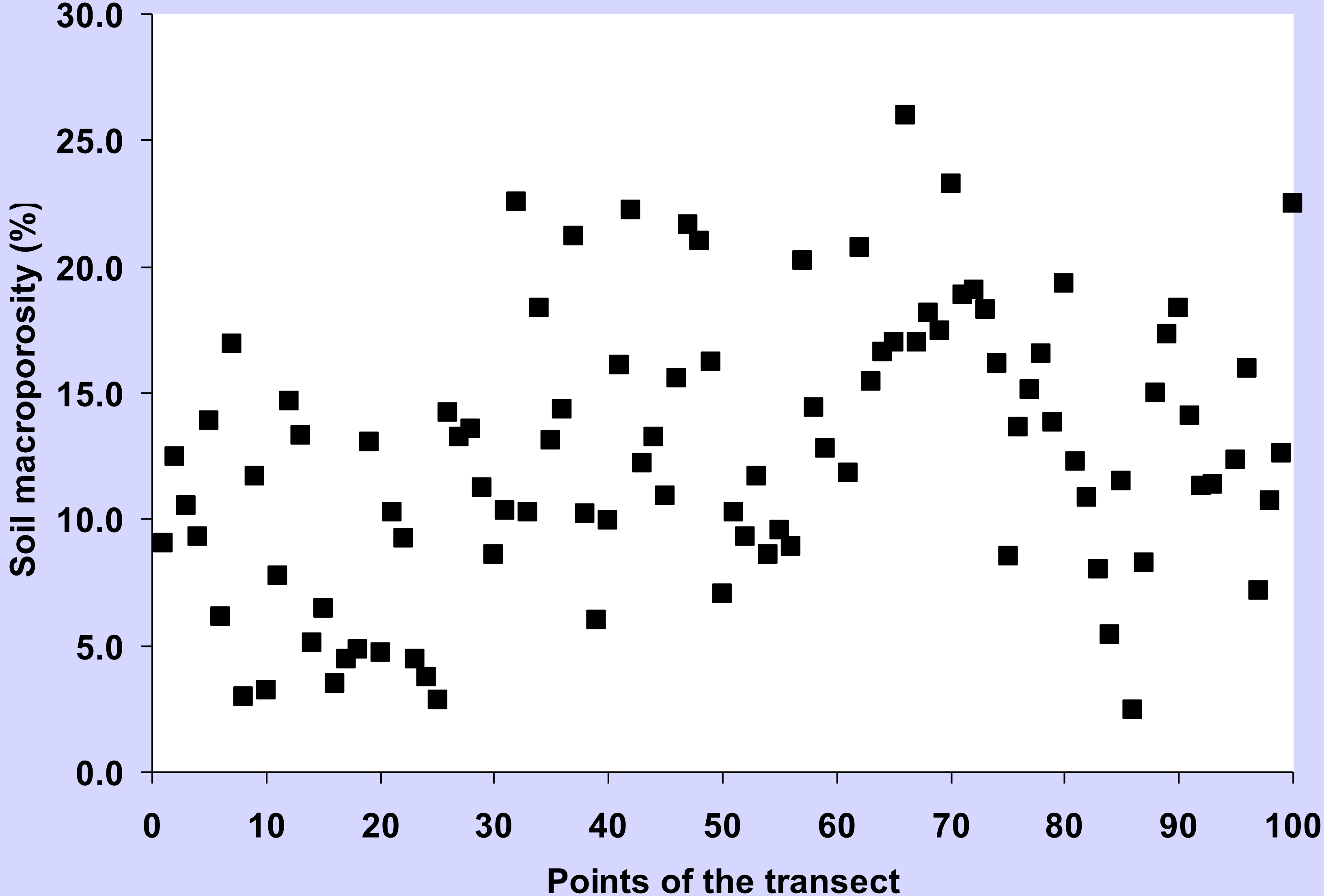
Soil bulk density spatial distribution

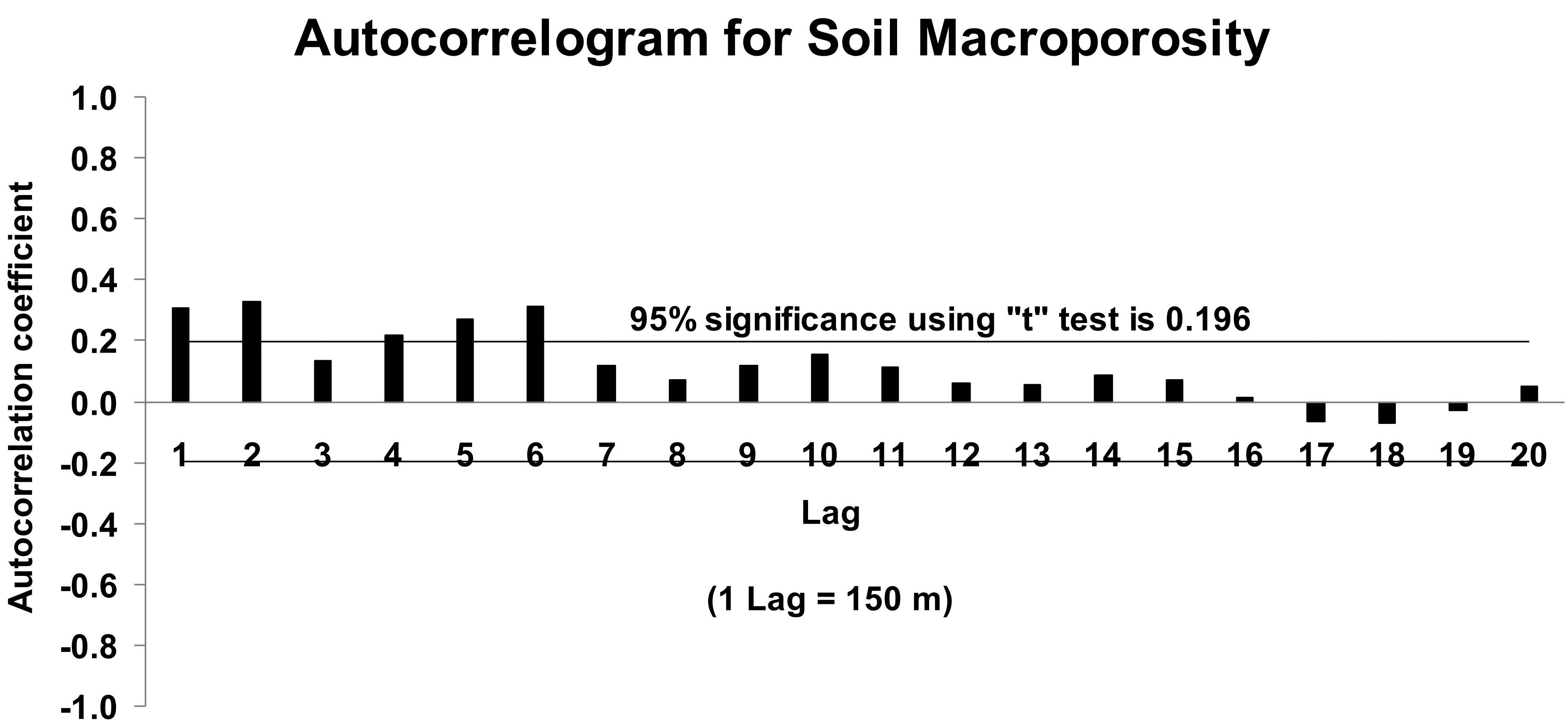


Autocorrelogram for Soil Bulk Density



Soil macroporosity spatial distribution





Autocorrelogram for Soil Macroporosity
Autocorrelation coefficient
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
95% significance using "t" test is 0.196
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Lag
(1 Lag = 150 m)

Crosscorrelogram

Soil macroporosity and Soil bulk density