class #25

Analysis of spatial data → Map Analysis

ref. Davis, J.C. (1986). Statistics and data analysis in Geology.
John Wiley and Sons, New York, NY, 646 pp.

very good and readable summary with many examples and practical applications of both temporal (somewhat skimpy) and spatial (much detail) analysis techniques

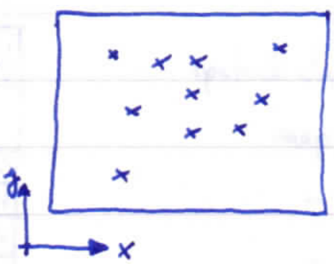
1 Introduction

types of data (a)

time series data
at location $\vec{x} = \{x_1, x_2 \ldots\}$

geography

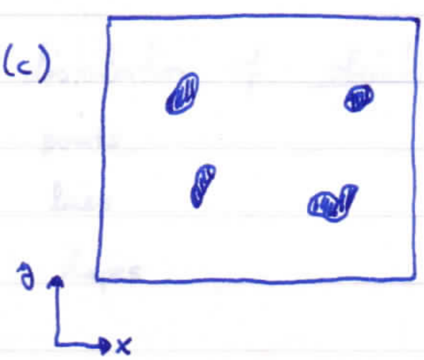
(b)



spatial data
maps of properties
"topography", vertical elevations
salinity
weather maps
$z = z(x, y)$

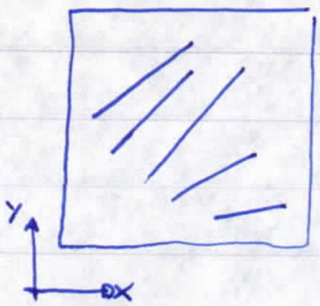
geology
mining

(c)



spatial data — shapes
maps of the presence/absence of a property such as
- mineral deposits
- distribution of ice
- plankton patchiness

(d)



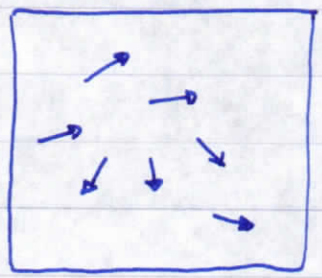
spatial data

maps of orientations

such as due to

- fault lines
- location of fronts
- leads in ice
- ripples of sand/sediment

(e)

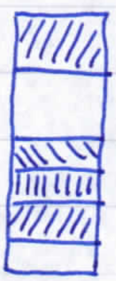


spatial data

maps of directions

- wind, velocity
- striations of glaciers
- river channels in a watershed

(f)



spatial data

map of sequences of properties

- sediment core
- ice core

⟹ spatial distribution of objects that are

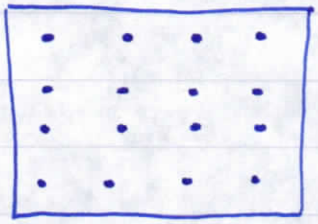
(1) points

(2) lines

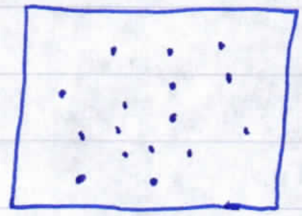
(3) shapes

wide range of techniques ; I'm just skimming the surface

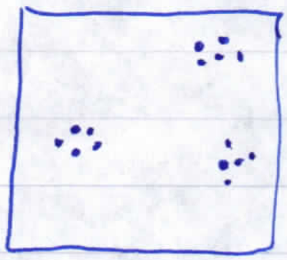
Distribution of "points" in space



regular grid



randomly scattered pts



clusters of pts.

Def.: <u>Uniform</u> density of pts in all subareas of the same size and shape is the same everywhere

<u>Regular</u> points are placed on a grid of some sort

Reliability of contour maps depends upon density AND uniformity of the observations. Often, the "contouring" is a 2 step process:

(a) generate a regular grid from quasi-uniform data → spatial interpolation

(b) interpolate contour lines through the regular grid → — " —

How to test (quantitatively) for "uniformity" and "randomness"?

construct a "statistic", e.g.,

$$\sum_{i=1}^{N} \frac{(O-E)^2}{E} = \chi^2_{N-2}$$

O - observed # of pts in subarea

E - expected # of pts in subarea

$E = \frac{\text{total \# of data}}{\text{\# of subareas } N}$

and test the hypothesis that data is unevenly distributed

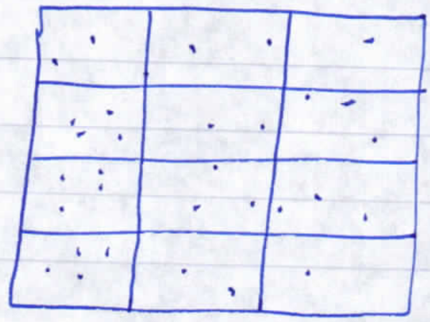
	Hypothesis correct	Hypothesis incorrect
Hypothesis is accepted	correct decision	type II error β
Hypothesis is rejected	type I error α	$\boxed{\text{correct decision}}$

- specify probability of type I error α, which is the "level of significance"
- formulate hypothesis in such a way that you hope to reject it

"... statistical tests cannot say what IS, but only what IS NOT ..."

1. Hypothesis (to be rejected, hopefully) : data is unevenly distributed
2. significance level : $\alpha = 5\%$
3. the statistic : $\chi_{N-2} = \sum_{i=1}^{N} \frac{(O-E)^2}{E}$

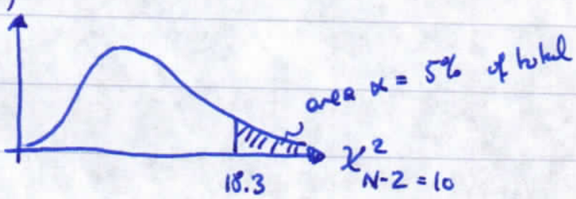
Hypothesis is rejected if $\chi^2_{N-2} = \sum_{i=1}^{N} \frac{(O-E)^2}{E} \leq \chi_{N-2,\alpha}$



$E = 123/12 \approx 10$

quadrant i	O	$(O-E^2)/E$
1	10	0
2	5	2.6
3	5	2.6
⋮	⋮	⋮
$N = 12$	8	0.48
	123	$\chi^2 = 15.23 = \sum_{i=1}^{N} (O-E)^2/E$

$\chi_{N-2;\ \alpha=0.05} = 18.3$



"Goodness of fit" leads us to conclude that the data is uniformly distributed

BUT depends upon the scale, i.e., the size of the quadrant → scale dependence

BUT random distributions can be uniform

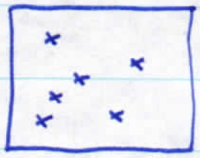
→ further test necessary to know more

to remove the "scale dependence" do a

Nearest - Neighbor Analysis

class #26

Contouring and Gridding



discrete measurement

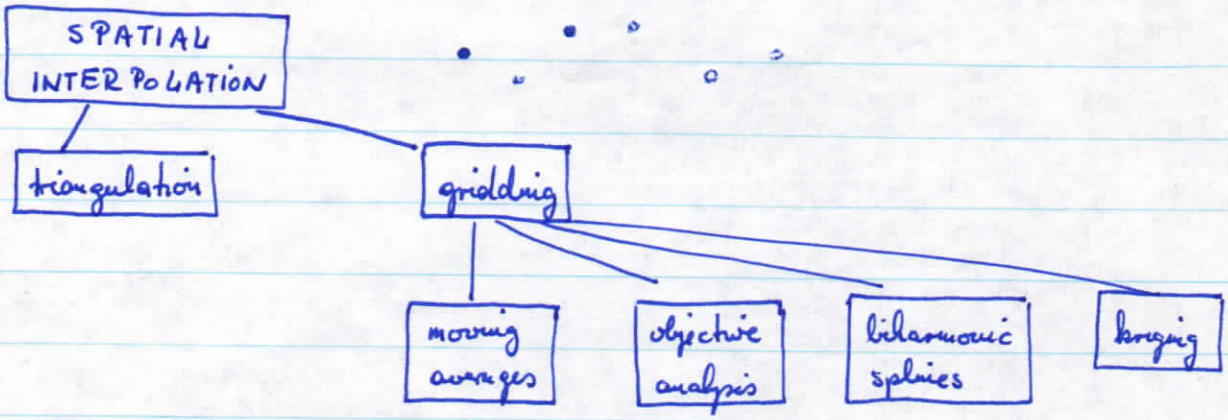
$z_i = z_i(x_i, y_i)$

continuous function

$z_i = z_i(x, y)$

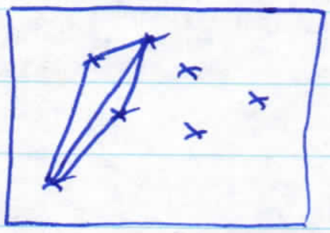
spatial interpolation: it's still more an art than a science even (or especially) if done by a computer

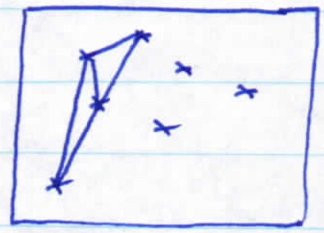
- always depict on the contour map where your measurements are located
- indicate what your measurement errors are
- indicate how the data are honored



Triangulation :

60 in
70 in





not uniquely defined
need excessive interactions to select an "optimal" set of triangles

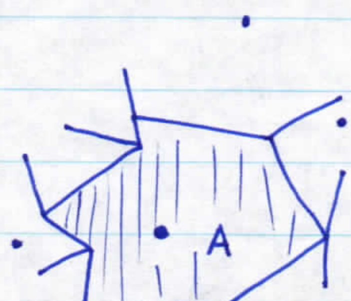
"optimal" : – near to equilateral as possible
– maximum height
– largest leg as short as possible

80 in Delaunay triangulations :

– uniquely defined
– largest sides as short as possible
– triangles are as equiangular as possible

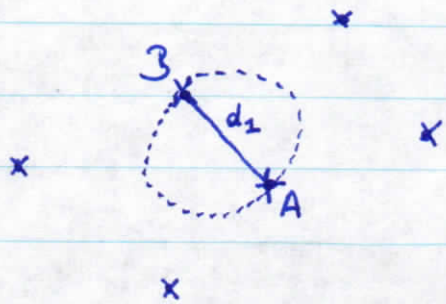
connects points in a triangular network that are Thiessen neighbors.

Thiessen polygon
≡
all points within are closer to A than to any other point



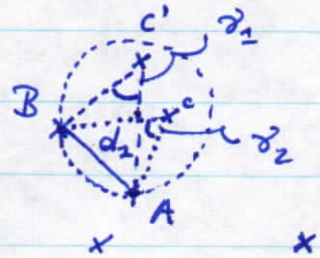
Determine the Thiessen neighbors

step 1:



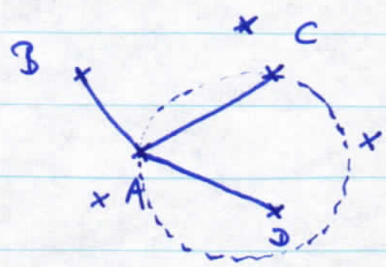
B is a neighbor if no pts are within the circle of diameter d_1

step 2: construct a circle with $d_2 > d_1$ ~~that has A and B on~~ that has A and B on its perimeter.

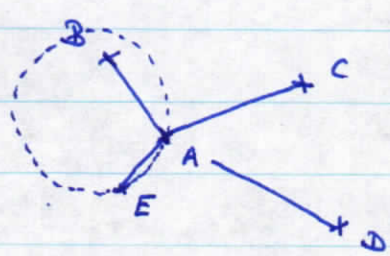


neighbor is the pt. with 'largest' γ_i

step 3 move on clockwise, i.e., construct successively larger circles that have A and the last neighbor on the perimeter



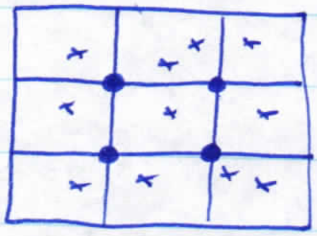
continue until you're back to pt. B as the closest neighbor



McCullagh, M.J. (1981). Creation of smooth contours over irregularly distributed local surface patches. Geographical Analysis, 13, 51-63.

Gridding

(a) Weighted (moving) averages



× measurements of $z_i(x_i, y_i)$

• grid nodes $(\hat{x}_k, \hat{y}_k)$

for each node k find a set of n nearest data to be used in forming an average

OR

for each node k find a set of data within a fixed radius (neighborhood) to form an average

average the found data and thus define $\hat{z}_k(\hat{x}_k, \hat{y}_k)$ at node

$$\hat{z}_k = \sum_{i=1}^{n} z_i$$

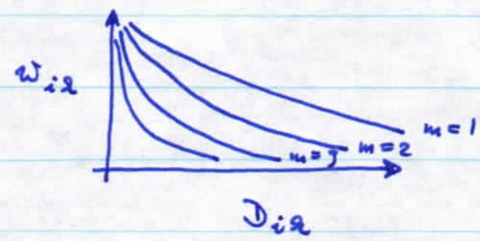
better yet, assign weights, e.g., weights proportional to the distance of the data at (x_i, y_i) to the grid location $(\hat{x}_k, \hat{y}_k)$

$$D_{ik} = \sqrt{(\hat{x}_k - x_i)^2 + (\hat{y}_k - y_i)^2}$$

and the weighted average is

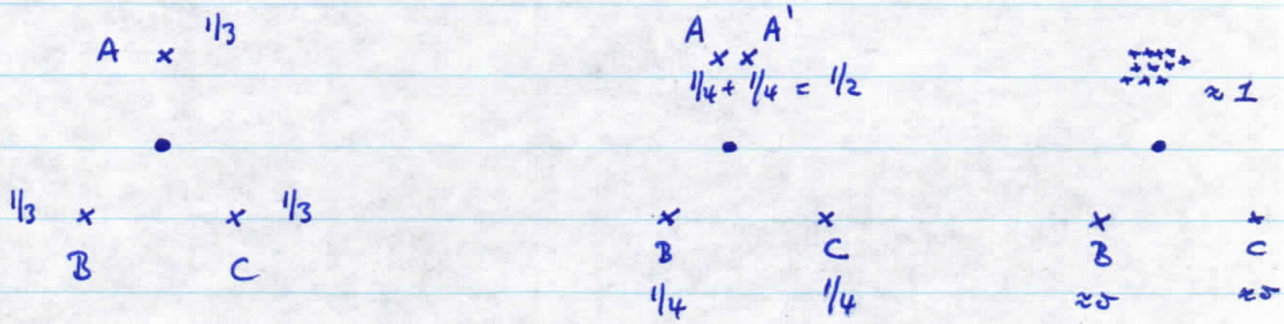
$$\hat{z}_k = \frac{\sum_{i=1}^{n} z_i / D_{ik}^{-m}}{\sum_{i=1}^{n} 1 / D_{ik}^{-m}} = \sum_{i=1}^{n} w_{ik} z_i \qquad w_{ik} = \frac{D_{ik}^{-m}}{\sum_{i=1}^{n} D_{ik}^{-m}}$$

where m is an arbitrary power



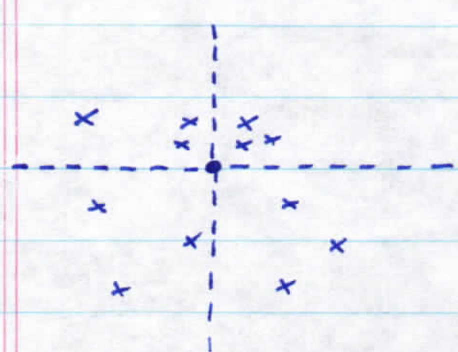
How to search for the n pts to be included in the interpolation?

1
three points are all equi= distance from node

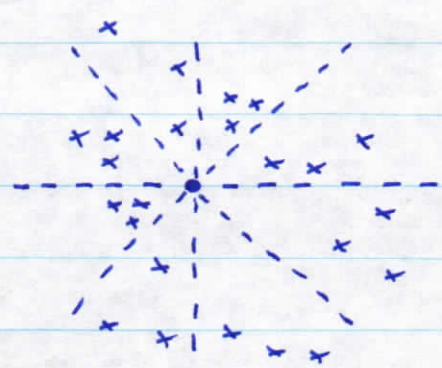


→ spatial arrangement important

→ search in angular segments, i.e.



quadrant search



octant search

Alternatives to weighted averaging

fit a least-square model locally, e.g.,

$$\hat{z}_k = a + b\,\hat{x}_k + c\,\hat{y}_k \qquad \text{linear}$$
$$+ d\,\hat{x}_k^2 + e\,\hat{y}_k^2 + f\,\hat{x}_k\,\hat{y}_k \qquad \text{quadratic}$$
$$\cdots \qquad \text{cubic}$$
$$\vdots$$

find coefficients, by minimizing

$$\varepsilon(a,b,\ldots) = \sum_{i=1}^{n} \left[z_i - (a + b x_i + c y_i + \ldots) \right]^2$$

This refered to as TREND SURFACE ANALYSIS

other "models" are possible, e.g.,

biharmonic splines (see notes on harmonic analysis in space)

which mathematically is equivalent to the constraint of finding a surface with MINIMUM CURVATURE over the domain

comment below

McIntosh, P.C. (1990). Oceanographic data interpolation: objective analysis and splines. J. Geophys. Res., 95, 13529–13541 biharmonic splines ⟷ minimum curvature

↕

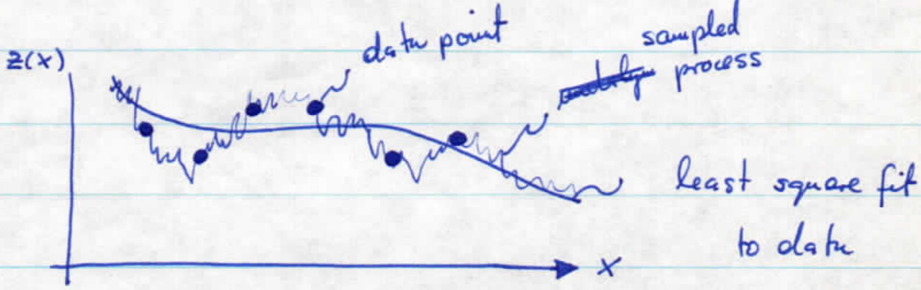
semi-norm interpolation $\|f(x,y)\| = \iint f_x^2 + f_y^2 + f_{xx}^2 + f_{xy}^2 + f_{yy}^2 \, dx\,dy$

↑

objective analysis (prescribe error and signal statistics) ⟷ norm interpolation $\|f(x,y)\| = \iint f^2 + f_x^2 + f_y^2 + f_{xx}^2 + f_{xy}^2 + f_{yy}^2 \, dx\,dy$

<u>COMMENT</u>: Note the complexity

1 D



the error at sampling points does NOT have to represent the error field correctly !