Spatial Interpolation

Lecture 10

Introduction

- **Definition:**

  "Spatial interpolation is the procedure of estimating the values of properties at unsampled sites within an area covered by existing observations." **(Waters, 1989)**

- **Application**
  - **wide range of applications**
  - **important in addressing problem of data availability**
  - **quick fix for partial data coverage**
  - **role of filling in the gaps between observations**

Interpolation: An essential skill

- Environmental data
  - often collected as discrete observations at points or along transects
  - example: soil properties, soil moisture, vegetation transects, pollution, ozone, groundwater characteristics, meteorological data, etc.
- Need to convert discrete data into continuous surface for use in hydrologic modelling

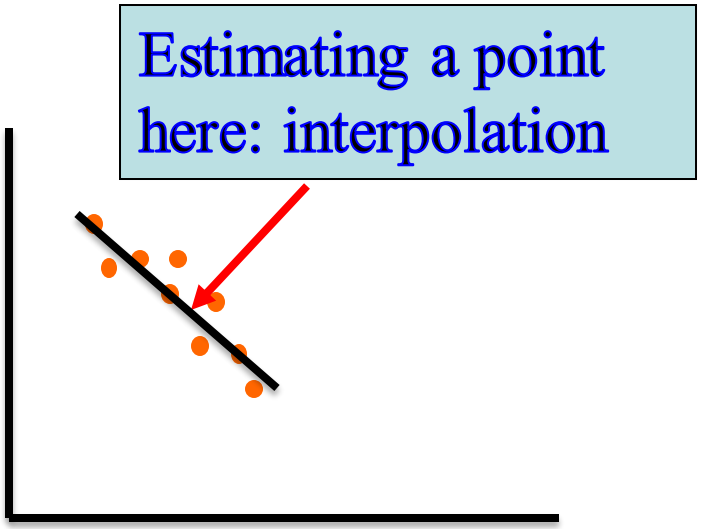
Interpolation vs Extrapolation

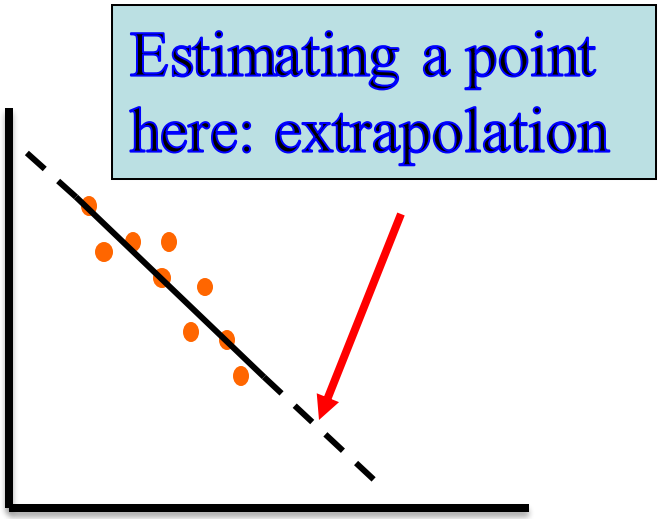
- Interpolation

Estimating the attribute values of locations that are within the range of available data using known data values

- Extrapolation

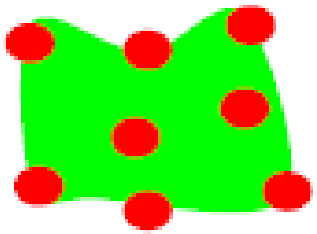
Estimating the attribute values of locations outside the range of available data using known data values



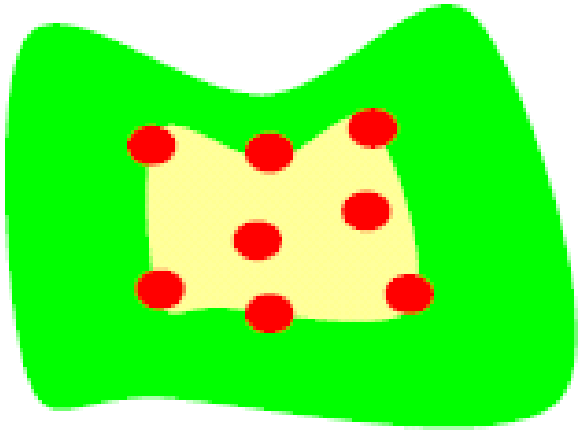


Interpolation vs Extrapolation

- Predicting the value of an attribute at sites outside the area is called "**extrapolation**".



Interpolation

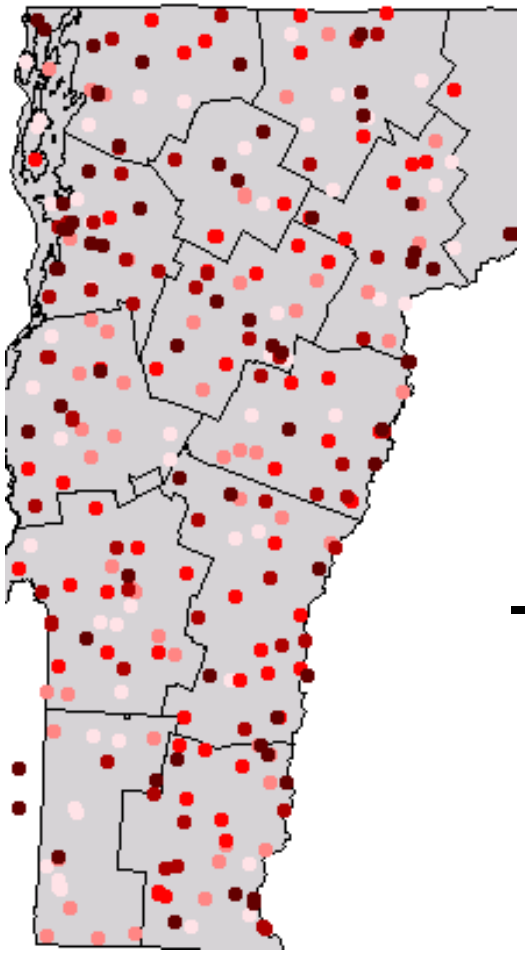


Extrapolation

Applications in Water Resources

- Precipitation map (hourly, daily, monthly, annual, etc.) and depth-area-duration curves
- Meteorological maps
- Soil property maps (eg. hydraulic conductivity, moisture, suction head, etc.)
- Elevation map (DEM)
- Groundwater characteristics (eg. depth to GW table)
- Measurement network design
- …….

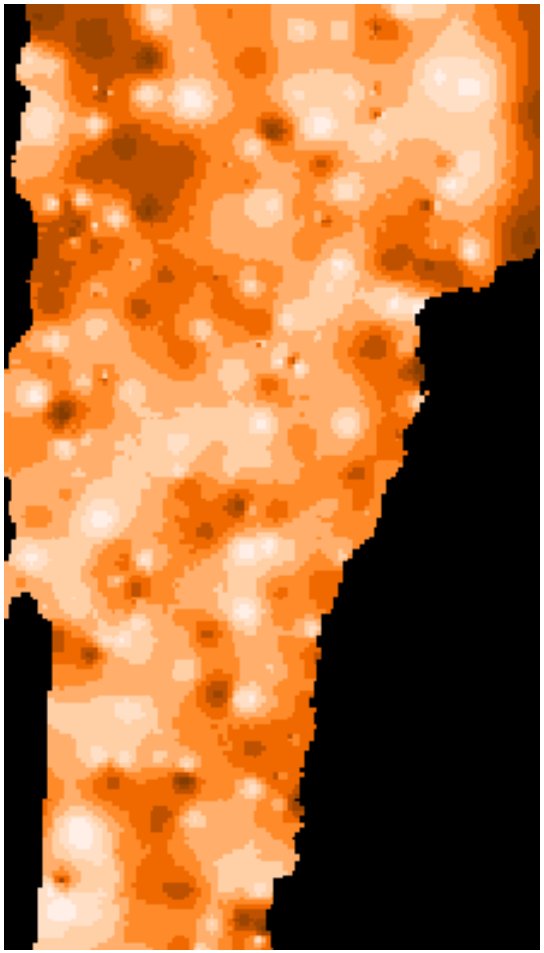
Ground water pollution



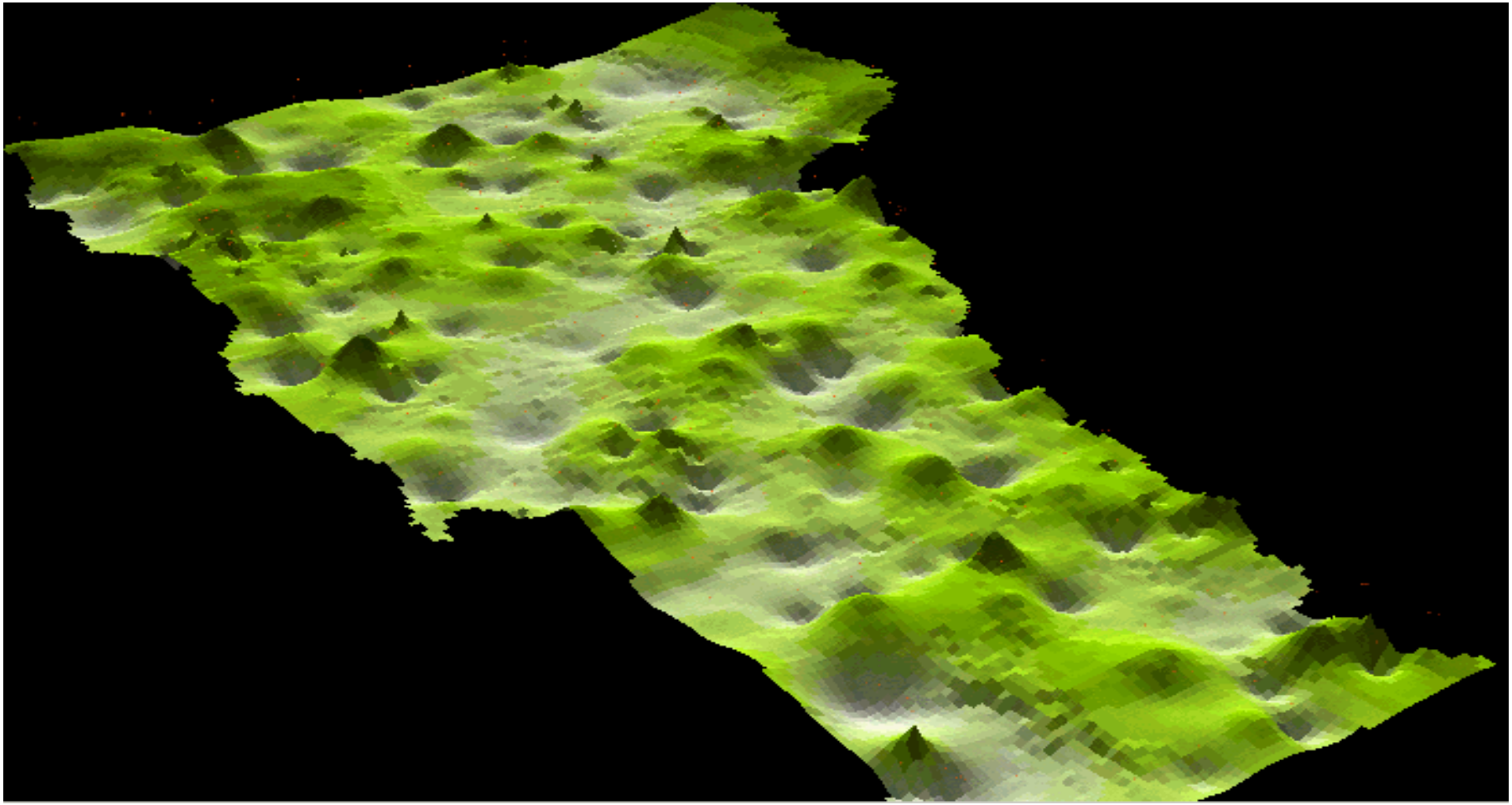
After

Interpolation

gives



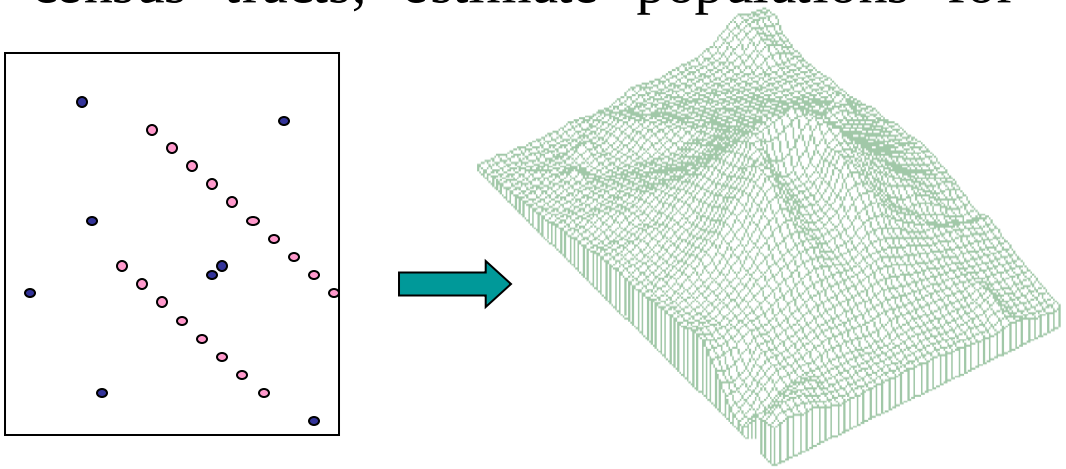
This can be displayed as a 3D surface



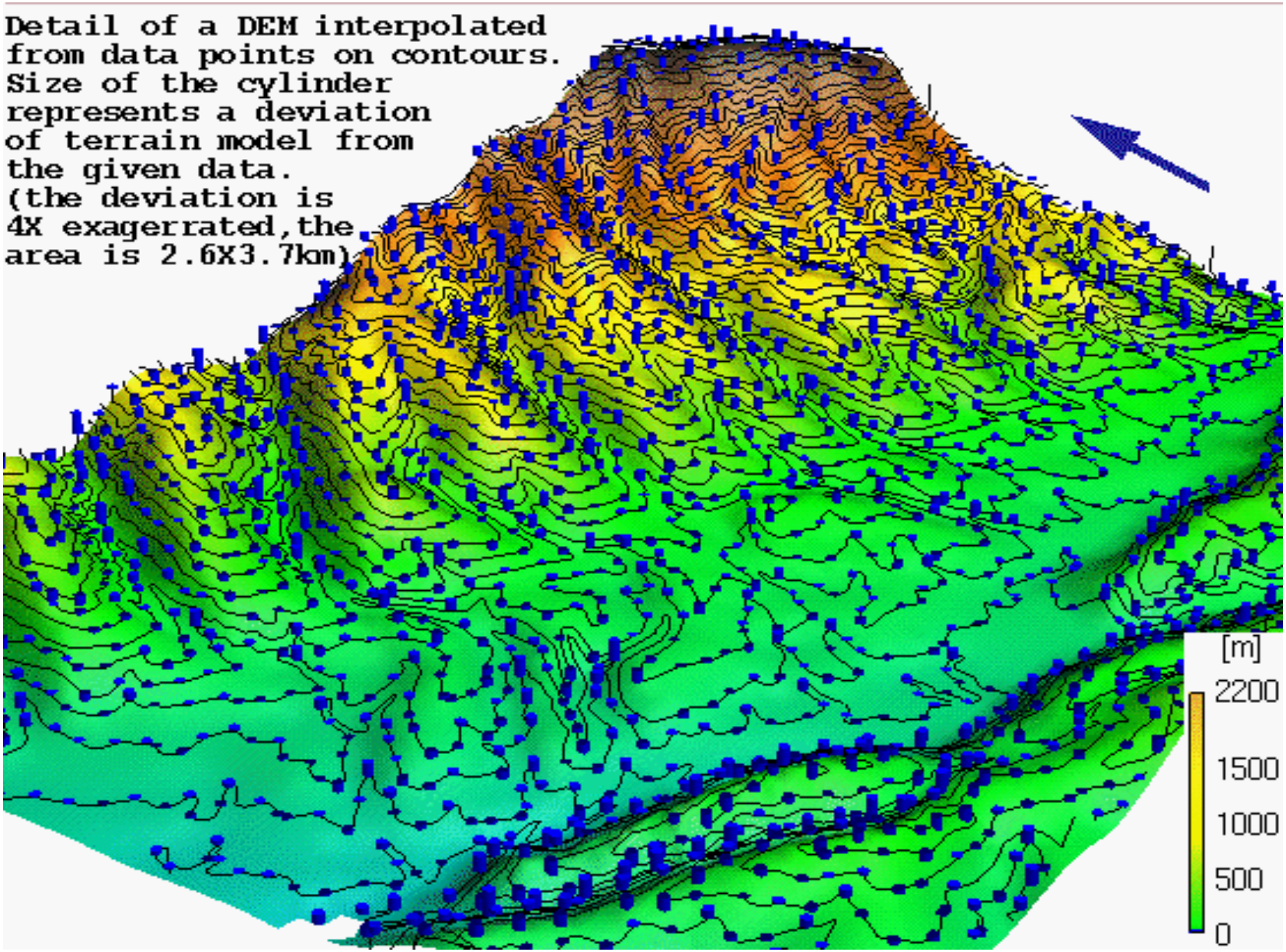
Forms of Interpolation

- Points to points (e.g. random points to regular grid)
- Points to lines (e.g. random points to contour lines)
- Lines to points (e.g. contours to a regular grid)
- Area to area (given a set of data mapped on one set of source zones, determine the values of the data for a different set of target zones, e.g. given population counts for census tracts, estimate populations for electoral districts)

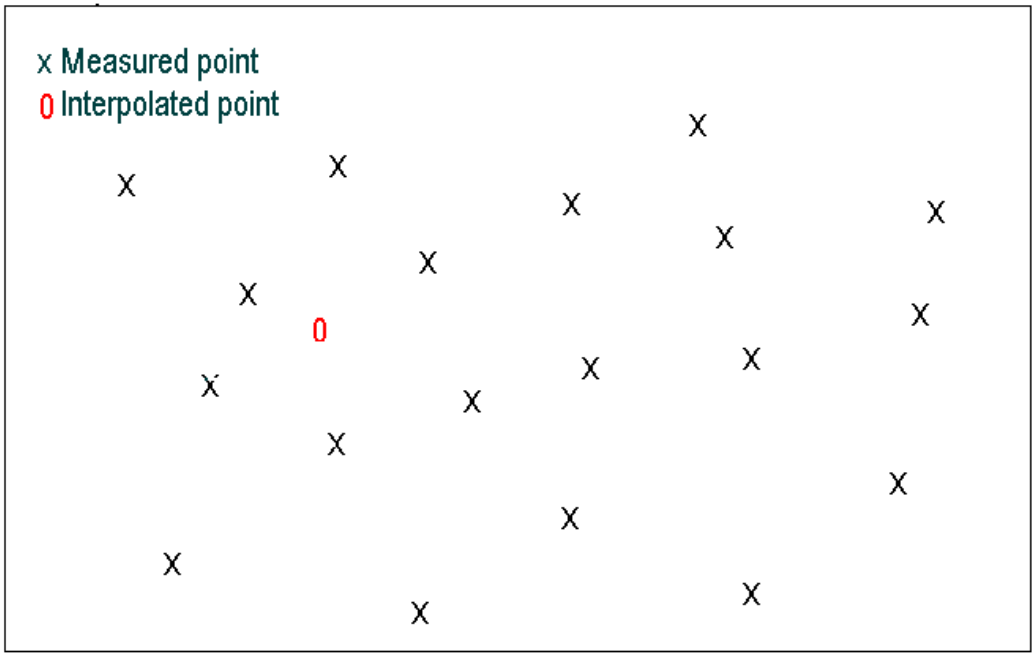
Grid surfaces from points

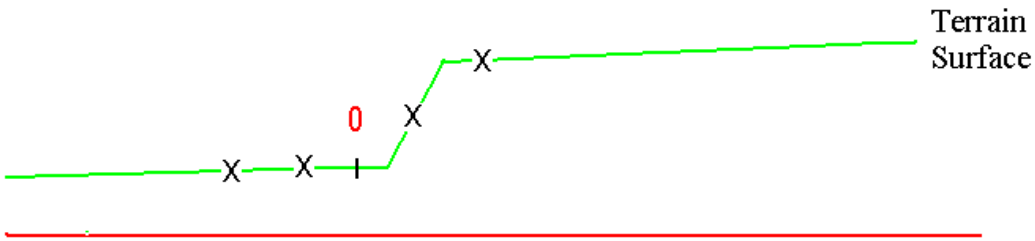


Points **Surface**


Detail of a DEM interpolated
from data points on contours.
Size of the cylinder
represents a deviation
of terrain model from
the given data.
(the deviation is
4X exagerrated,the
area is 2.6X3.7km)
[m]
2200
1500
1000
500
0

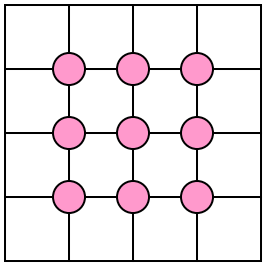
Interpolation



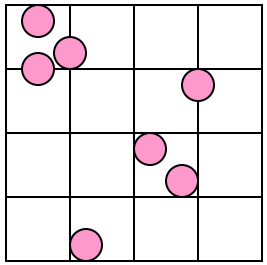


Data sampling

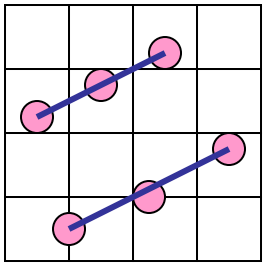
Method of sampling is critical for subsequent interpolation...



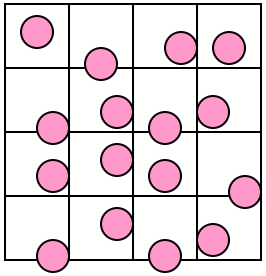
Regular



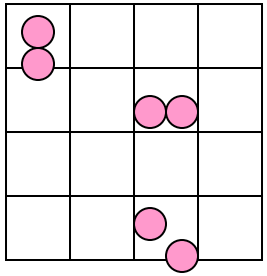
Random



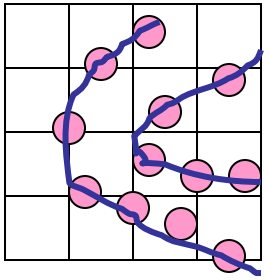
Transect



Stratified random



Cluster



Contour

Systematic sampling pattern

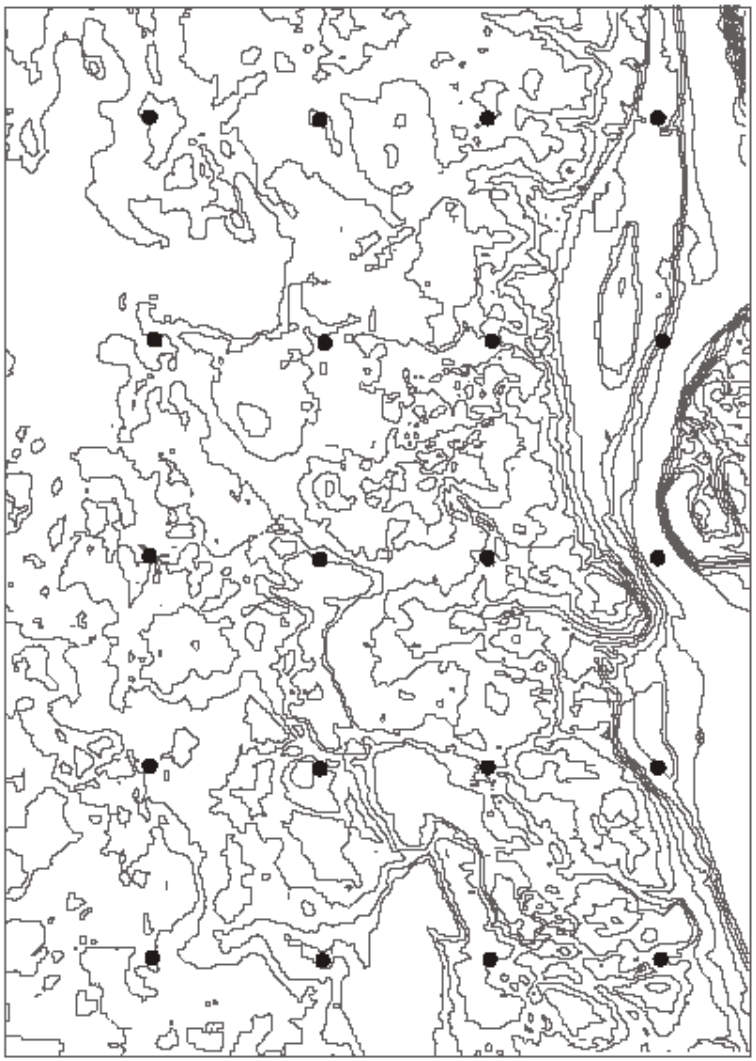
Easy Samples spaced uniformly at fixed X, Y intervals Parallel lines

Advantages

Easy to understand

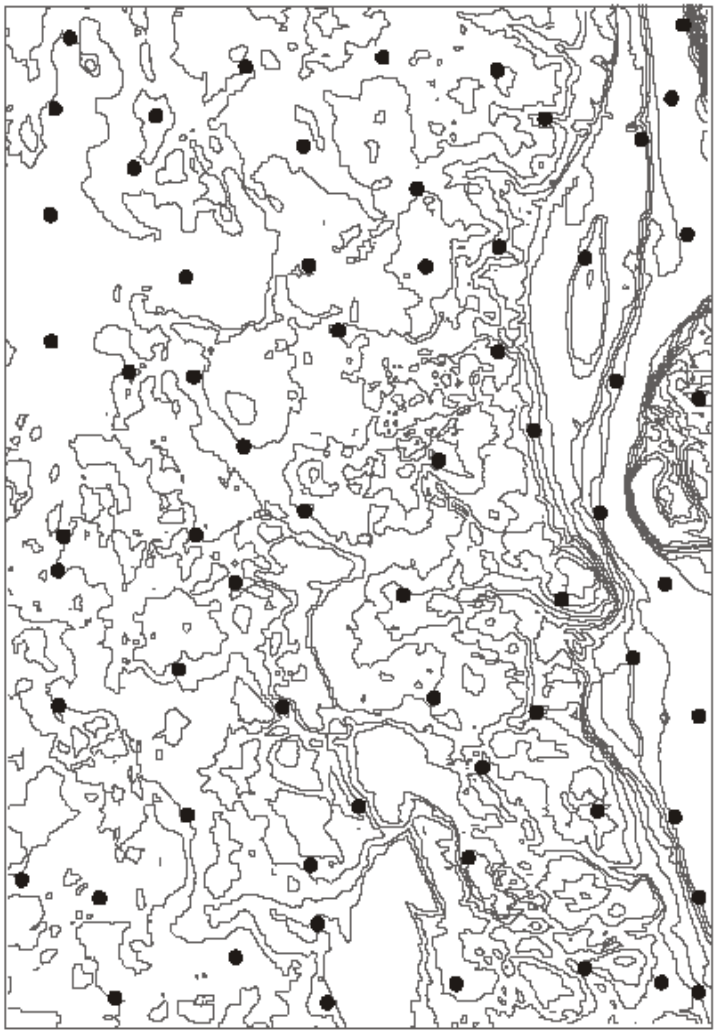
Disadvantages

All receive same attention
Difficult to stay on lines
May be biases



Random Sampling

Select point based on random number process.
Plot on map, visit sample



Advantages

Less biased

Disadvantages

Difficult to explain,
Location of points may be a problem

Cluster Sampling

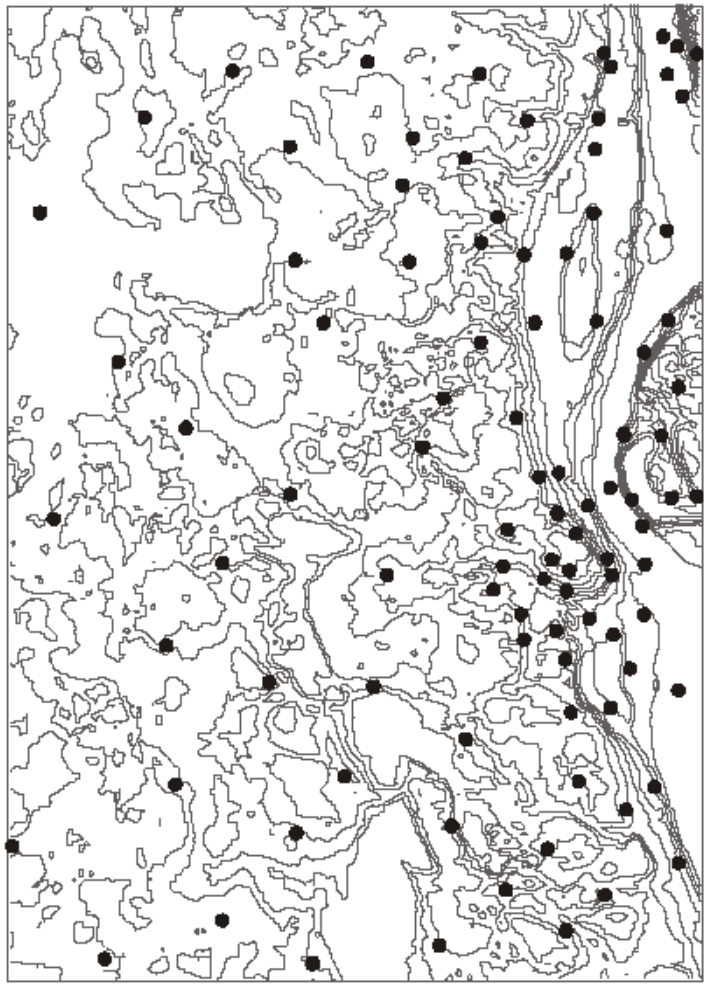
Cluster centers are established (random or systematic)
Samples arranged around each center

Advantages

Reduced travel time

Adaptive sampling

Higher density sampling where the feature of interest is more variable. Requires some method of estimating feature variation



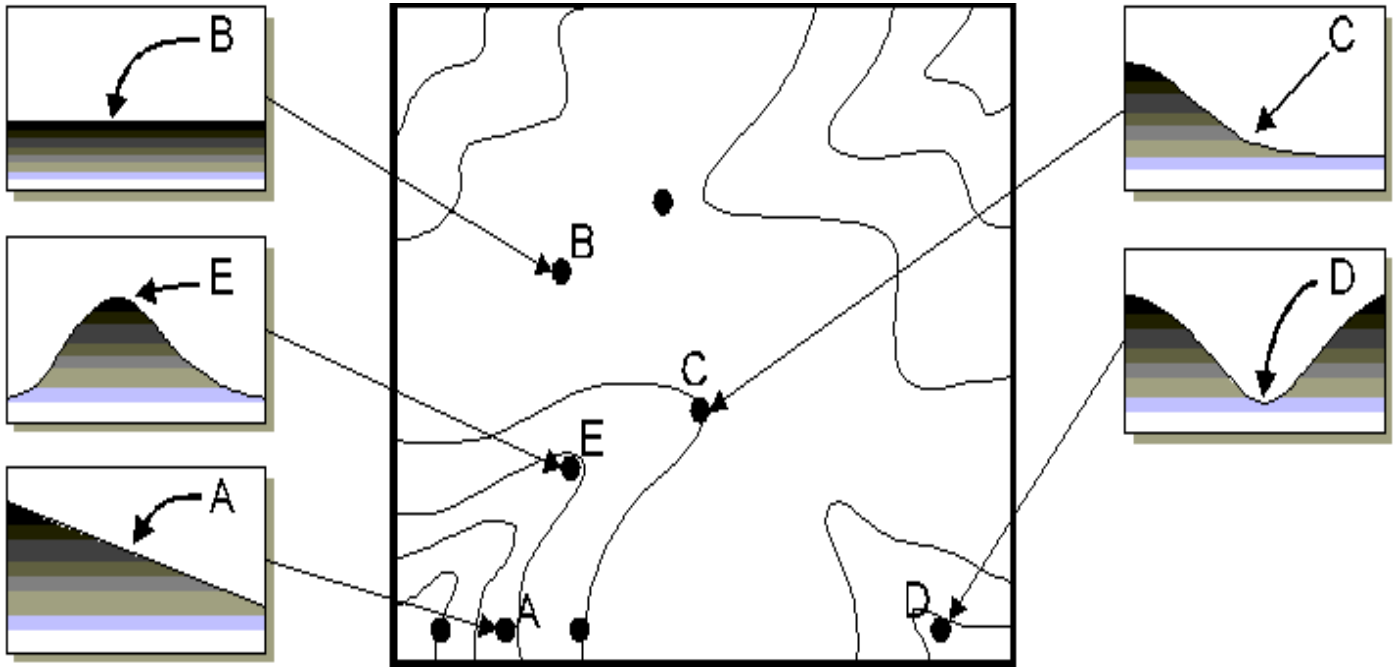
Advantages

Often efficient as large homogeneous areas have few samples reserving more for areas with higher spatial variation.

Disadvantages

If no method of identifying where features are most variable, then you need to make several sampling visits.

Very Important Points



Interpolation Types

- Correlation with another variable not discussed as interpolation methods (eg. rainfall=f(elevation))
- many different methods available
- classification according to:
  - exact or approximate
  - deterministic or stochastic
  - local or global
  - gradual or abrupt

Classification: global or local

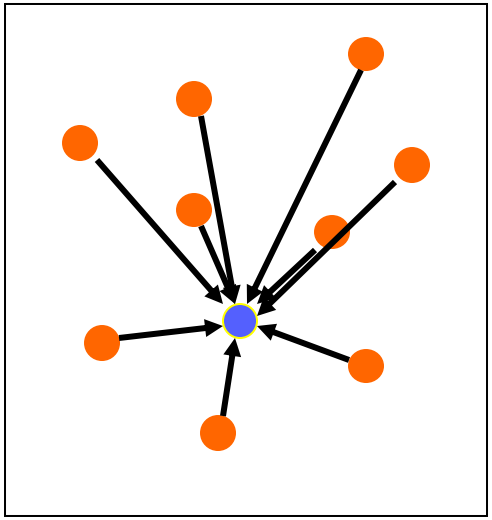
- Global methods:
  - single mathematical function applied to all points
  - tends to produce smooth surfaces
- Local methods:
  - single mathematical function applied repeatedly to subsets of the total observed points
  - link regional surfaces into composite surface

Classification: global or local

- global algorithms tend to produce smoother surfaces with less abrupt changes
- are used when there is an hypothesis about the form of the surface, e.g. a trend
- some local interpolators may be extended to include a large proportion of the data points in set, thus making them in a sense global
- the distinction between global and local interpolators is thus a continuum

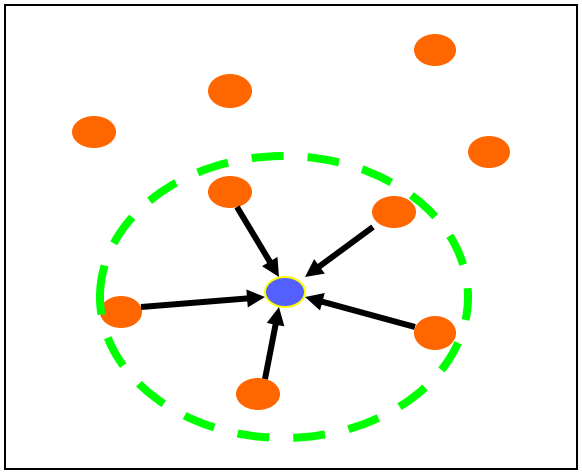
Global vs. Local

- Global



Sample data

- Local



Sample data

Classification: exact or approximate

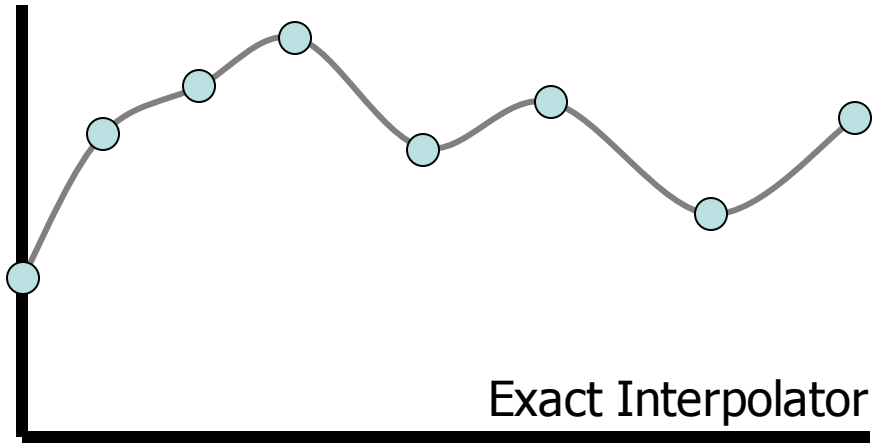
- Exact methods:
  - honor all data points such that the resulting surface passes exactly through all data points
  - appropriate for use with accurate data
- Approximate (inexact) methods:
  - do not honor all measured data points
  - more appropriate when there is high degree of uncertainty about data points

Classification: exact or approximate

- Kriging may incorporate a nugget effect and in that case, it becomes an inexact interpolator.
- Approximate methods look at the data sets to have a global trends, which vary slowly, overlain by local fluctuations, which vary rapidly and produce uncertainty (error) in the recorded values.
- The effect of smoothing will therefore be to reduce the effects of error on the resulting surface.

Inexact vs Exact Interpolators

Predicts value that is identical to sampled value

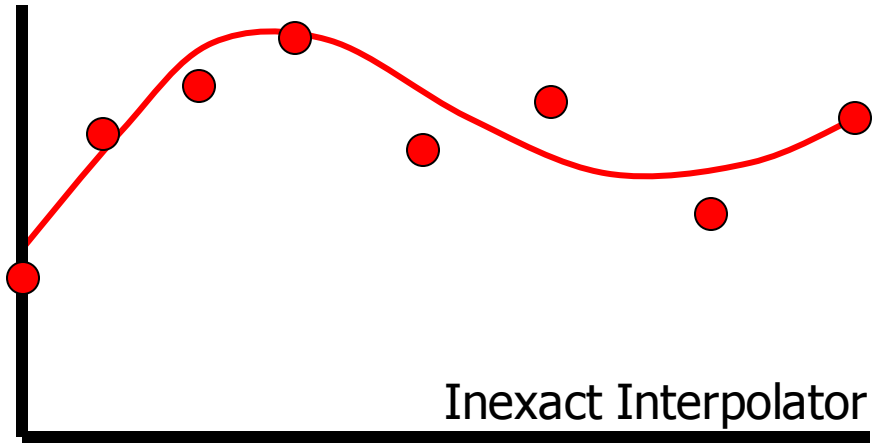


e.g.

Inverse Weighted Method

Radial Basis Function

Predicts a value that is different from the sampled value



e.g.

Global Polynomial Method

Local Polynomial Method

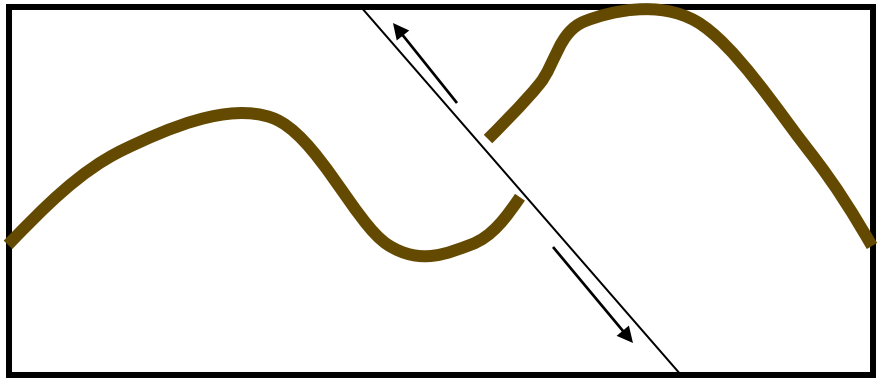
Classification: gradual or abrupt

- Gradual (smooth) methods:
  - produce smooth surface between data points
  - appropriate for interpolating data of low local variability

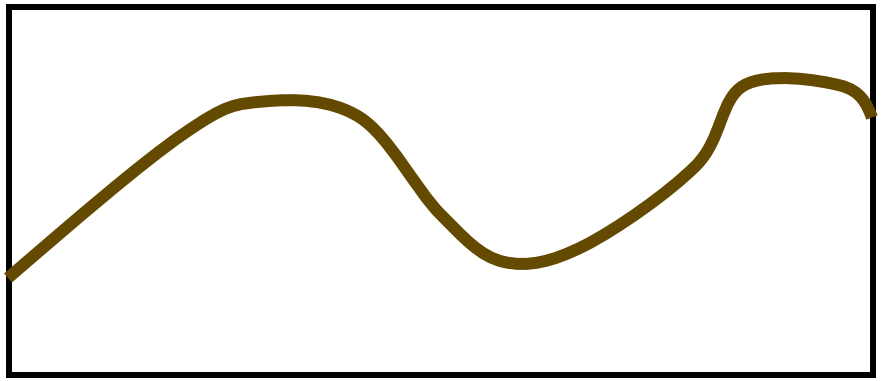
- Abrupt methods:
  - produce surfaces with a stepped appearance
  - appropriate for interpolating data of high local variability or data with discontinuities

Abrupt vs. Smooth

- Abrupt: interpolators that allow for barriers
  - Faults
  - Breakwalls
  - Atmospheric Fronts



- Smooth: interpolators that produce a smooth surface



Classification: deterministic or stochastic

- Deterministic methods:
  - used when there is sufficient knowledge about the surface being modelled
  - allows it to be modelled as a mathematical surface
  - methods do not use probability theory
- Stochastic methods:
  - used to incorporate random variation in the interpolated surface
  - the interpolated surface is conceptualized as one of many that might have been observed, all of which could have produced the known data points.
  - Geostatistical interpolation techniques (e.g., Kriging) utilize the statistical properties of the measured points
  - Will be subject of an independent future lecture

General Interpolation Equation

$$z_o' = \sum_{i=1}^{n} w_i * z_i \qquad \sum_{i=1}^{n} w_i = 1$$

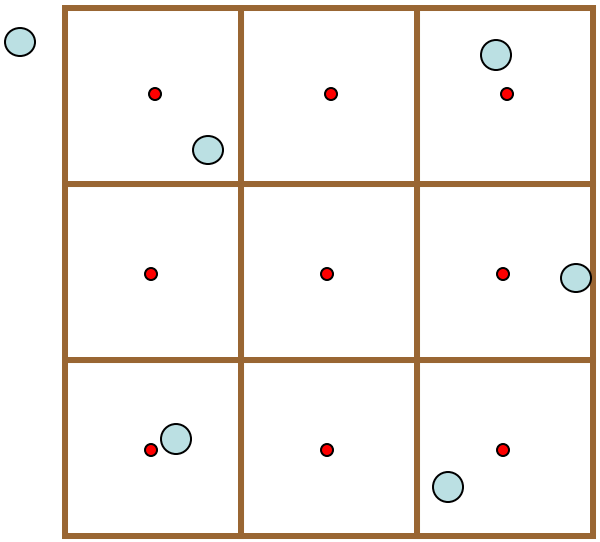
Z_0' is the attribute value to be predicted at unsampled site
Z_i is the attribute value at the i point of the nearby locations
w_i is the weight assigned to the attribute at point i, w_i should sum up to 1 (to be unbiased)
n is the total number of nearby locations involved

Key issues:

1) How many nearby points should we include for a given unsampled site?
2) How to select these nearby points?
3) How to allocate the weight for each nearby point?

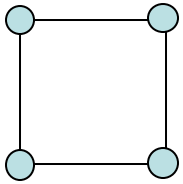
Interpolation

Estimate values between known values.



A set of spatial analyst functions that predict values for a surface from a limited number of sample points creating a continuous raster.

- Nearest neighbor
- Inverse distance weight $z = \sum \frac{1}{r_i} z_i$
- Bilinear interpolation $z = (a + bx)(c + dy)$
- Kriging (best linear unbiased estimator) $z = \sum w_i z_i$
- Spline $z = \sum c_i x^{e_i} y^{e_i}$



Usual Interpolation Questions

- What interpolation method should we use?
- How many samples should we include in the estimation of unsampled locations?
- How do we compensate for irregularly spaced or highly clustered sampling?
- How far should we go to include samples in our estimation process?
- Should we honor the sample values?
- How reliable is the estimate when we have it?
- Why is our map too smooth?
- What happens if there is a strong trend in the values?

Interpolation methods

NON-stochastic

Nearest Neighbors
Triangular Irregular Network
Inverse Distance Weighting
Regression Models (???)
Trend Surface
Splines
Thin Plate Splines
…….
……..
…….
……..
……

GEOSTATISTICS

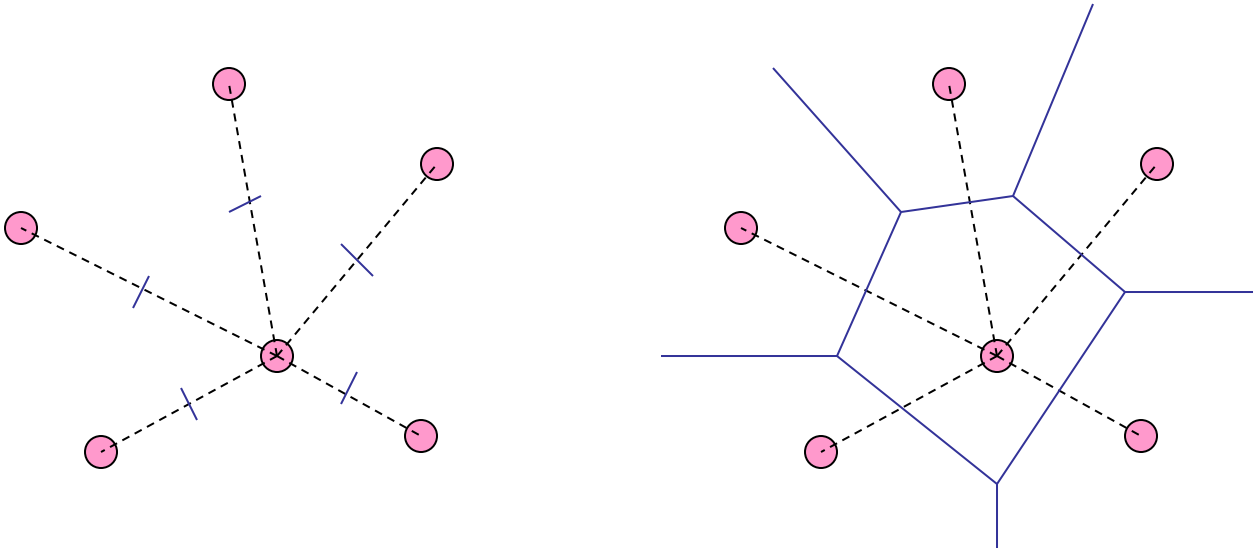
Geostatistics:
Kriging Estimator

1. Thiessen Polygons (Nearest Neighbors)

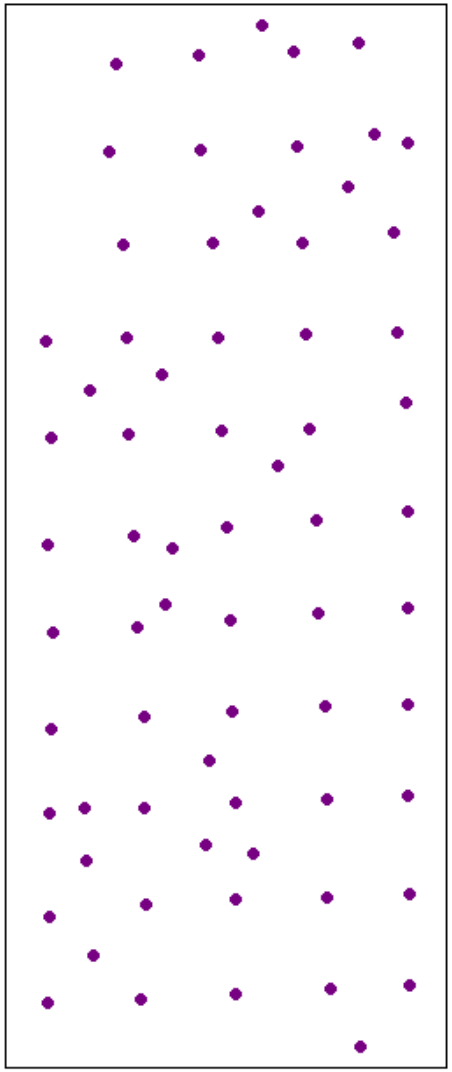
Thiessen polygons:

- assumes values of unsampled locations are equal to the value of the nearest sampled point
- regularly spaced points produces a regular mesh
- irregularly spaced points produces an network of irregular polygons
- could be performed in raster data model
- local, abrupt, exact, and deterministic

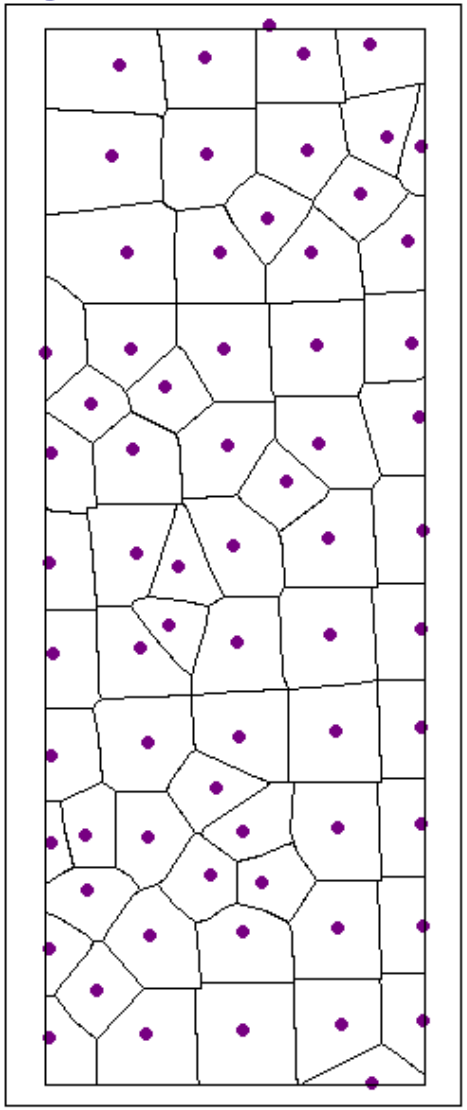
Thiessen polygon construction



Deterministic methods: Proximity Mapping



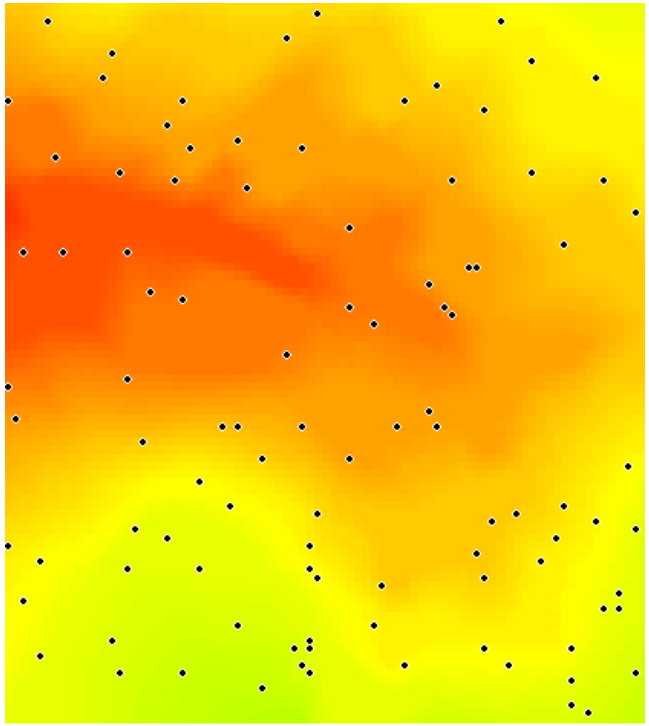
Thiessen Polygon or



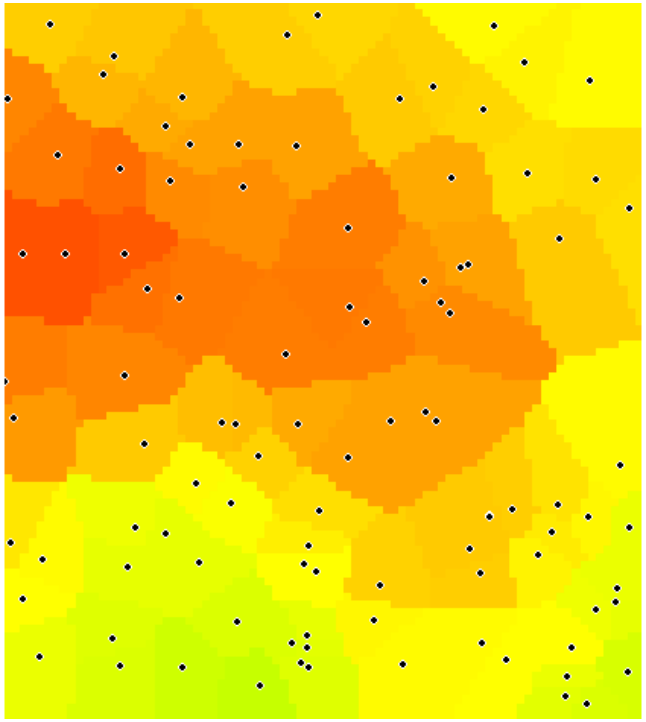
- constant value for each cell
- simple method (nearest neighbor for raster)
- jumps on the cell borders

Example Thiessen polygon

Source surface with sample points



Thiessen polygons with sample points

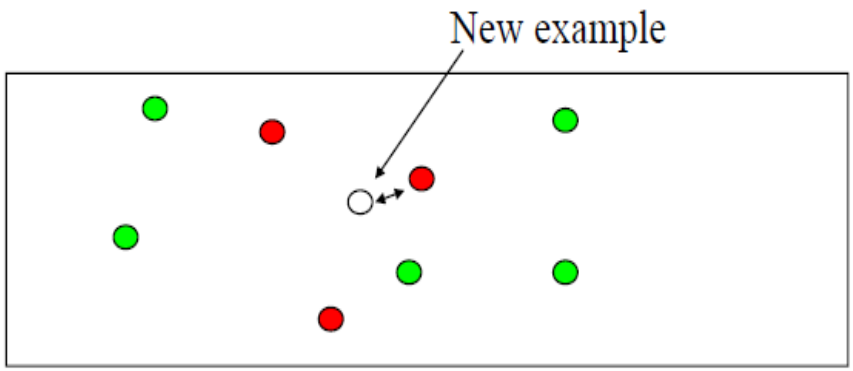


Nearest Neighbors (Thiessen Polygons)

- The nearest neighbors (NN) method predicts the value of an attribute at an unsampled point based on the value of the nearest sample by drawing perpendicular bisectors between sampled points (*n*), forming such as Thiessen (or Dirichlet/Voronoi) polygons. (*Vi*, *i*=1,2,…, *n*).
- This produces one polygon per sample and the sample is located in the center of the polygon, such that in each polygon all points are nearer to its enclosed sample point than to any other sample points (Isaaks and Srivastava, 1989; Ripley, 1981; Webster and Oliver, 2001).
- The estimations of the attribute at unsampled points within polygon *Vi* are the measured value at the nearest single sampled data point *xi* that is $\hat{z}(x0) = z(xi)$. The weights are:

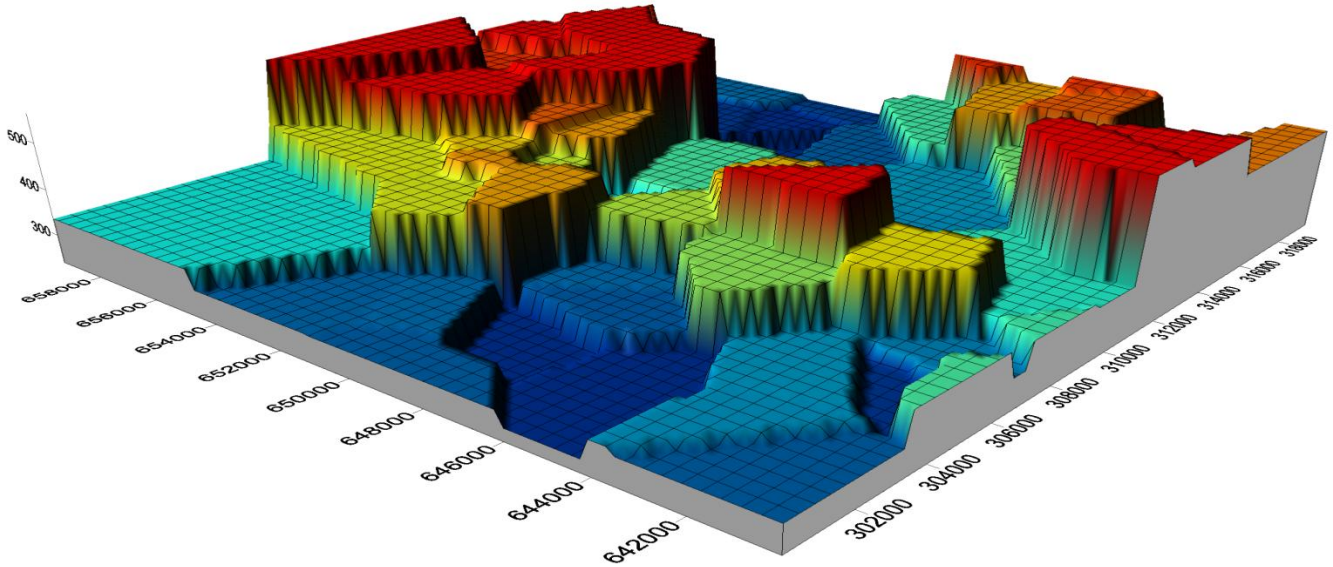
$$\lambda_i = \begin{cases} 1 & if \quad x_i \in V_i, \\ 0 & otherwise. \end{cases}$$

$$\left\| \mathbf{x} - \mathbf{x}^i \right\|^2 = \sum_j (x_j - x_j^i)^2$$



Nearest Neighbors

- Deterministic interpolation: Nearest neighbour
  - Simple assignment on nearest neighbour basis
  - Uniform within zones, stepped between zones



Notes on Nearest Neighbor

- Use if there is a direct interaction between features
- Good for data analyzed along a line (ie plants or wildlife observations collected along a transect)
- Stat is calculated using the distance between features

مثال: روش Thiessen

Point1	Point2	Point3	Point4	Point5	Point6	Point7	Point8	Point9
Point10	Point11	Point12	Point13	**190**	Point15	Point16	Point17	Point18
Point19	Point20	Point21	Point22	Point23	Point24	Point25	Point26	Point27
Point28	**140**	Point30	Point31	Point32	Point33	Point34	**230**	Point36
Point37	Point38	Point39	Point40	Point41	Point42	Point43	Point44	Point45
Point46	Point47	Point48	Point49	Point50	Point51	Point52	Point53	Point54
90	Point56	Point57	Point58	Point59	Point60	Point61	Point62	Point63
Point64	Point65	Point66	Point67	Point68	Point69	Point70	Point71	Point72
Point73	Point74	Point75	Point76	Point77	Point78	**120**	Point80	Point81

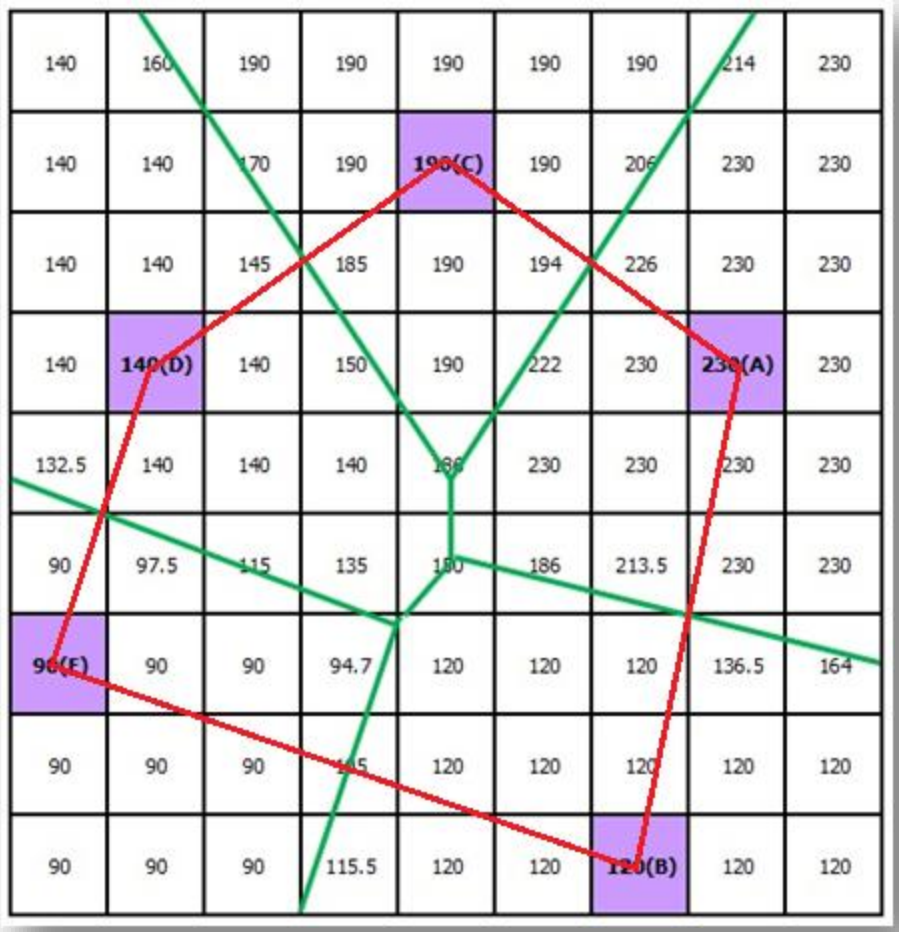
مثال: عمق بارندگی را در کلیه سلول‌ها به روش Thiessen تعیین نمائید.

(اندازه سلول‌ها اهمیتی ندارد!)

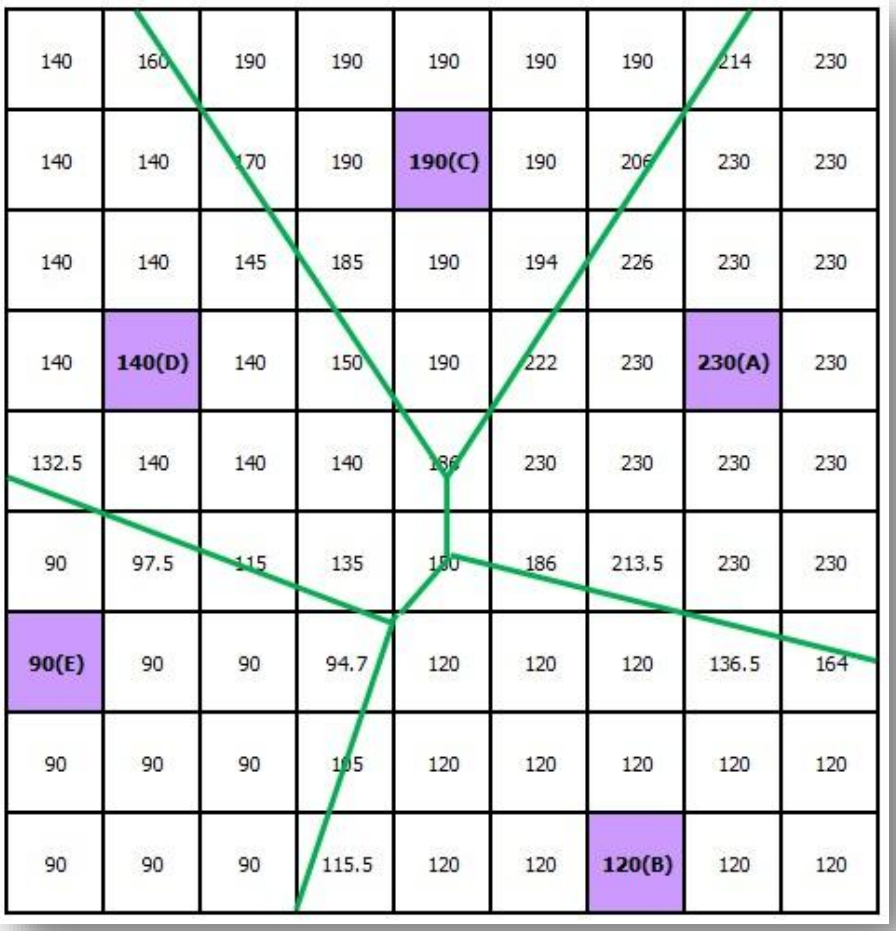
مثال روش Thiessen

حل : در ابتدا ایستگاه به یکدیگر متصل شده و از اتصال این نقـاط مثلـث‌هـایی پدیـد می‌آید.

سپس با ترسیم عمود منصف اضلاع مثلث‌ها، چندضلعی‌هایی با مرکزیت ایستگاه‌ها ایجاد می‌شود.



مثال روش Thiessen



حل: مقدار کمیت سلول‌های قرار گرفته در هر چندضلعی، برابر با مقدار ایستگاه مرکزی تعیین می شود.

2. Triangulation

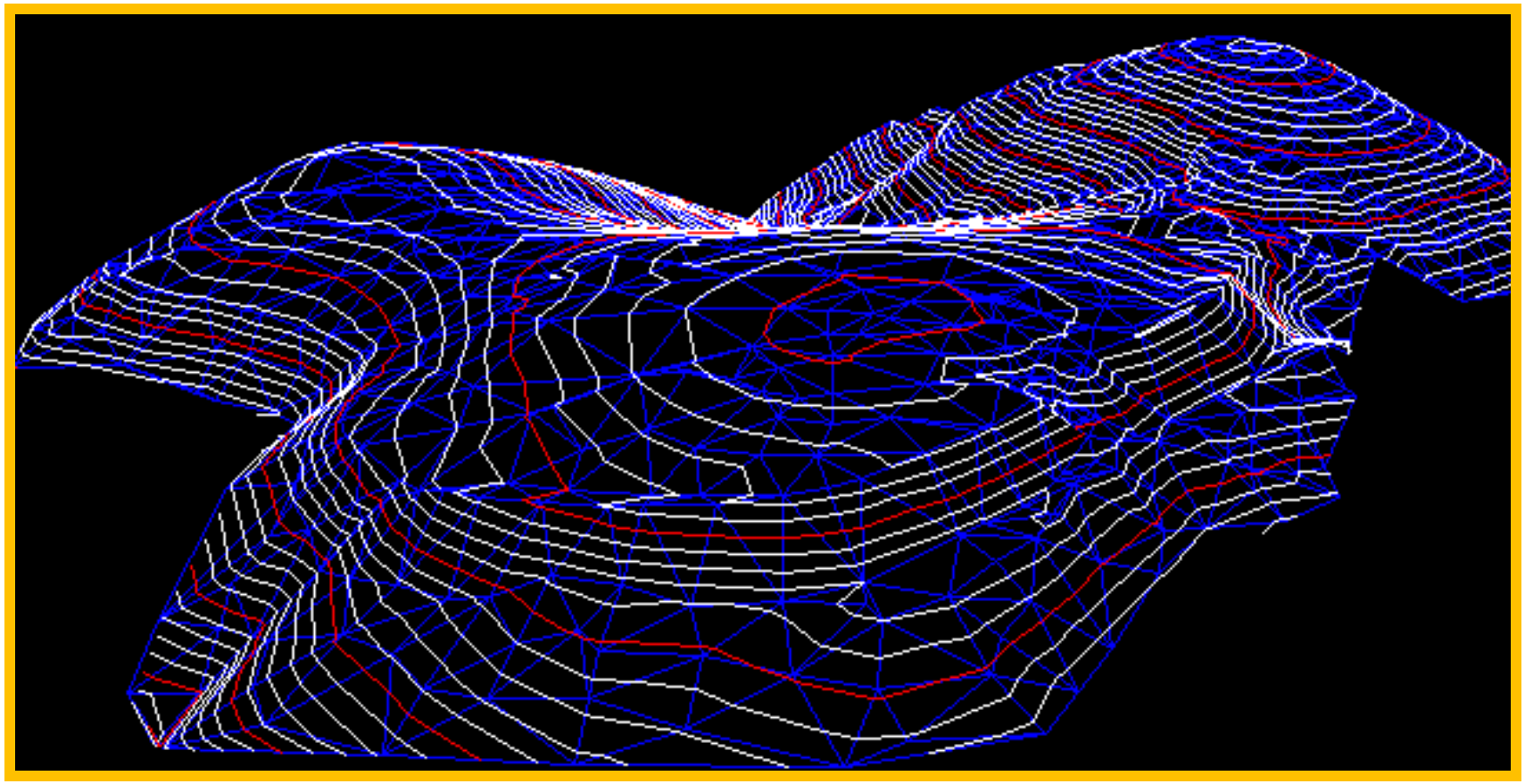
- A vector-based TIN model
- adjacent data points connected by lines to create a network of irregular triangles
- calculate real 3D distance between data points along vertices using trigonometry
- calculate interpolated value along facets between three vertices
- local, exact, and deterministic

روش مثلث‌بندی

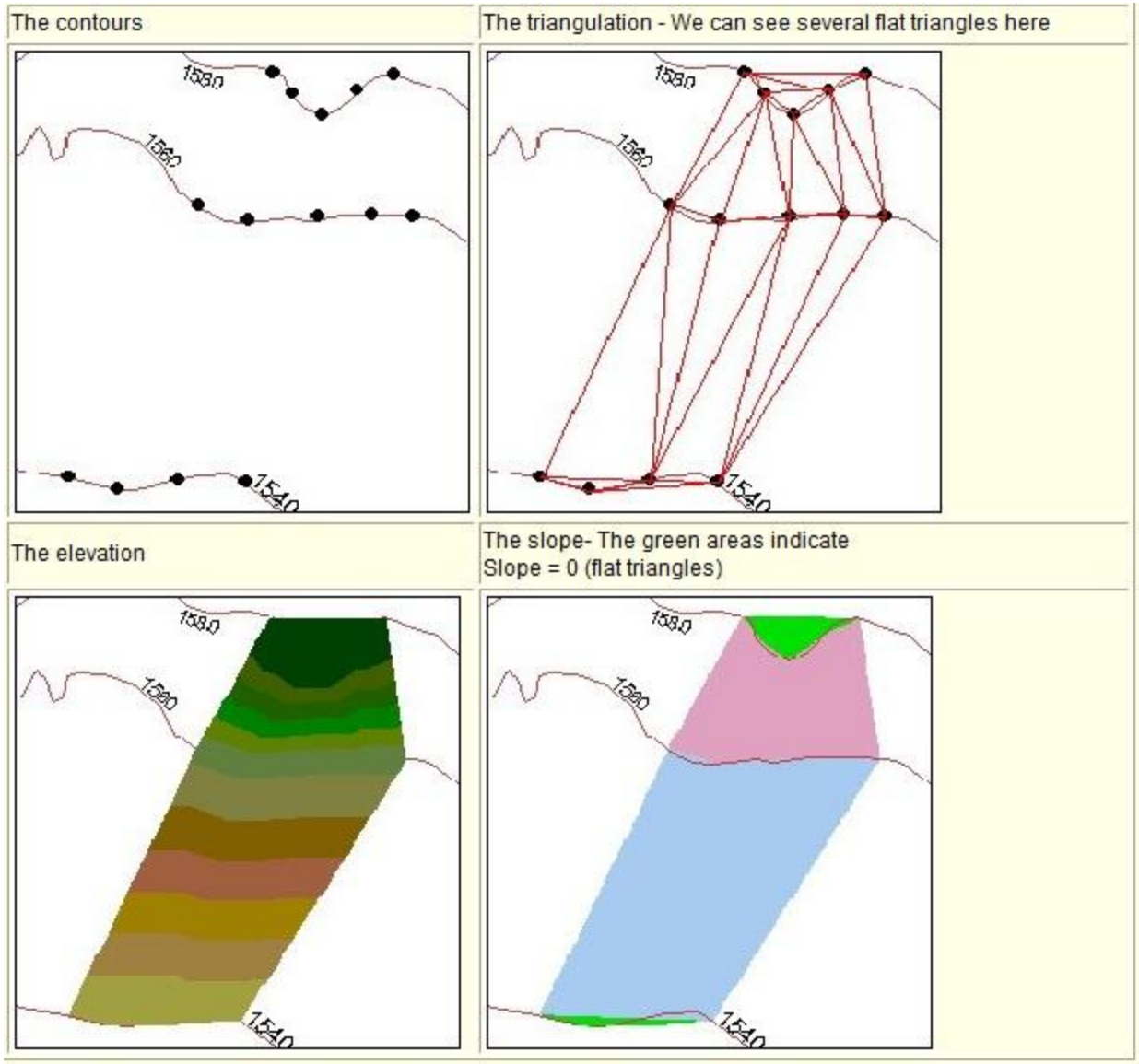
- در این روش، با اتصال نقاط اندازه‌گیری مجموعه ای از مثلث‌ها حاصل می‌شود.
- از اطلاعات رئوس هر مثلث، رابطه صفحه مثلث بدست می‌آید.
- با حل دستگاه سه معادله و سه مجهول، ثابت‌های رابطه تعیین می‌شود.
- با استفاده از رابطه هر مثلث، مقادیر کلیه نقاط درون مثلث قابل درون‌یابی است.

روش مثلث‌بندی

نمونه‌ای از روش مثلث‌بندی در فضای سه بعدی



روش مثلث‌بندی

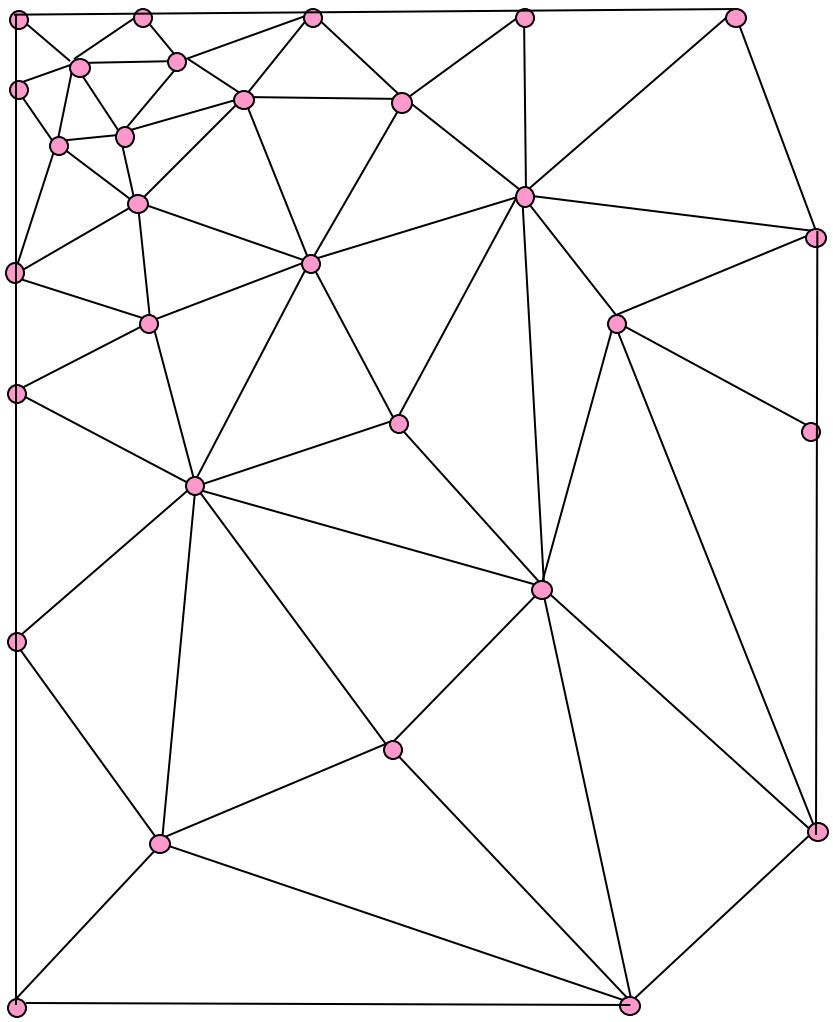


روش مثلث‌بندی

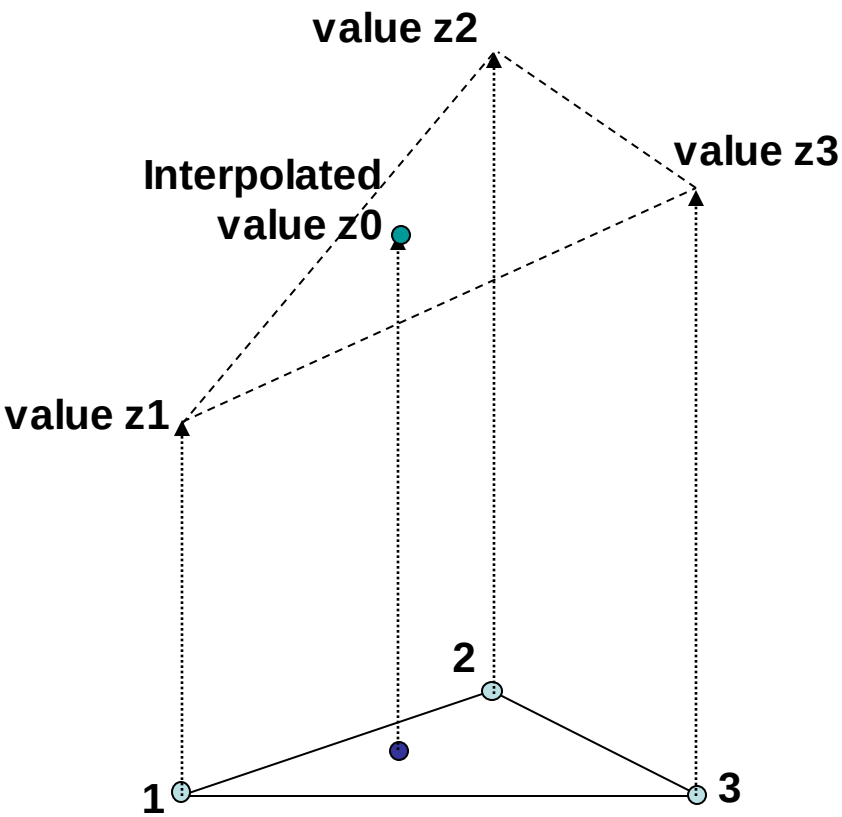
در صفحه دو

بعدی کانتورها

TIN construction

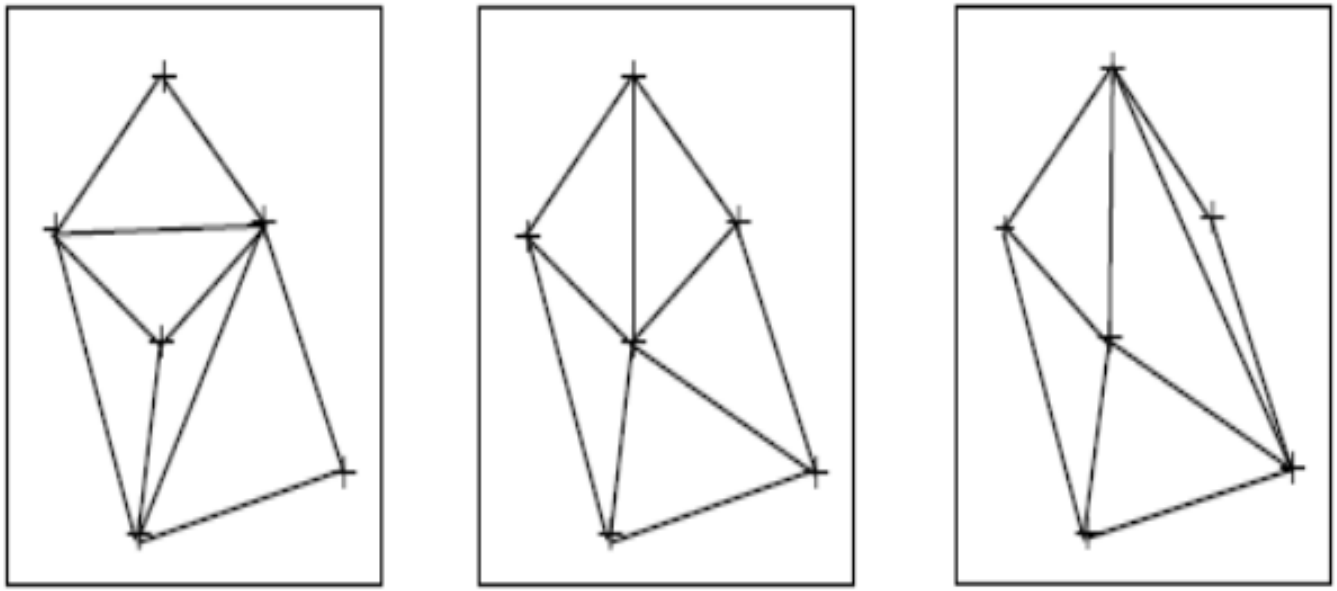


Plan view

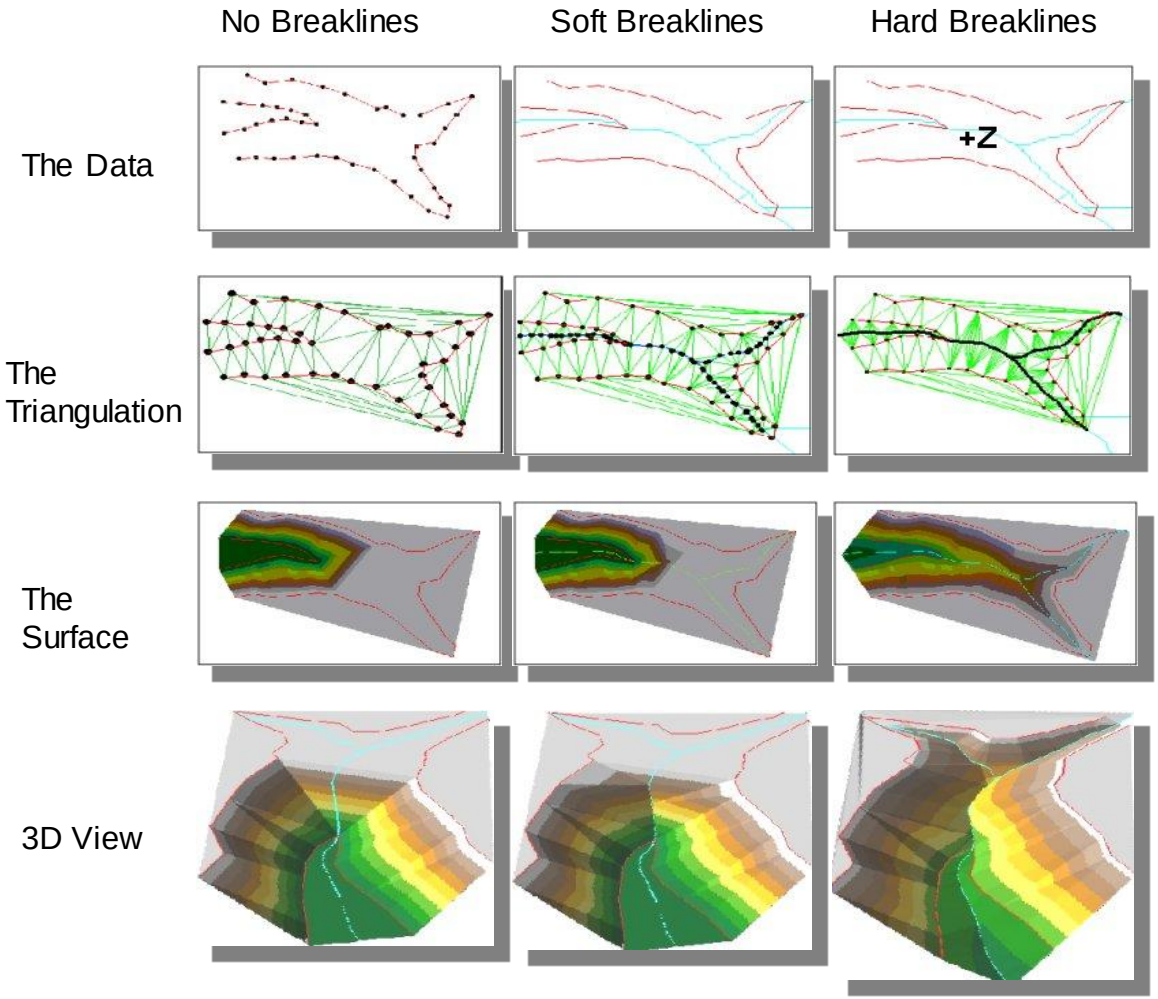


Isometric view

Different triangulation network using same data points



Creating a TIN



Break lines
Linear features which define and control surface behavior in terms of smoothness and continuity.

TIN Equations

$$z = ax + by + c$$

General Plane Equation

$$z_1 = ax_1 + by_1 + c$$

$$z_2 = ax_2 + by_2 + c$$

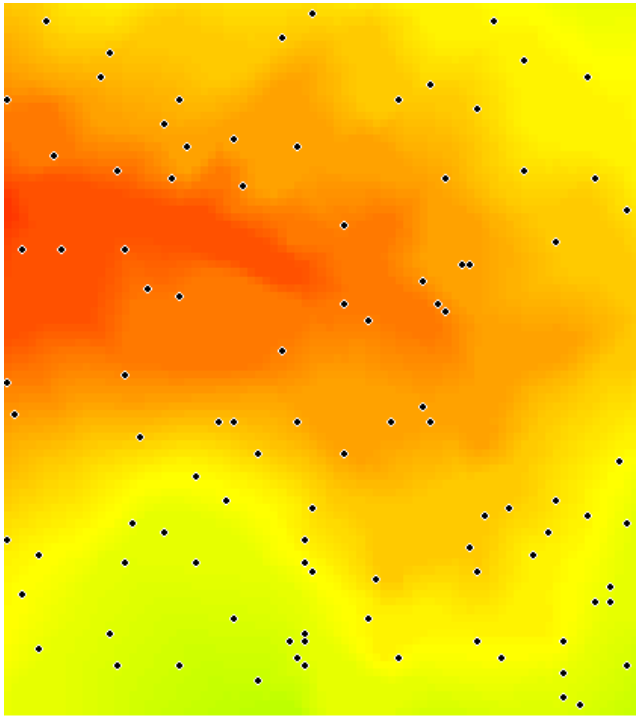
$$z_3 = ax_3 + by_3 + c$$

$$z_0 = ax_0 + by_0 + c$$

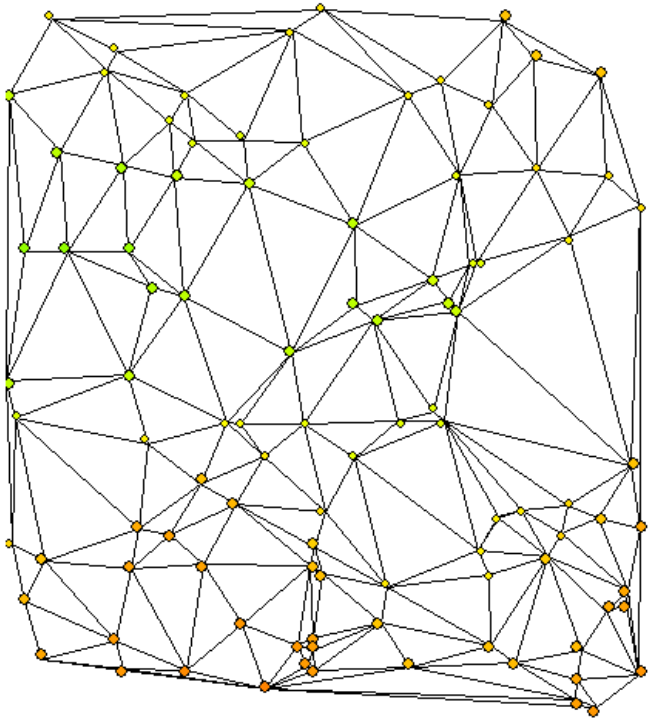
Estimation Equation

Example TIN

Source surface with sample points



Resulting TIN



مثال روش مثلث‌بندی

مثال: عمق بارندگی را با استفاده از موقعیت و مقادیر در شبکه ایستگاه‌های باران‌سنجی زیر، به روش مثلث‌بندی تعیین کنید.

ابعاد سلول‌ها را یک کیلومتر در نظر بگیرید.

Annual Rainfall Depth (mm)								
Point1	Point2	Point3	Point4	Point5	Point6	Point7	Point8	Point9
Point10	Point11	Point12	Point13	**190(C)**	Point15	Point16	Point17	Point18
Point19	Point20	Point21	Point22	Point23	Point24	Point25	Point26	Point27
Point28	**140(D)**	Point30	Point31	Point32	Point33	Point34	**230(A)**	Point36
Point37	Point38	Point39	Point40	Point41	Point42	Point43	Point44	Point45
Point46	Point47	Point48	Point49	Point50	Point51	Point52	Point53	Point54
90(E)	Point56	Point57	Point58	Point59	Point60	Point61	Point62	Point63
Point64	Point65	Point66	Point67	Point68	Point69	Point70	Point71	Point72
Point73	Point74	Point75	Point76	Point77	Point78	**120(B)**	Point80	Point81

مثال روش مثلث‌بندی

حل : نقاط ایستگاه‌های اندازه‌گیری را به یکدیگر متصل کرده که در مجموع ۳ مثلث حاصل می‌شود. این روش برون‌یابی را انجام نمی‌دهد. به همین جهت، درون‌یابی فقط در سلول‌های داخل مثلث‌ها انجام خواهد گرفت.

Annual Rainfall Depth (mm)								
Point1	Point2	Point3	Point4	Point5	Point6	Point7	Point8	Point9
Point10	Point11	Point12	Point13	190(C)	Point15	Point16	Point17	Point18
Point19	Point20	Point21	Point22	Point23	Point24	Point25	Point26	Point27
Point28	140(D)	Point30	Point31	Point32	Point33	Point34	230(A)	Point36
Point37	Point38	Point39	Point40	Point41	Point42	Point43	Point44	Point45
Point46	Point47	Point48	Point49	Point50	Point51	Point52	Point53	Point54
90(E)	Point56	Point57	Point58	Point59	Point60	Point61	Point62	Point63
Point64	Point65	Point66	Point67	Point68	Point69	Point70	Point71	Point72
Point73	Point74	Point75	Point76	Point77	Point78	120(B)	Point80	Point81
Pixel size = 1 km								

1 2 3

مثال روش مثلث‌بندی

حل : فرض کنید که رابطه صفحه مثلث به شکل $Zi=aXi+bYi+c$ باشد. در این رابطه داریم:

i : شماره هر راس مثلث (شامل ۳ راس)

Zi : بارندگی در راس i

Xi : مختصات راستای X

Yi : مختصات راستای Y

مثال روش مثلث‌بندی

حل:

- ۳ ضریب این رابطه (a، b و c) به عنوان مجهولات هر معادله درنظر گرفته شده‌اند.
- با حل ۳ معادله و ۳ مجهول به ازای هر مثلث، مقادیر a، b و c تعیین می‌شود.
- با استفاده از رابطه Z0=aX0+bY0+c و قرار دادن مقادیر a، b و c (که در مرحله قبلی تعیین شد) و مختصات نقطه یا سلول مورد نظر در داخل هر مثلث (Y0,X0)، مقدار Z0 تعیین می گردد.

مثال روش مثلث‌بندی

حل: در زیر، مشخصات مثلث‌ها شامل اطلاعات سه راس مثلث و دستگاه معادلات و جواب آن در هر مثلث، نشان داده شده است.

		Z	X	Y
مثلث ۱	نقطه ۱	90	0.5	2.5
	نقطه ۲	120	6.5	0.5
	نقطه ۳	140	1.5	5.5
مثلث ۲	نقطه ۱	140	1.5	5.5
	نقطه ۲	120	6.5	0.5
	نقطه ۳	190	4.5	7.5
مثلث ۳	نقطه ۱	190	4.5	7.5
	نقطه ۲	120	6.5	0.5
	نقطه ۳	230	7.5	5.5

دستگاه معادلات	جواب دستگاه معادلات
Z1=aX1+bY1+c Z2=aX2+bY2+c Z3=aX3+bY3+c	a=9.5 b=13.5 c=51.5
Z1=aX1+bY1+c Z2=aX2+bY2+c Z3=aX3+bY3+c	a=8.4 b=12.4 c=59.2
Z1=aX1+bY1+c Z2=aX2+bY2+c Z3=aX3+bY3+c	a=24.7059 b=17.0588 c= - 49.1176

مثال روش مثلث‌بندی

حل: رابطه مربوط به پیکسل‌های هر مثلث به صورت زیر خواهد بود:

دستگاه معادلات	جواب دستگاه معادلات			رابطه مربوط به پیکسلهای هر مثلث
Z1=aX1+bY1+c Z2=aX2+bY2+c Z3=aX3+bY3+c		a=9.5 b=13.5 c=51.5	مثلث ۱	Z0=9.5X0+13.5Y0+51.5
Z1=aX1+bY1+c Z2=aX2+bY2+c Z3=aX3+bY3+c		a=8.4 b=12.4 c=59.2	مثلث ۲	Z0=8.4X0+12.4Y0+59.2
Z1=aX1+bY1+c Z2=aX2+bY2+c Z3=aX3+bY3+c		a=24.7059 b=17.0588 c= - 49.1176	مثلث ۳	Z0=24.71X0+17.06Y0-49.11

با قرار دادن مقادیر مختصات هر نقطه یا سلول (Y0,X0) در رابطه مثلث مربوطه، مقدار Z0 بدست می‌آید.

مثال روش مثلث‌بندی

First Triangular (mm) Z0=9.5X0+13.5Y0+51.5								
Point1	Point2	Point3	Point4	Point5	Point6	Point7	Point8	Point9
Point10	Point11	Point12	Point13	190(C)	Point15	Point16	Point17	Point18
Point19	Point20	Point21	Point22	Point23	Point24	Point25	Point26	Point27
Point28	140(D)	Point30	Point31	Point32	Point33	Point34	230(A)	Point36
Point37	126.5	136	Point40	Point41	Point42	Point43	Point44	Point45
Point46	113	122.5	132	Point50	Point51	Point52	Point53	Point54
90(E)	99.5	109	118.5	128	Point60	Point61	Point62	Point63
Point64	Point65	Point66	105	114.5	124	Point70	Point71	Point72
Point73	Point74	Point75	Point76	Point77	Point78	120(B)	Point80	Point81

Second Triangular (mm) Z0=8.4X0+12.4Y0+59.2								
Point1	Point2	Point3	Point4	Point5	Point6	Point7	Point8	Point9
Point10	Point11	Point12	Point13	190(C)	Point15	Point16	Point17	Point18
Point19	Point20	Point21	169.2	177.6	Point24	Point25	Point26	Point27
Point28	140(D)	148.4	156.8	165.2	Point33	Point34	230(A)	Point36
Point37	Point38	136	144.4	152.8	Point42	Point43	Point44	Point45
Point46	Point47	Point48	132	140.4	148.8	Point52	Point53	Point54
90(E)	Point56	Point57	Point58	128	136.4	Point61	Point62	Point63
Point64	Point65	Point66	Point67	Point68	124	Point70	Point71	Point72
Point73	Point74	Point75	Point76	Point77	Point78	120(B)	Point80	Point81

Third Triangular (mm) Z0=24.71X0+17.06Y0-49.11								
Point1	Point2	Point3	Point4	Point5	Point6	Point7	Point8	Point9
Point10	Point11	Point12	Point13	190(C)	Point15	Point16	Point17	Point18
Point19	Point20	Point21	Point22	Point23	197.685	Point25	Point26	Point27
Point28	140(D)	Point30	Point31	Point32	180.625	205.335	230(A)	Point36
Point37	Point38	Point39	Point40	Point41	163.565	188.275	Point44	Point45
Point46	Point47	Point48	Point49	Point50	Point51	171.215	Point53	Point54
90(E)	Point56	Point57	Point58	Point59	Point60	154.155	Point62	Point63
Point64	Point65	Point66	Point67	Point68	Point69	137.095	Point71	Point72
Point73	Point74	Point75	Point76	Point77	Point78	120(B)	Point80	Point81

حل: در تصویر اول، دوم و سوم به ترتیب، سلول‌های مثلث اول، دوم و سوم درون‌یابی شده است.

مثال روش مثلث‌بندی

حل: نتیجه نهایی روش مثلث‌بندی بدین صورت است:

Interpolated Annual Rainfall Depth (mm)								
Point1	Point2	Point3	Point4	Point5	Point6	Point7	Point8	Point9
Point10	Point11	Point12	Point13	**190(C)**	Point15	Point16	Point17	Point18
Point19	Point20	Point21	169.2	177.6	197.685	Point25	Point26	Point27
Point28	**140(D)**	148.4	156.8	165.2	180.625	205.335	**230(A)**	Point36
Point37	126.5	136	144.4	152.8	163.565	188.275	Point44	Point45
Point46	113	122.5	132	140.4	148.8	171.215	Point53	Point54
90(E)	99.5	109	118.5	128	136.4	154.155	Point62	Point63
Point64	Point65	Point66	105	114.5	124	137.095	Point71	Point72
Point73	Point74	Point75	Point76	Point77	Point78	**120(B)**	Point80	Point81

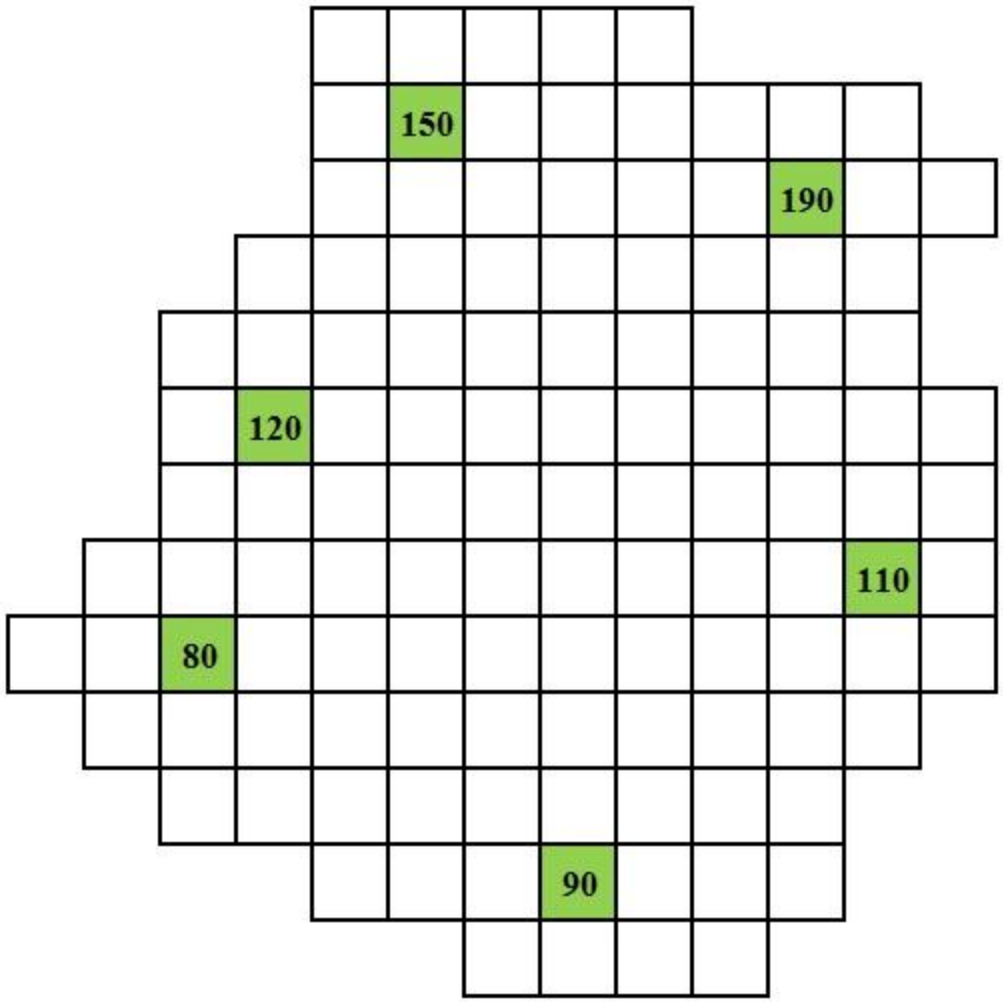
تمرین روش مثلث‌بندی

۱. الف. در حالت کلی، آیا برای نقاط در مرز مشترک مثلث‌ها پس از درونیابی، مقادیر یکسان یا متفاوتی بدست می‌آید؟

۱.ب. فرم معادله کلی درون یابی $z = ax + by + c$ را به شکل زیر نمایش دهید:

$$z_o' = \sum_{i=1}^{n} w_i * z_i \qquad \sum_{i=1}^{n} w_i = 1$$

تمرین روش مثلث‌بندی



۲. عمق بارندگی در ۶ ایستگاه باران‌سنجی مطابق شکل داده شده است. به کمک روش مثلث‌بندی، مقدار بارش را در سایر سلول‌ها تعیین نمائید. ابعاد سلول را یک کیلومتر در نظر بگیرید.

3. Moving averages

- A vector or raster method
- most common GIS method
- calculates value at unsampled location based on the values associated with neighbouring points
- local/global, gradual, exact, and deterministic
- neighbourhood and operation could be specified by a filter
- size, shape and character of filter?

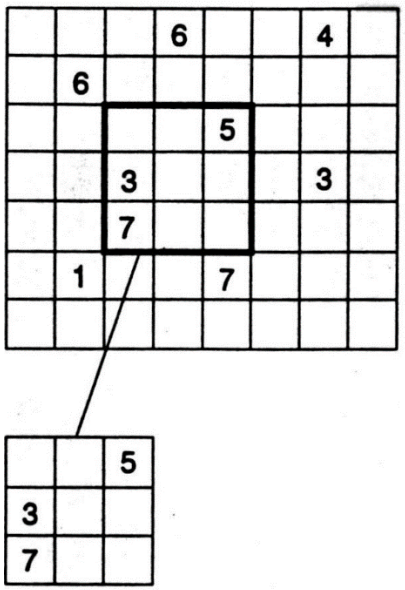
(Weighted) Moving average: Inverse Distance Weighting (IDW)

- "Everything is related to everything else, but near things are more related than distant things." **Tobler's first law of geography.**
- IDW works by using an unbiased weight matrix based on the distances from an unsampled location to sampled locations.
- Weights may be defined in a number of different ways.
- Almost infinite variety of algorithms may be used
- is the most widely used method

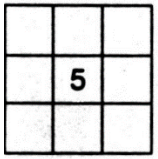
Inverse Distance Weighted

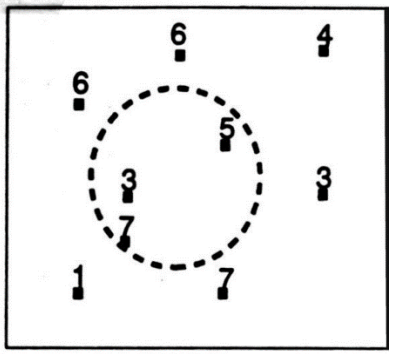
- در این روش فرض بر این است که فاصله از نمونه‌ها تعیین کننده است.
- به بیان دیگر، سلول‌های نزدیک به نقاط نمونه اثر بیشتری نسبت به سلول‌های دورتر از این نقاط می‌گیرند.

Moving average: Arithmetic



$$\frac{3+7+5}{3} = 5$$





$$\frac{3+7+5}{3} = 5$$

Inverse Distance Weighted

Fixed Radius

۱- برای داده‌های مشاهده‌ای زیاد که به طور منطقی پخش شده‌اند.

۲- این روش به یک فاصله همسایگی و حداقل نقاط نیاز دارد.

۳- در این روش اندازه شعاع جستجو یکنواخت خواهد بود و فقط نقاطی که درون دایره با شعاع تعیین شده می‌افتند در پیش‌بینی ارزش نقاط مجهول موثر واقع خواهند شد.

۴- در این روش اگر تعداد نقاط واقع در اطراف نقطه مجهول از تعداد حداقل کمتر باشد شعاع جستجو به صورت خودکار افزایش خواهد یافت تا به حداقل نقاط مورد نیاز جهت پیش‌بینی دست یابد.

Inverse Distance Weighted

Variable Radius

- در این روش شعاع جستجو برای عمل درون‌یابی هر نقطه مجهول متفاوت خواهد بود.
- در روش variable برای هر نقطه مجهول خاص شعاع جستجو به صورت خودکار آنقدر تغییر می‌کند تا به حداقل نقاط مورد نیاز پیش‌بینی دست یابد. به عبارتی، شعاع جستجو بستگی به تراکم نقاط نمونه در اطراف نقطه مجهول مورد نظر دارد.

IDW parameters

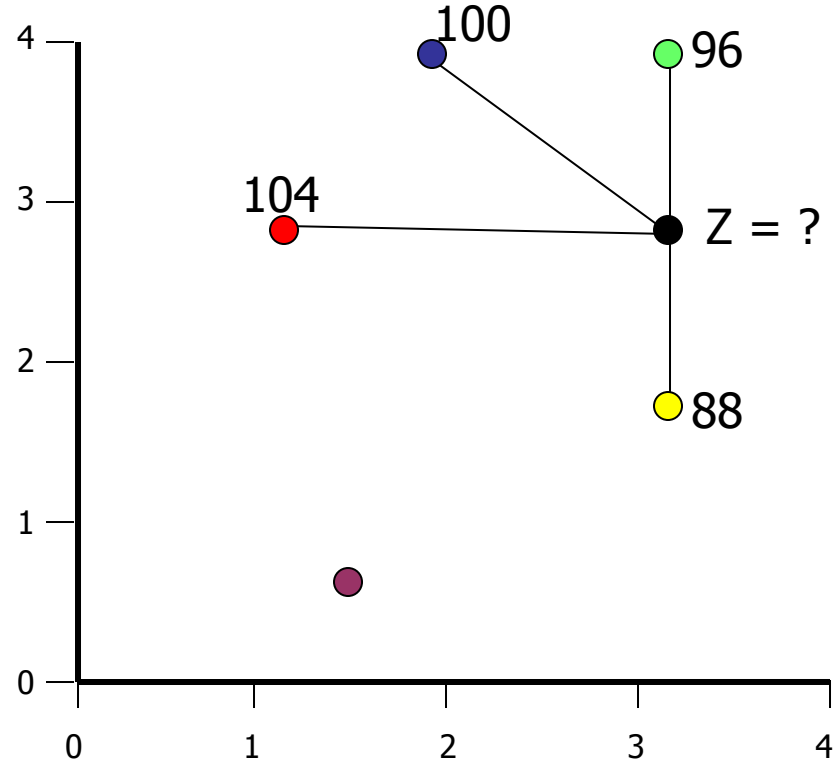
- Exponent to specify distance decay
- Neighborhood:
  1. fixed distance, variable points
  2. variable distance, fixed points
  3. the direction from which they are selected

$$w_i = \frac{D_i^{-\alpha}}{\sum_{j=1}^{n} D_j^{-\alpha}}$$

- objections to this method arise from the fact that the range of interpolated values is limited by the range of the data: no interpolated value will be outside the observed range of z values

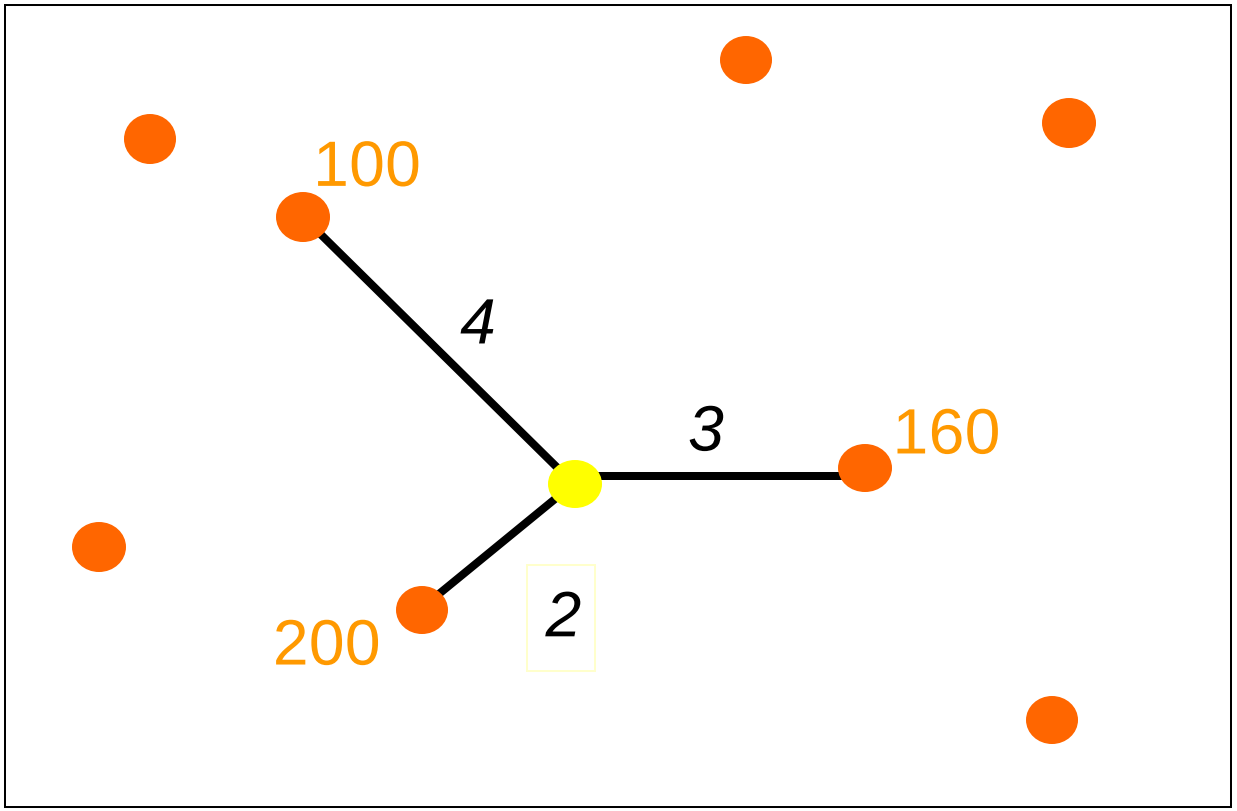
IDW Example 1

- method “honors” real values (exact interpolator)
- user friendly (or, little flexibility)
- fast computational capabilities
- creates a “bull’s-eye” around data locations
- sensitive to outliers



Control Points	Height	Distance	Inverse Distance	Weight	Weighted value
1	104	2.000	0.50	0.1559	16.21
2	100	1.414	0.71	0.2205	22.05
3	96	1.000	1.00	0.3118	29.93
4	88	1.000	1.00	0.3118	27.44
Total			3.21	1.0000	95.63

IDW Example 2

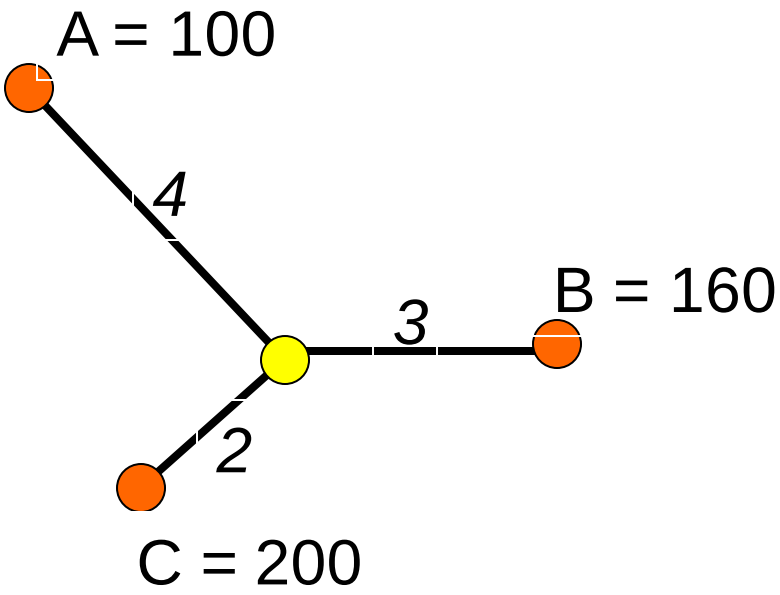


IDW: Closest 3 neighbors, alpha = 2

IDW Example 2

Weights

A	$1 / (4^2) = .0625$
B	$1 / (3^2) = .1111$
C	$1 / (2^2) = .2500$



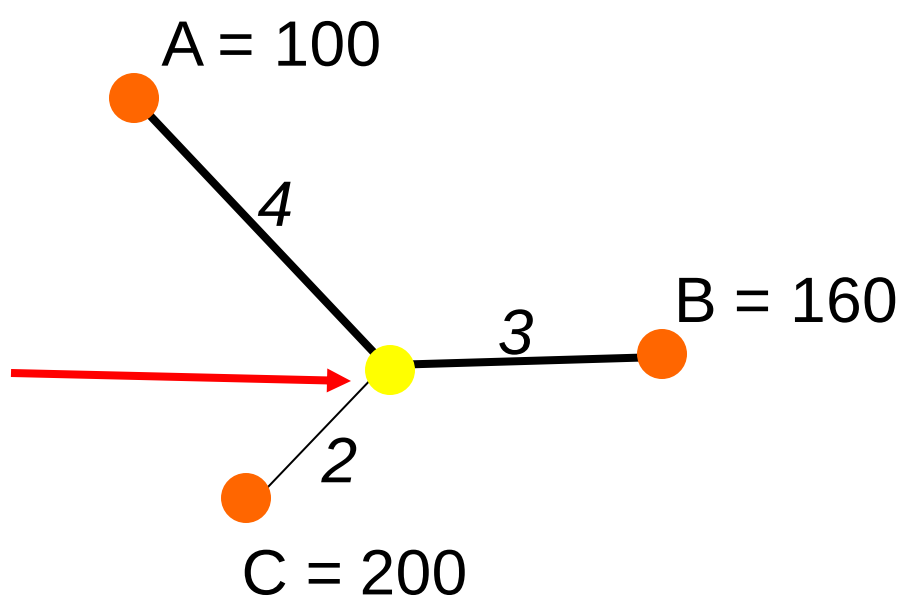
IDW Example 2

	Weights	Weights * Value
A	$1 / (4^2) = .0625$	$.0625 * 100 = 6.25$
B	$1 / (3^2) = .1111$	$.1111 * 160 = 17.76$
C	$1 / (2^2) = .2500$	$.2500 * 200 = 50.00$

Total = .4236

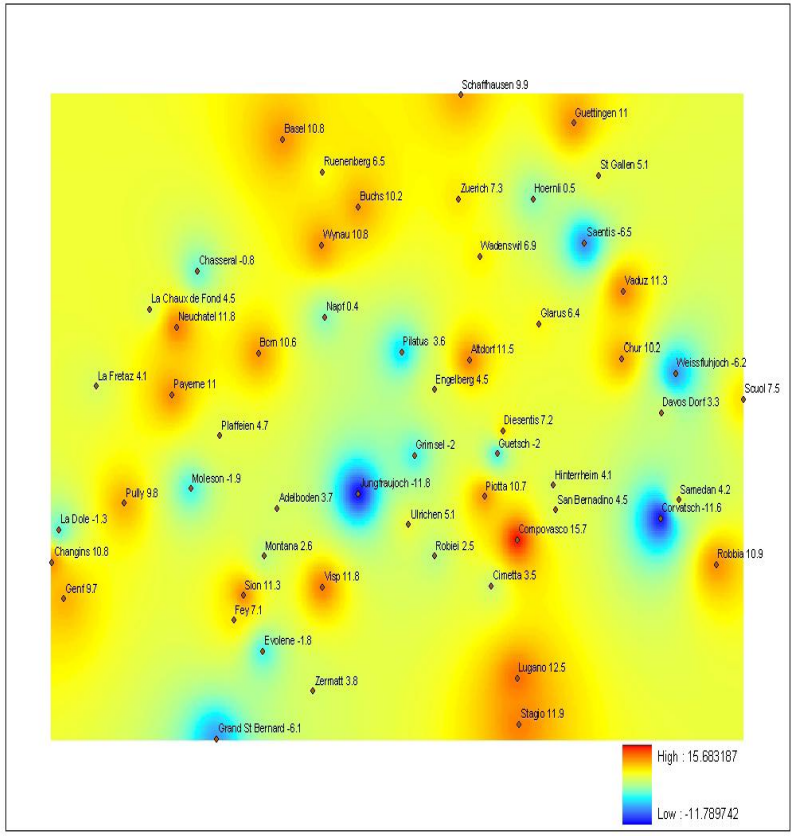
6.25 +17.76 + 50.00 = 74.01

74.01 / .4236 = **175**

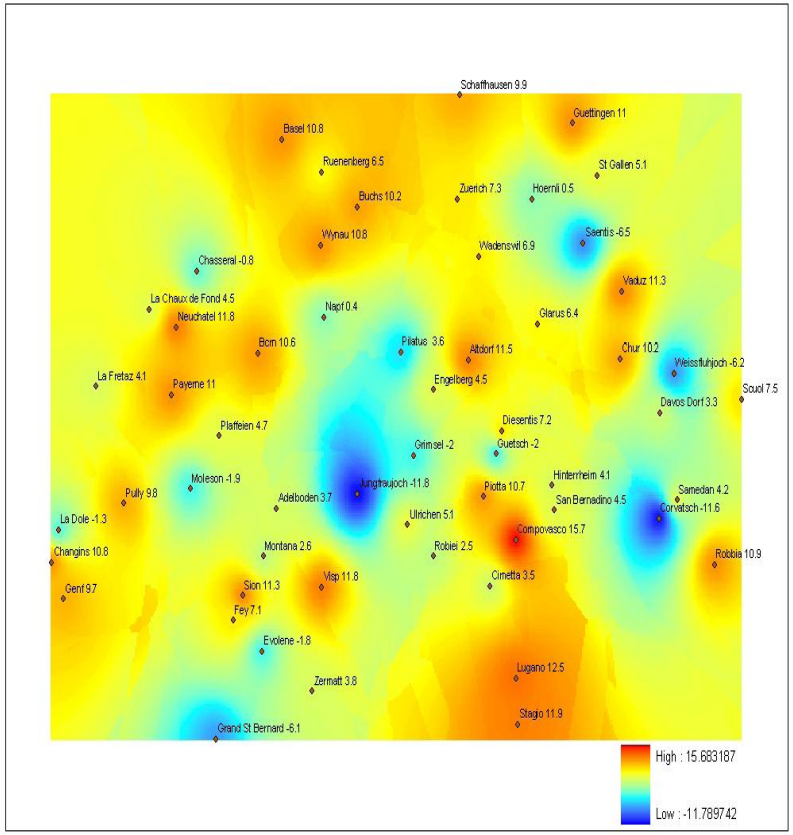


Effect of neighborhood

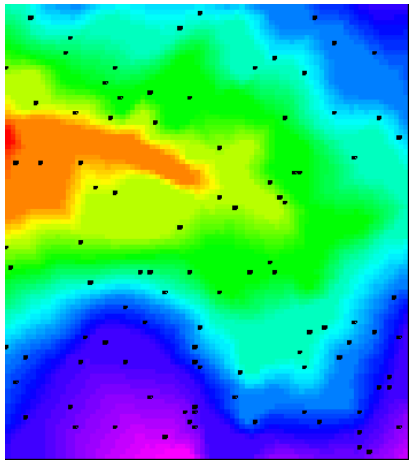
IDW, 60 nearest neighbours, d2 function



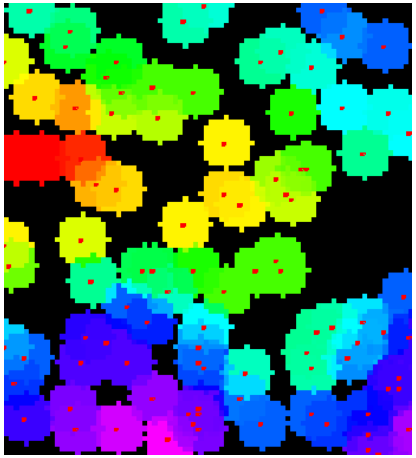
IDW, 6 nearest neighbours, d2 function



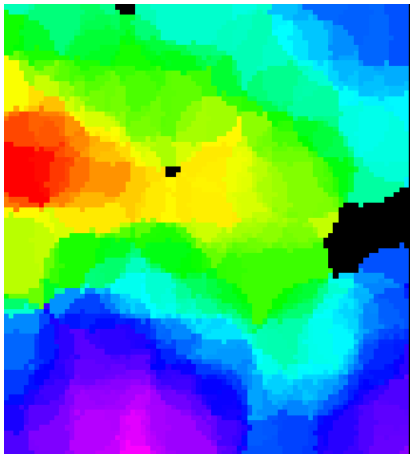
Effect of neighbourhood size



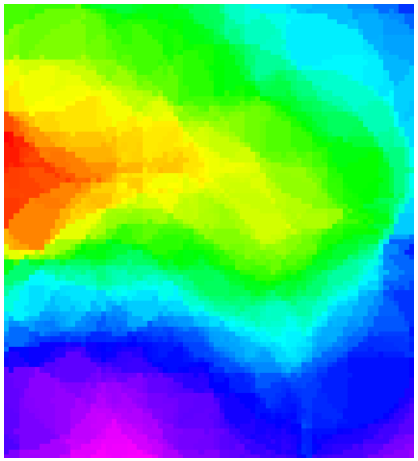
Source surface with sample points



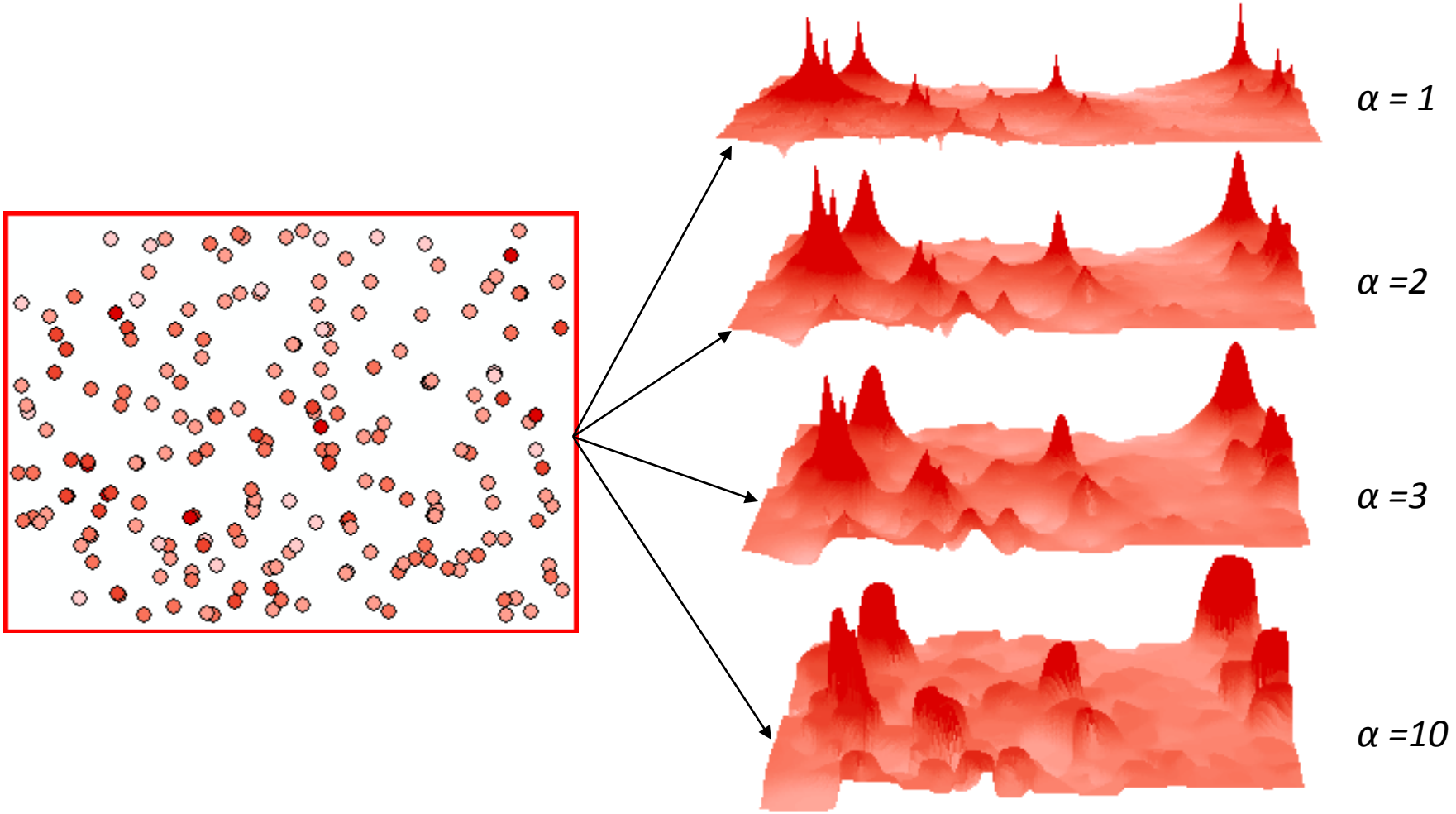
11x11 circular filter SMA with sample points



21x21 circular filter SMA

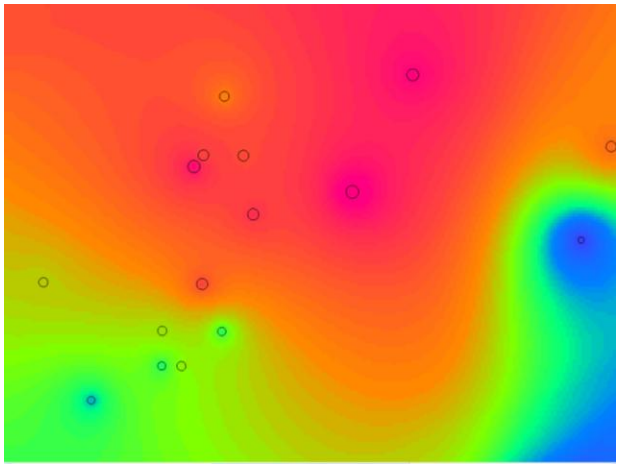


41x41 circular filter SMA

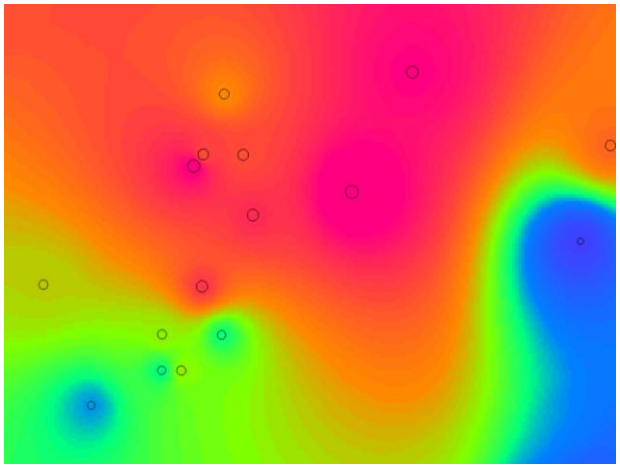

α = 1
α =2
α =3
α =10

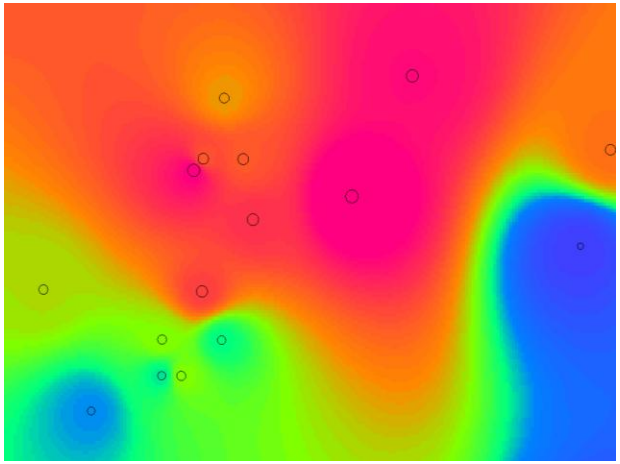
Effect of IDW Exponent

IDW1



IDW2





IDW3

مثال ۳ روش IDW

مثال: عمق بارندگی در محل ایستگاه‌های باران سنجی داده شده است. به روش IDW و با در نظر گرفتن فاصله تاثیر ۶ کیلومتر و توان ۱، نقشه بارندگی منطقه را ترسیم کنید. ابعاد سلول را یک کیلومتر در نظر بگیرید.

Annual Rainfall Depth (mm)								
Point1	Point2	Point3	Point4	Point5	Point6	Point7	Point8	Point9
Point10	Point11	Point12	Point13	**190(C)**	Point15	Point16	Point17	Point18
Point19	Point20	Point21	Point22	Point23	Point24	Point25	Point26	Point27
Point28	**140(D)**	Point30	Point31	Point32	Point33	Point34	**230(A)**	Point36
Point37	Point38	Point39	Point40	Point41	Point42	Point43	Point44	Point45
Point46	Point47	Point48	Point49	Point50	Point51	Point52	Point53	Point54
90(E)	Point56	Point57	Point58	Point59	Point60	Point61	Point62	Point63
Point64	Point65	Point66	Point67	Point68	Point69	Point70	Point71	Point72
Point73	Point74	Point75	Point76	Point77	Point78	**120(B)**	Point80	Point81

مثال ۳ روش IDW

حل: نمونه‌ای از محاسبات برای سلول‌های ۱ تا ۴ به صورت مقابل است.

Point Num	Control Points	Rain Depth	Distance	Inverse Distance	Weight	Weighted Value
1	1	190	4.12	0.243	0.33	63.52
	2	140	3.16	0.316	0.44	61.03
	3	90	6	0.167	0.23	20.68
	4					
	5					
	∑			0.725	1	**145.23**

Point Num	Control Points	Rain Depth	Distance	Inverse Distance	Weight	Weighted Value
2	1	190	3.16	0.316	0.49	92.50
	2	140	3.00	0.333	0.51	71.84
	3					
	4					
	5					
	∑			0.650	1	**164.34**

Point Num	Control Points	Rain Depth	Distance	Inverse Distance	Weight	Weighted Value
3	1	190	2.24	0.447	0.48	90.88
	2	140	3.16	0.316	0.34	47.35
	3	230	5.83	0.171	0.18	42.19
	4					
	5					
	∑			0.935	1	**180.43**

Point Num	Control Points	Rain Depth	Distance	Inverse Distance	Weight	Weighted Value
4	1	190	1.41	0.707	0.60	113.43
	2	140	3.61	0.277	0.23	32.78
	3	230	5.00	0.200	0.17	38.84
	4					
	5					
	∑			1.184	1	**185.05**

مثال ۳ روش IDW

حل: نتیجه درون‌یابی به روش IDW برای به صورت زیر است:

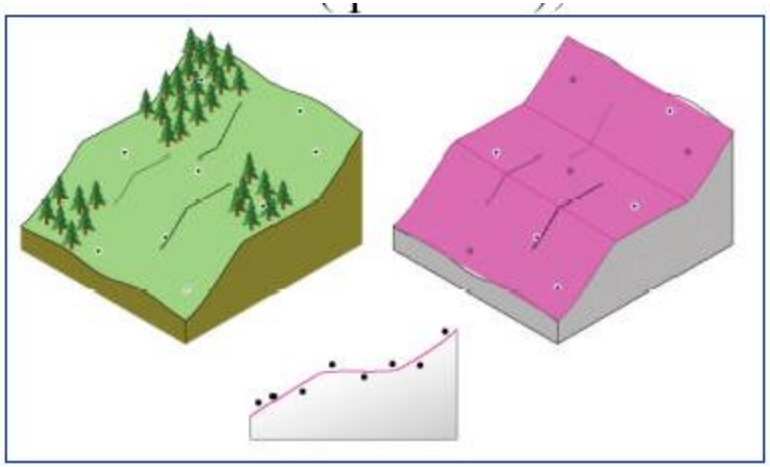
Interpolated Annual Rainfall Depth (mm) IDW D=6 Km a=1								
145.23	164.34	180.43	185.05	188.40	190.92	194.36	210.53	212.64
142.79	146.66	164.59	175.20	**190(C)**	191.88	197.59	214.00	215.66
139.69	142.36	158.32	169.55	178.48	194.12	194.49	220.39	219.78
136.48	**140(D)**	151.91	158.48	166.60	175.06	199.12	**230(A)**	209.16
129.34	133.72	146.02	153.33	160.86	169.23	190.68	205.57	201.17
114.65	121.77	141.02	148.55	154.94	160.88	177.62	187.90	188.31
90(E)	113.30	136.60	144.35	149.13	151.98	154.49	171.12	175.35
99.76	105.94	110.81	133.97	144.19	137.07	141.31	148.73	158.68
107.16	108.89	111.64	114.22	134.61	131.77	**120(B)**	138.33	150.99

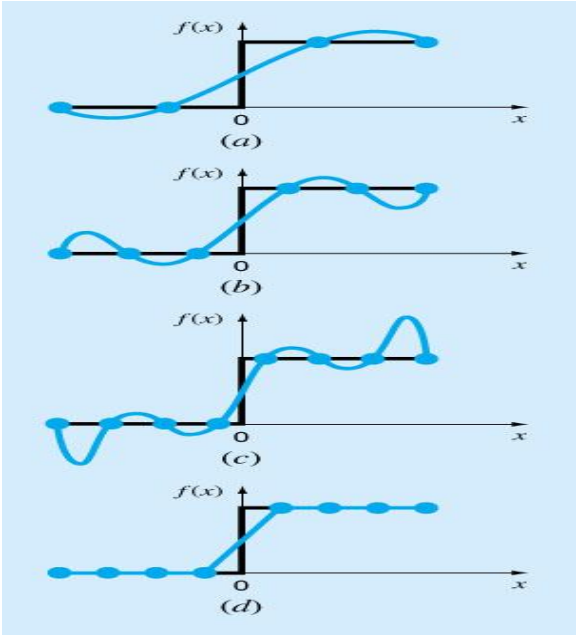
4. B-splines (optional)

- uses a piecewise (local) polynomial to provide a series of patches resulting in a surface that has continuous first and second derivatives
- ensures continuity in:
  - elevation (zero-order continuity) - surface has no cliffs
  - slope (first-order continuity) - slopes do not change abruptly, in case of quadratic polynomials
  - curvature (second order continuity) - minimum curvature is achieved in case of cubic polynomial

B-splines

- note that maxima and minima do not necessarily occur at the data points
- best for very smooth surfaces
- poor for surfaces which show marked fluctuations, this can cause wild oscillations in the spline
- are popular in general surface interpolation packages but are not common in GISs

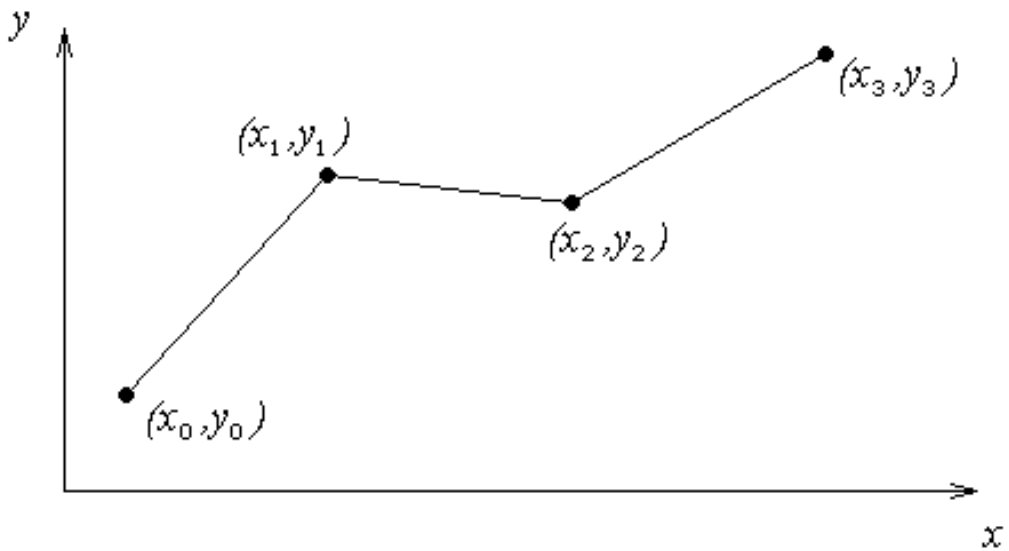




Local Linear Interpolation

Given $(x_0, y_0), (x_1, y_1), \ldots\ldots, (x_{n-1}, y_{n-1})(x_n, y_n)$, fit linear splines to the data. This simply involves forming the consecutive data through straight lines. So if the above data is given in an ascending order, the linear splines are given by $(y_i = f(x_i))$

Figure : Linear splines



Local Linear Interpolation (contd)

$$f(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0), \qquad x_0 \le x \le x_1$$

$$= f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1), \qquad x_1 \le x \le x_2$$

$$\vdots$$

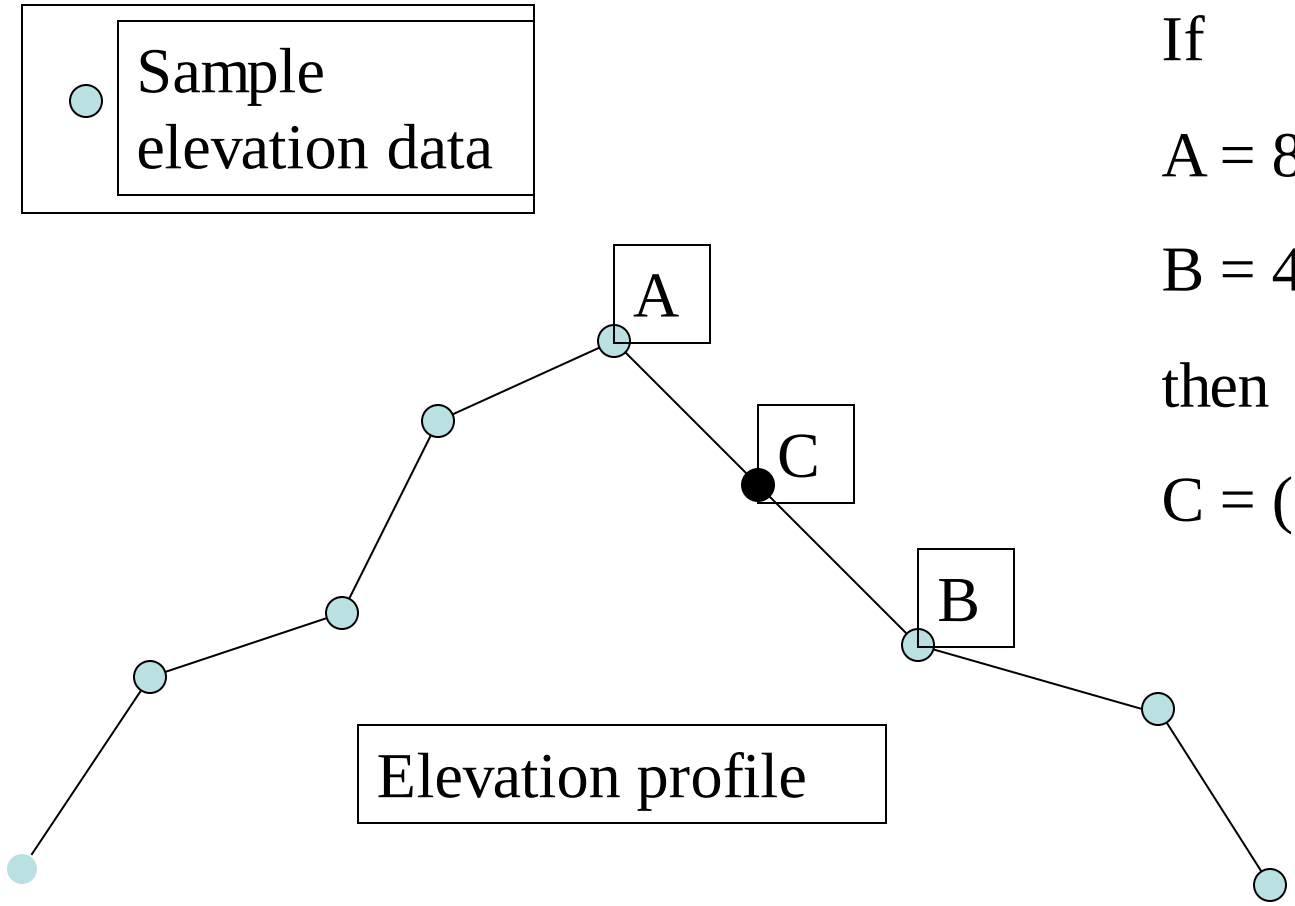
$$= f(x_{n-1}) + \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}(x - x_{n-1}), \quad x_{n-1} \le x \le x_n$$

Note the terms of

$$\frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}$$

in the above function are simply slopes between x_{i-1} and x_i.

Local Linear Interpolation



If

A = 8 feet and

B = 4 feet

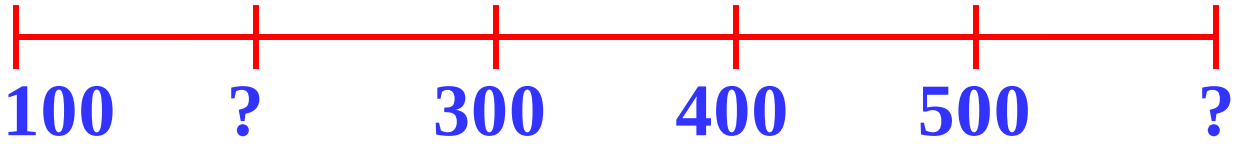
then

C = (8 + 4) / 2 = 6 feet

Linear Interpolation

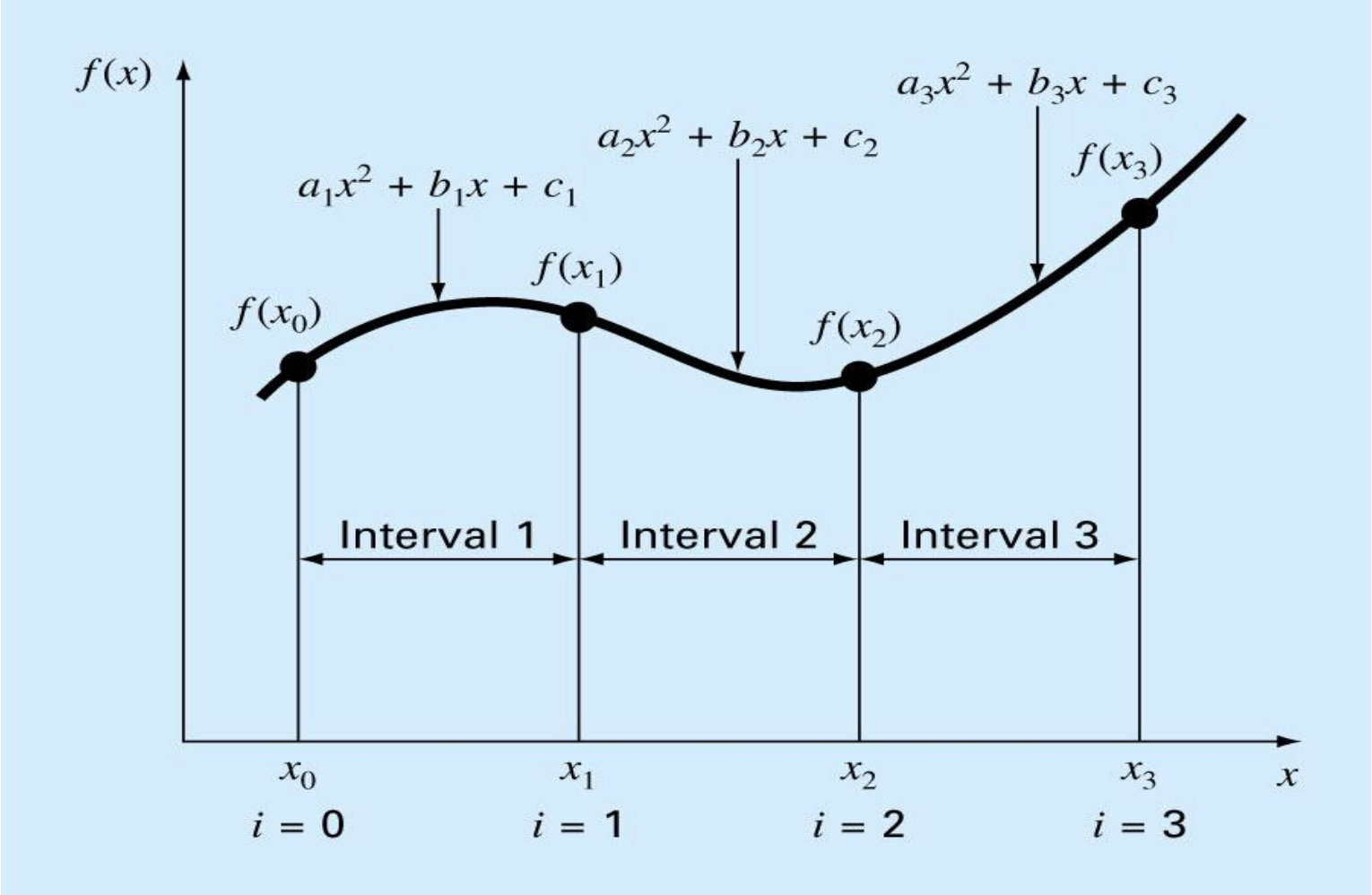
- Simple!

Linear relationship; easy to assign values between known values



Answer: 200 and 600!

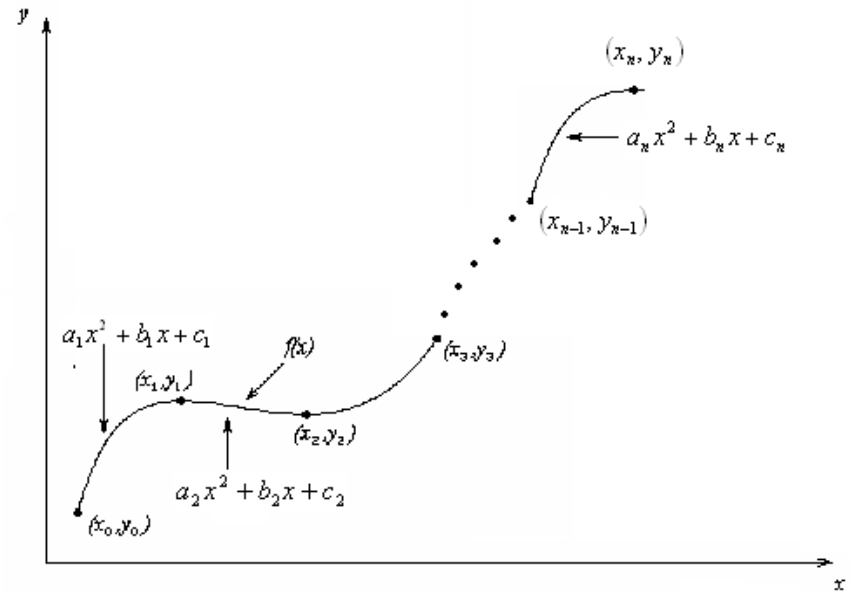
Quadratic Interpolation



Local Quadratic Interpolation

Given $(x_0, y_0), (x_1, y_1), \ldots\ldots, (x_{n-1}, y_{n-1}), (x_n, y_n)$, fit quadratic splines through the data. The splines are given by

$$
\begin{aligned}
f(x) &= a_1 x^2 + b_1 x + c_1, && x_0 \le x \le x_1 \\
&= a_2 x^2 + b_2 x + c_2, && x_1 \le x \le x_2 \\
&\;\;\vdots \\
&= a_n x^2 + b_n x + c_n, && x_{n-1} \le x \le x_n
\end{aligned}
$$



Find a_i, b_i, c_i, $i = 1, 2, \ldots, n$

Quadratic Interpolation (contd)

Each quadratic spline goes through two consecutive data points

$$a_1 x_0^2 + b_1 x_0 + c_1 = f(x_0)$$

$$a_1 x_1^2 + b_1 x_1 + c_1 = f(x_1) \quad .$$

$$\vdots$$

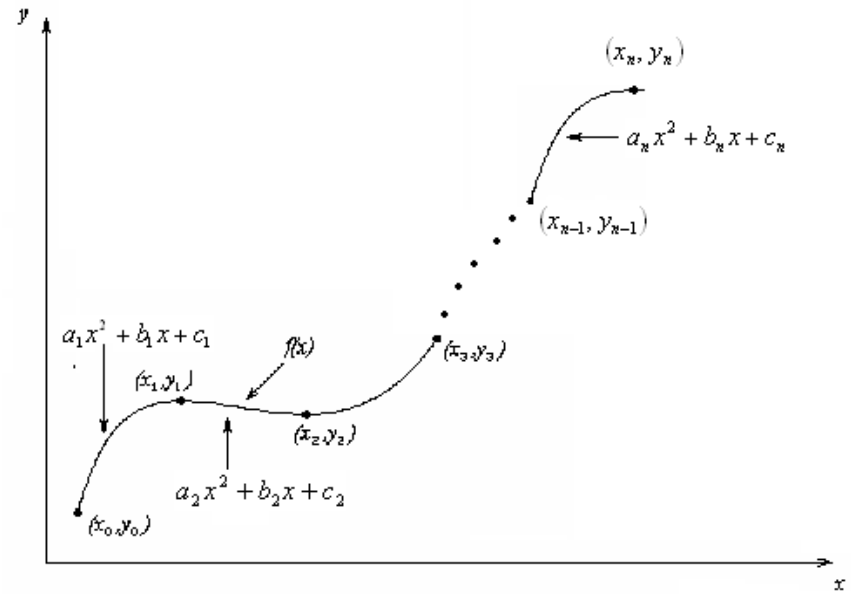
$$a_i x_{i-1}^2 + b_i x_{i-1} + c_i = f(x_{i-1})$$

$$a_i x_i^2 + b_i x_i + c_i = f(x_i) \quad .$$

$$\vdots$$

$$a_n x_{n-1}^2 + b_n x_{n-1} + c_n = f(x_{n-1})$$

$$a_n x_n^2 + b_n x_n + c_n = f(x_n)$$



This condition gives 2n equations

Quadratic Splines (contd)

The first derivatives of two quadratic splines are continuous at the interior points.

For example, the derivative of the first spline

$$a_1 x^2 + b_1 x + c_1 \quad \text{is} \quad 2a_1 x + b_1$$

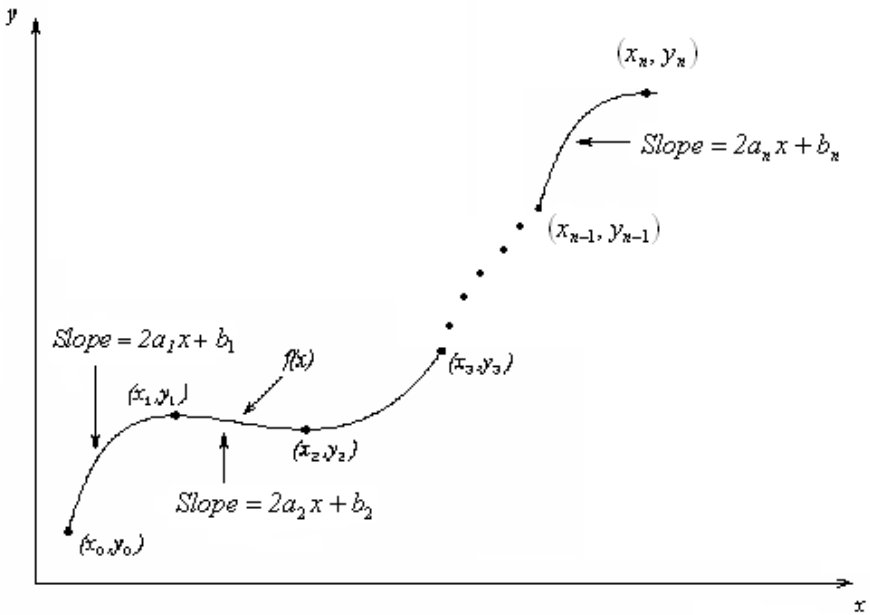
The derivative of the second spline

$$a_2 x^2 + b_2 x + c_2 \quad \text{is} \quad 2a_2 x + b_2$$

and the two are equal at $x = x_1$ giving

$$2a_1 x_1 + b_1 = 2a_2 x_1 + b_2$$

$$2a_1 x_1 + b_1 - 2a_2 x_1 - b_2 = 0$$



Quadratic Splines (contd)

Similarly at the other interior points,

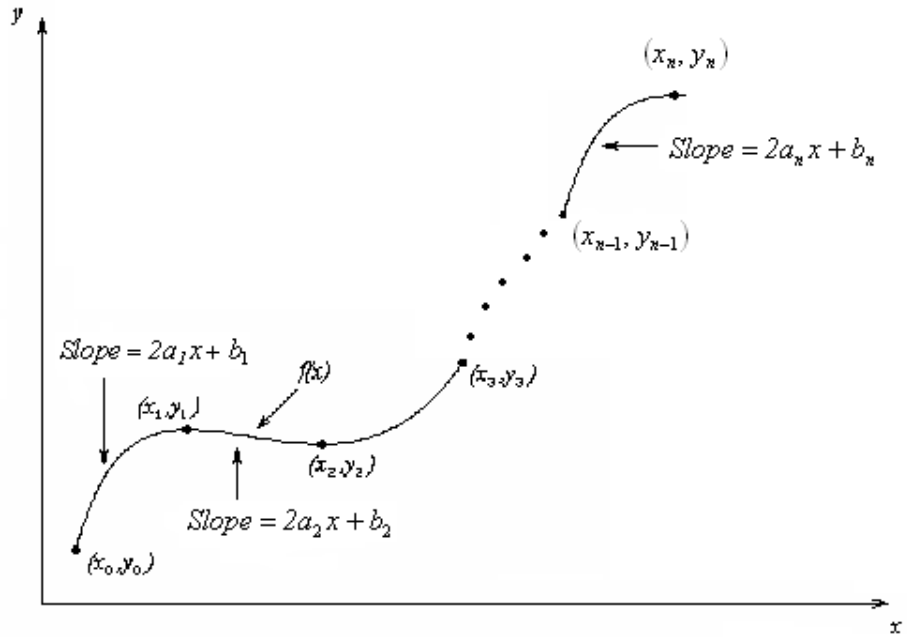
$$2a_2x_2 + b_2 - 2a_3x_2 - b_3 = 0$$

$$\vdots$$

$$2a_ix_i + b_i - 2a_{i+1}x_i - b_{i+1} = 0$$

$$\vdots$$

$$2a_{n-1}x_{n-1} + b_{n-1} - 2a_nx_{n-1} - b_n = 0$$



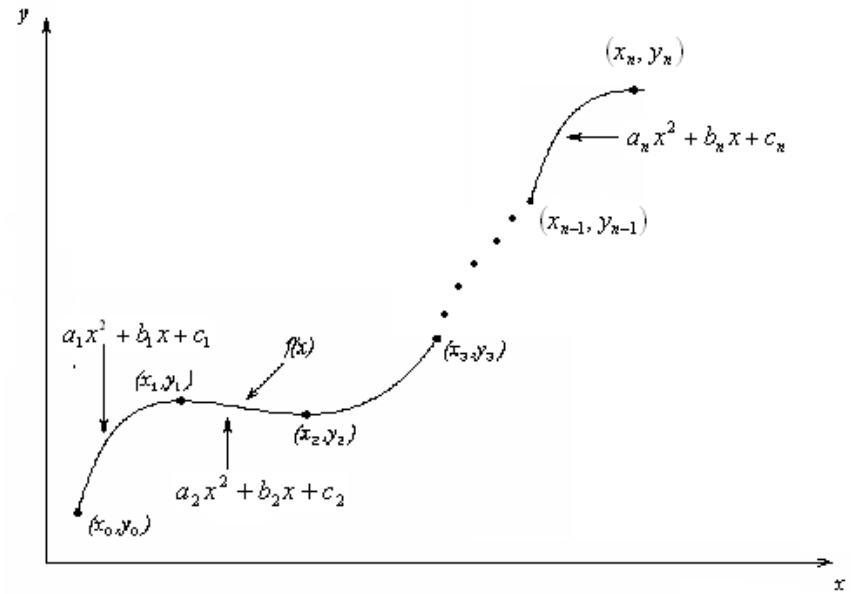
We have (n-1) such equations. The total number of equations is $(2n) + (n-1) = (3n-1)$.

We can assume that the first spline is linear, that is $a_1 = 0$

Quadratic Splines (contd)

This gives us ‘3n’ equations and ‘3n’ unknowns. Once we find the ‘3n’ constants, we can find the function at any value of ‘x’ using the splines,

$$
\begin{aligned}
f(x) &= a_1 x^2 + b_1 x + c_1, && x_0 \le x \le x_1 \\
&= a_2 x^2 + b_2 x + c_2, && x_1 \le x \le x_2 \\
&\;\;\vdots \\
&= a_n x^2 + b_n x + c_n, && x_{n-1} \le x \le x_n
\end{aligned}
$$



Why Splines ?

$$f(x) = \frac{1}{1+25x^2}$$

Table : Six equidistantly spaced points in [-1, 1]

x	$y = \frac{1}{1+25x^2}$
-1.0	0.038461
-0.6	0.1
-0.2	0.5
0.2	0.5
0.6	0.1
1.0	0.038461

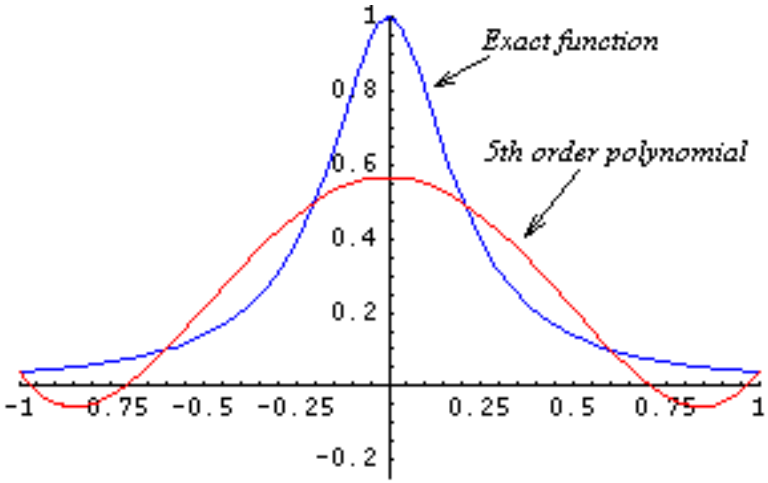


Figure : 5th order polynomial vs. exact function

Higher order polynomial interpolation

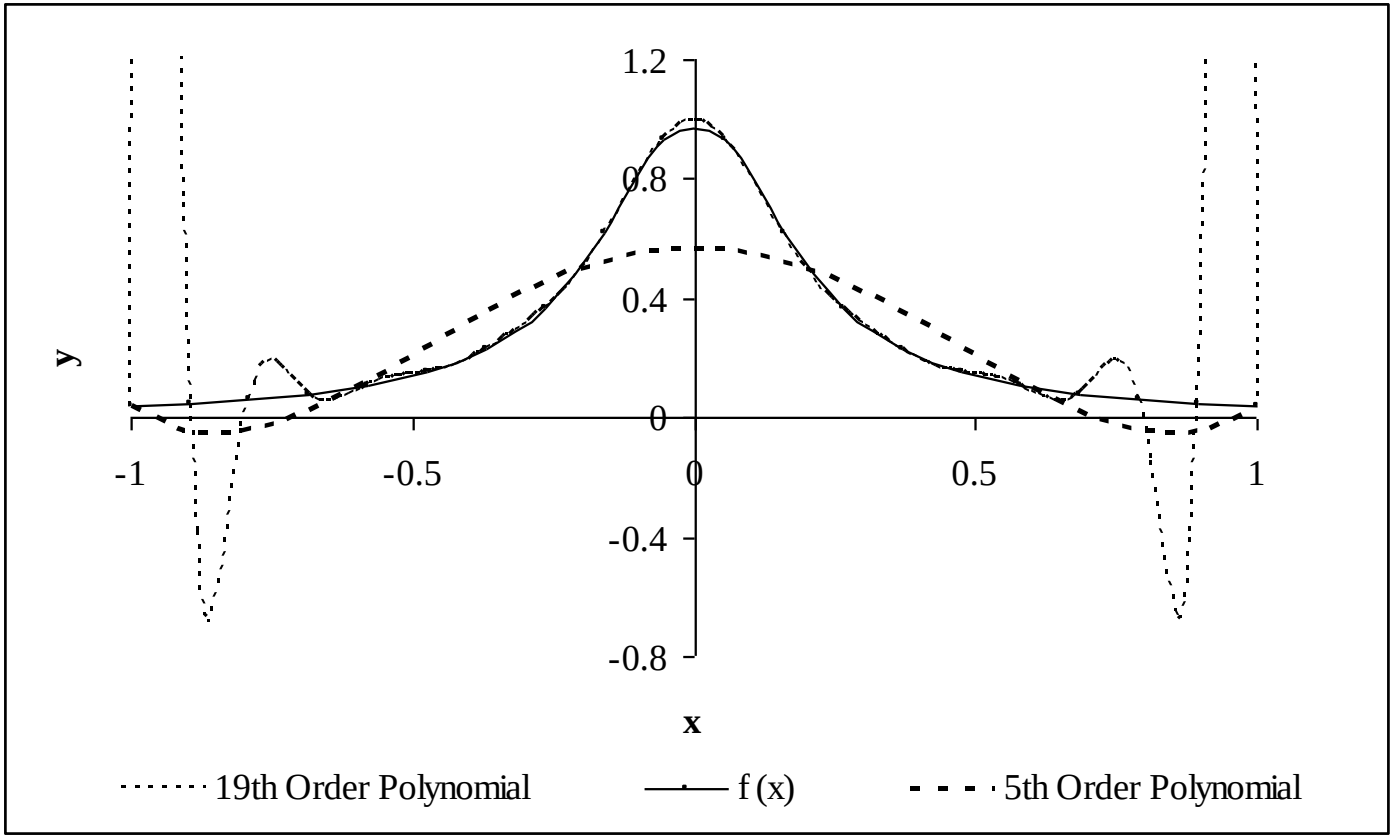


Figure : Higher order polynomial interpolation is a bad idea

Example

The upward velocity of a rocket is given as a function of time in Table 1. Find the velocity at t=16 seconds using linear splines.

Table Velocity as a function of time

(s)	(m/s)
t_0	$v(t)_0$
10	227.04
15	362.78
20	517.35
22.5	602.97
30	901.67

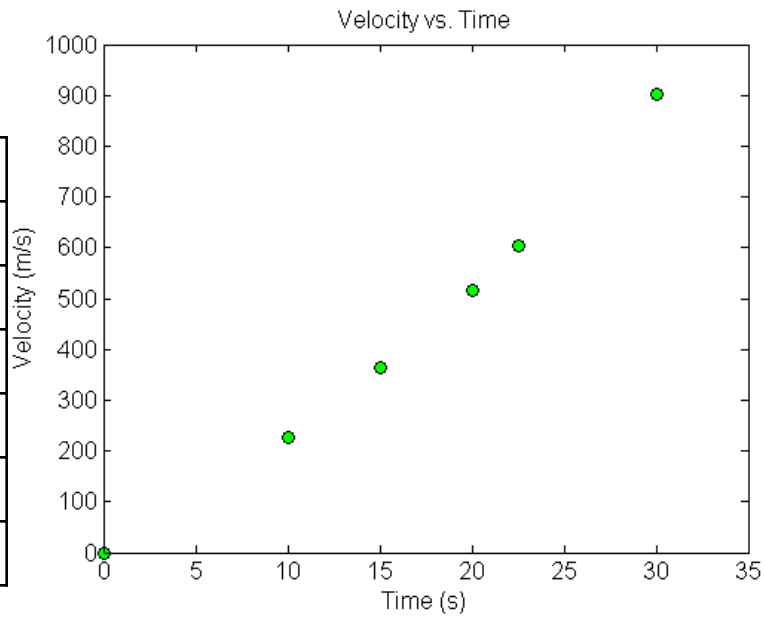


Figure. Velocity vs. time data for the rocket example

Linear Interpolation

$$t_0 = 15, \qquad v(t_0) = 362.78$$

$$t_1 = 20, \qquad v(t_1) = 517.35$$

$$v(t) = v(t_0) + \frac{v(t_1) - v(t_0)}{t_1 - t_0}(t - t_0)$$

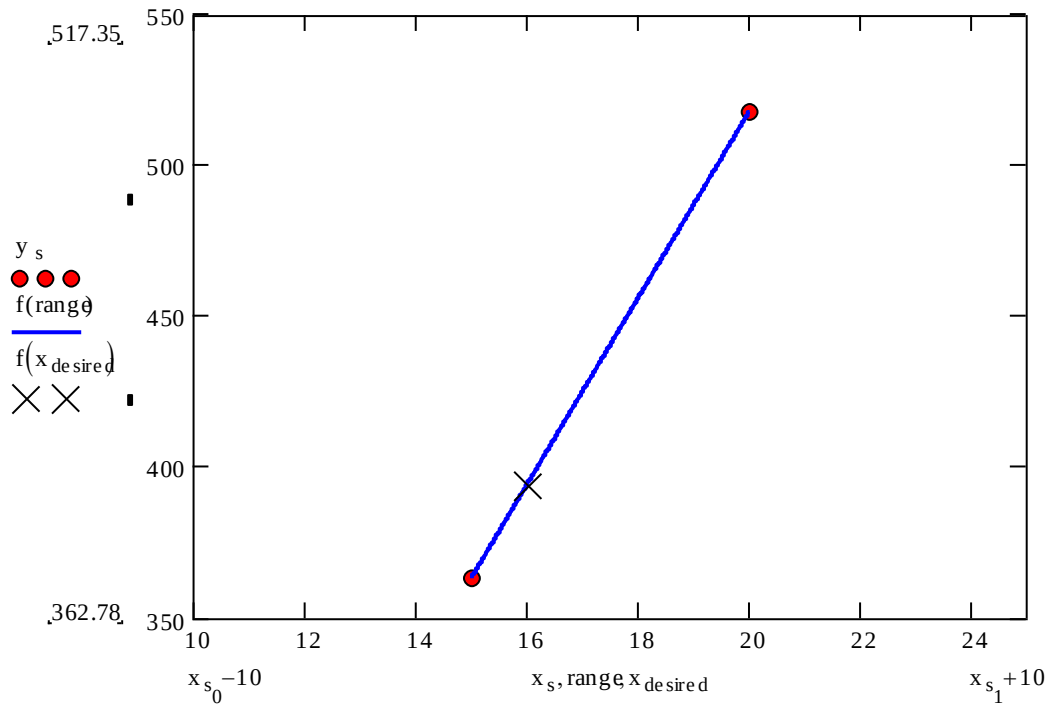
$$= 362.78 + \frac{517.35 - 362.78}{20 - 15}(t - 15)$$

$$v(t) = 362.78 + 30.913(t - 15)$$

At $t = 16$,

$$v(16) = 362.78 + 30.913(16 - 15)$$

$$= 393.7 \text{ m/s}$$



Solution

$$
\begin{aligned}
v(t) &= a_1 t^2 + b_1 t + c_1, & 0 \le t \le 10 \\
&= a_2 t^2 + b_2 t + c_2, & 10 \le t \le 15 \\
&= a_3 t^2 + b_3 t + c_3, & 15 \le t \le 20 \\
&= a_4 t^2 + b_4 t + c_4, & 20 \le t \le 22.5 \\
&= a_5 t^2 + b_5 t + c_5, & 22.5 \le t \le 30
\end{aligned}
$$

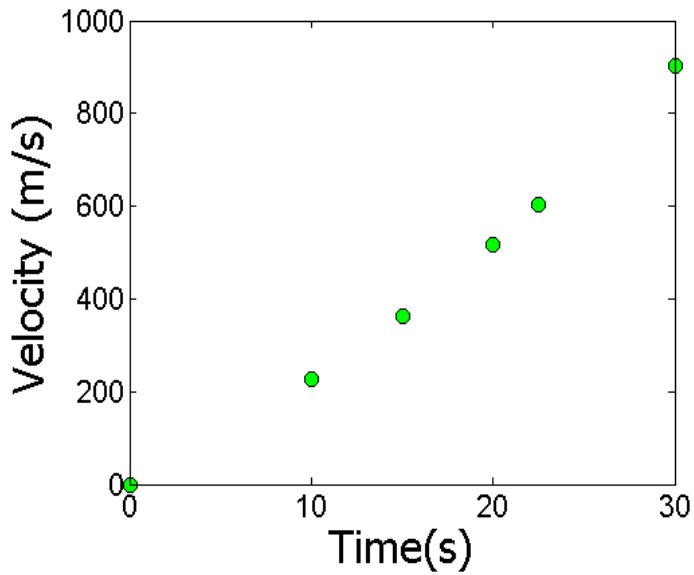
Let us set up the equations

Each Spline Goes Through Two Consecutive Data Points

$$v(t) = a_1 t^2 + b_1 t + c_1, \qquad 0 \le t \le 10$$

$$a_1 (0)^2 + b_1 (0) + c_1 = 0$$

$$a_1 (10)^2 + b_1 (10) + c_1 = 227.04$$



Each Spline Goes Through Two Consecutive Data Points

t	v(t)
s	m/s
0	0
10	227.04
15	362.78
20	517.35
22.5	602.97
30	901.67

$$a_2(10)^2 + b_2(10) + c_2 = 227.04$$

$$a_2(15)^2 + b_2(15) + c_2 = 362.78$$

$$a_3(15)^2 + b_3(15) + c_3 = 362.78$$

$$a_3(20)^2 + b_3(20) + c_3 = 517.35$$

$$a_4(20)^2 + b_4(20) + c_4 = 517.35$$

$$a_4(22.5)^2 + b_4(22.5) + c_4 = 602.97$$

$$a_5(22.5)^2 + b_5(22.5) + c_5 = 602.97$$

$$a_5(30)^2 + b_5(30) + c_5 = 901.67$$

Derivatives are Continuous at Interior Data Points

$$v(t) = a_1 t^2 + b_1 t + c_1,\ 0 \le t \le 10$$

$$= a_2 t^2 + b_2 t + c_2, 10 \le t \le 15$$

$$\left.\frac{d}{dt}\left(a_1 t^2 + b_1 t + c_1\right)\right|_{t=10} = \left.\frac{d}{dt}\left(a_2 t^2 + b_2 t + c_2\right)\right|_{t=10}$$

$$\left.\left(2a_1 t + b_1\right)\right|_{t=10} = \left.\left(2a_2 t + b_2\right)\right|_{t=10}$$

$$2a_1(10) + b_1 = 2a_2(10) + b_2$$

$$20a_1 + b_1 - 20a_2 - b_2 = 0$$

Derivatives are Continuous at Interior Data Points

At t=10

$$2a_1(10) + b_1 - 2a_2(10) - b_2 = 0$$

At t=15

$$2a_2(15) + b_2 - 2a_3(15) - b_3 = 0$$

At t=20

$$2a_3(20) + b_3 - 2a_4(20) - b_4 = 0$$

At t=22.5

$$2a_4(22.5) + b_4 - 2a_5(22.5) - b_5 = 0$$

$$a_1 = 0$$

Final Set of Equations

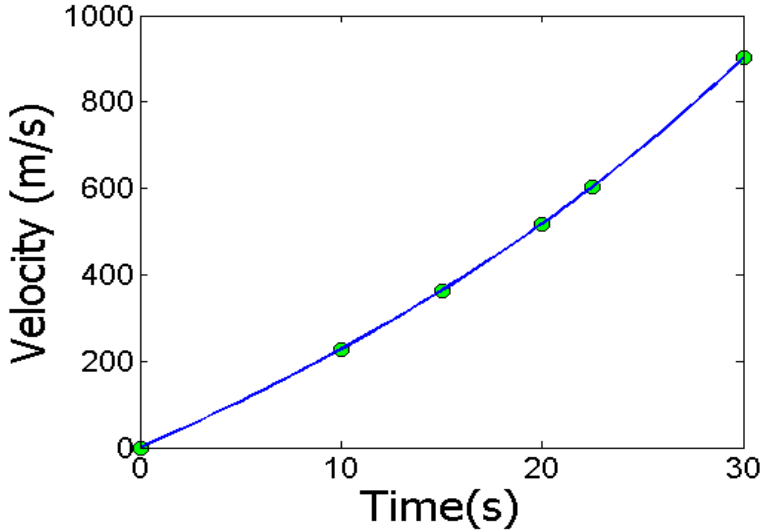
$$
\begin{bmatrix}
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
100 & 10 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 100 & 10 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 225 & 15 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 225 & 15 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 400 & 20 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 400 & 20 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 506.25 & 22.5 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 506.25 & 22.5 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 900 & 30 & 1 \\
20 & 1 & 0 & -20 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 30 & 1 & 0 & -30 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 40 & 1 & 0 & -40 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 45 & 1 & 0 & -45 & -1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ a_2 \\ b_2 \\ c_2 \\ a_3 \\ b_3 \\ c_3 \\ a_4 \\ b_4 \\ c_4 \\ a_5 \\ b_5 \\ c_5 \end{bmatrix}
=
\begin{bmatrix} 0 \\ 227.04 \\ 227.04 \\ 362.78 \\ 362.78 \\ 517.35 \\ 517.35 \\ 602.97 \\ 602.97 \\ 901.67 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
$$

Coefficients of Spline

i	a_i	b_i	c_i
1	0	22.704	0
2	0.8888	4.928	88.88
3	−0.1356	35.66	−141.61
4	1.6048	−33.956	554.55
5	0.20889	28.86	−152.13

Final Solution

$$
\begin{aligned}
v(t) &= 22.704t, && 0 \le t \le 10 \\
&= 0.8888t^2 + 4.928t + 88.88, && 10 \le t \le 15 \\
&= -0.1356t^2 + 35.66t - 141.61, && 15 \le t \le 20 \\
&= 1.6048t^2 - 33.956t + 554.55, && 20 \le t \le 22.5 \\
&= 0.20889t^2 + 28.86t - 152.13, && 22.5 \le t \le 30
\end{aligned}
$$



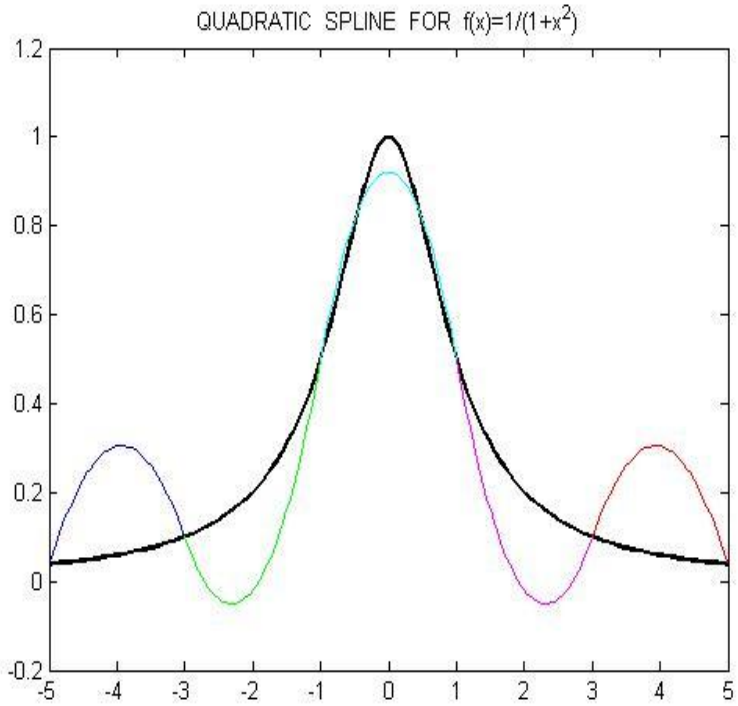
Velocity at a Particular Point

a) Velocity at t=16

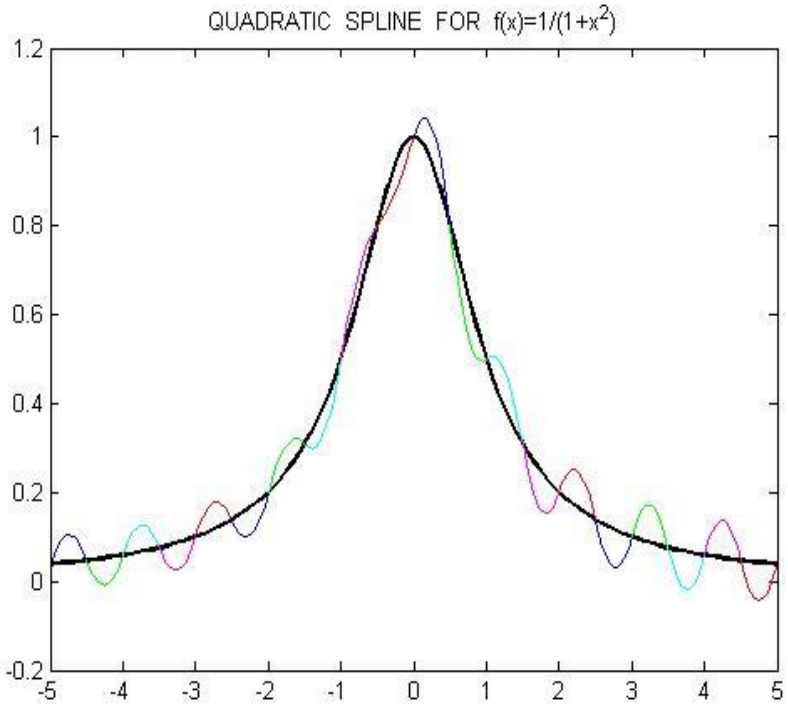
$$
\begin{aligned}
v(t) &= 22.704t, && 0 \le t \le 10 \\
&= 0.8888t^2 + 4.928t + 88.88, && 10 \le t \le 15 \\
&= -0.1356t^2 + 35.66t - 141.61, && 15 \le t \le 20 \\
&= 1.6048t^2 - 33.956t + 554.55, && 20 \le t \le 22.5 \\
&= 0.20889t^2 + 28.86t - 152.13, && 22.5 \le t \le 30
\end{aligned}
$$

$$
\begin{aligned}
v(16) &= -0.1356(16)^2 + 35.66(16) - 141.61 \\
&= 394.24 \text{ m/s}
\end{aligned}
$$

Quadratic Spline Graph



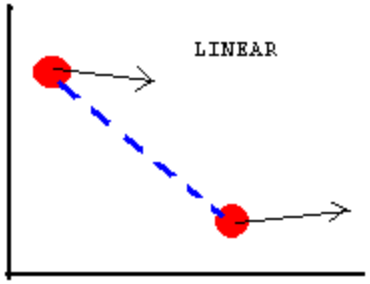
t=a:2:b;



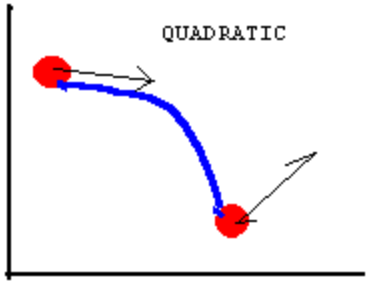
t=a:0.5:b;

5. Polynomial Interpolation: Trend Surfaces

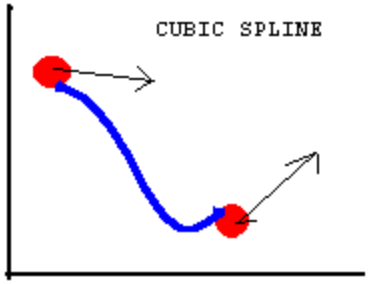
- Linear, nonlinear (eg. Quadratic, Cubic, etc.)
- 1 to 3rd degree polynomials are preferred
- More efficient, smoother with smaller error
- High-degree polynomials are very computationally expensive



The path is unrealistic considering the inital velocity vector.

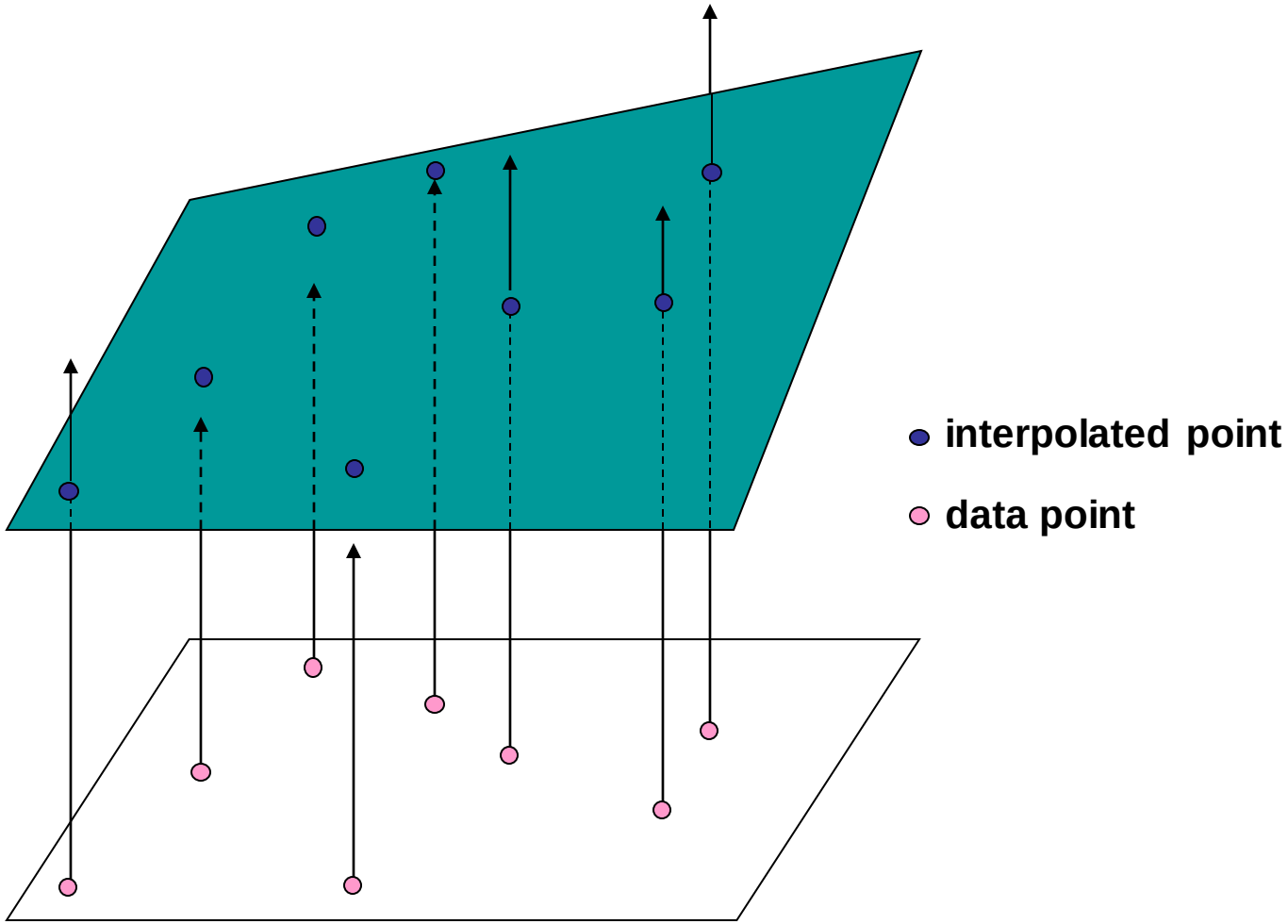


The path doesn't account for the final velocity.

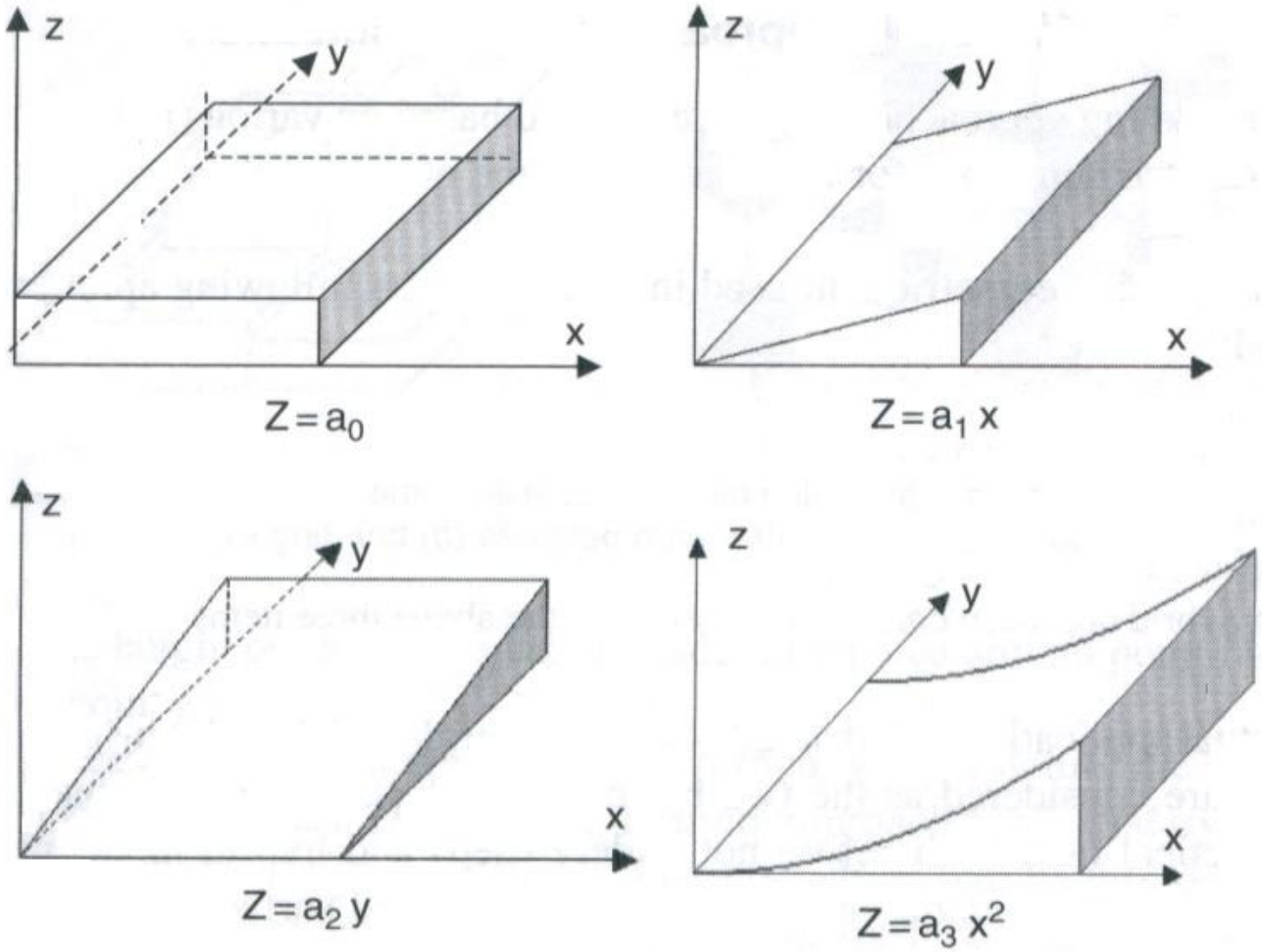


Notice that the path accounts for initial and final velocities.

Global Linear trend surface



Trend Surfaces



Trend Surfaces

- linear equation (degree 1) describes a tilted plane surface:

  $z = a + bx + cy$

- a quadratic equation (degree 2) describes a simple hill or valley

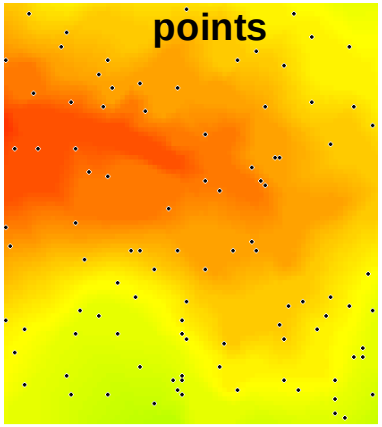
  $z = a + bx + cy + dx^2 + exy + fy^2$

- a cubic surface can have one maximum and one minimum in any cross-section

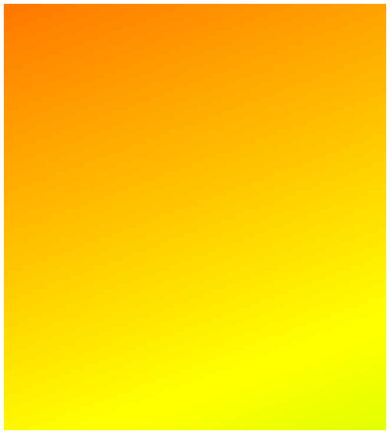
  $z = a + bx + cy + dx^2 + exy + fy^2 + gx^3 + hx^2y + ixy^2 + jy^3$

Example: trend surfaces

Source surface with sample points

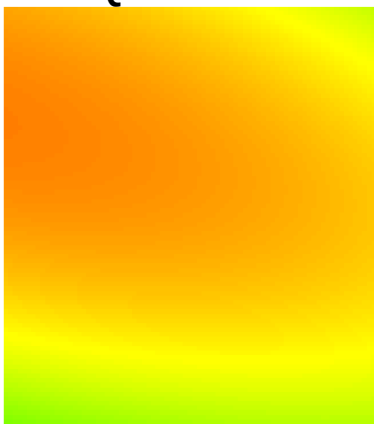


Linear



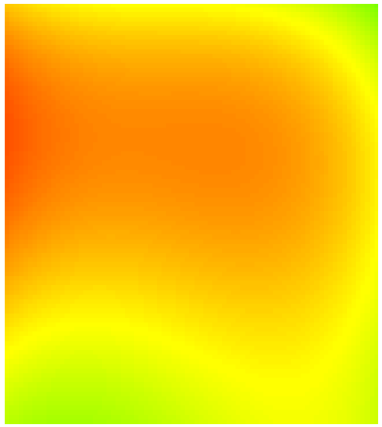
Goodness of fit (R2) = 45.42 %

Quadratic



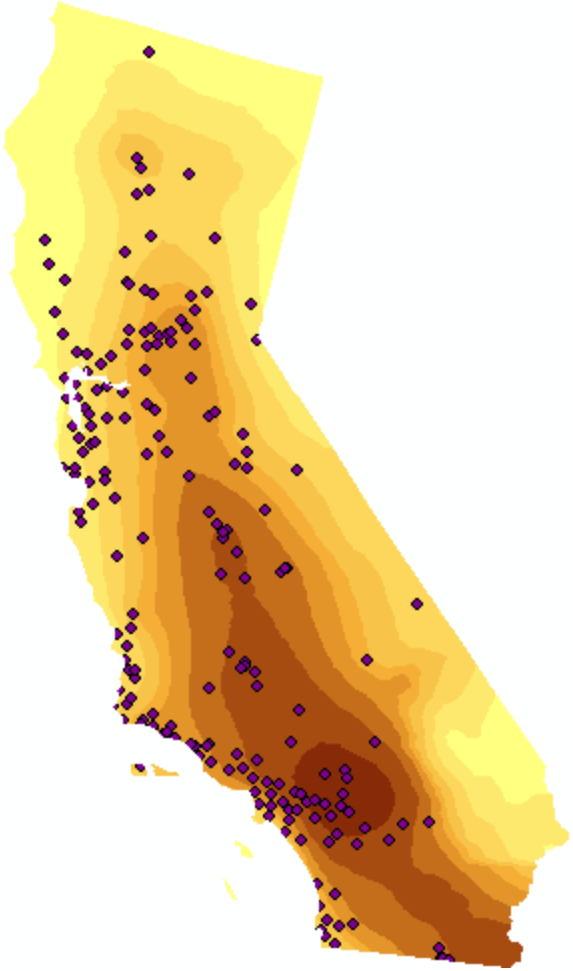
Goodness of fit (R2) = 82.11 %

Cubic



Goodness of fit (R2) = 92.72 %

Local



Global

Trend surfaces

- Uses a polynomial regression to fit a least-squares surface to the data points
- normally allows user control over the order of the polynomial used to fit the surface
- as the order of the polynomial is increased, the surface being fitted becomes progressively more complex
- higher order polynomial will not always generate the most accurate surface, it is dependent upon the data
- the lower the RMS error, the more closely the interpolated surface represents the input points
- most common order of polynomials is 1 through 3.
- edge effects may be severe.
- global, gradual, exact/approximate, and deterministic

6. Regression Techniques

- Global, inexact
- Like the classification approach, regression techniques make use of 'external' (i.e. predictor) variables.
- Example model is:

$$z(x) = b_0 + b_1 P_1 + b_2 P_2 + \varepsilon$$

Sample Linear Regression Model

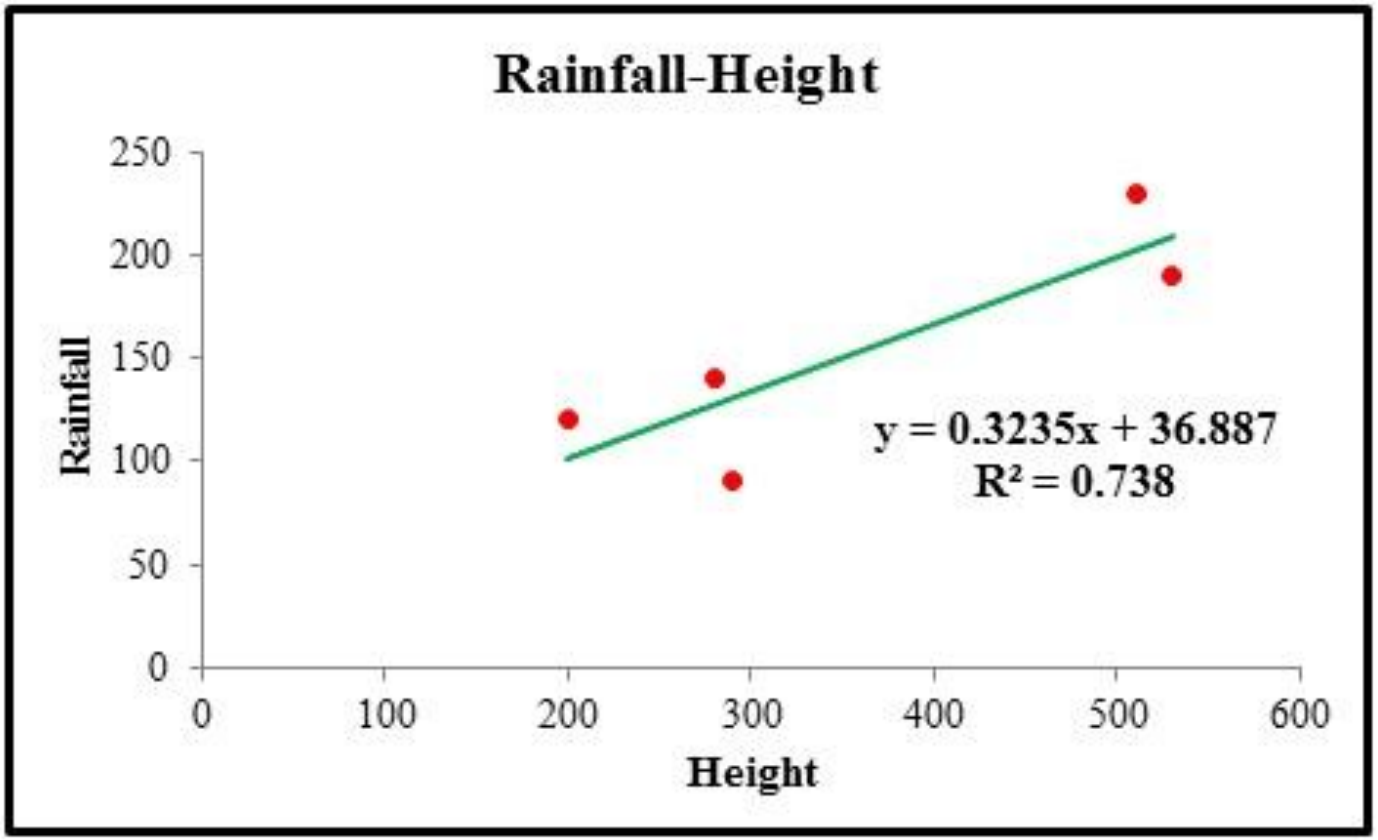
$\hat{\varepsilon}_i$ = Random error

Unsampled observation

Observed value

روش رگرسیون خطی (بارش با ارتفاع)

در این روش بین ارتفاع ایستگاه باران‌سنجی مطابق با DEM و عمق بارش در هر ایستگاه، خطی برازش می‌یابد. آنگاه با استفاده از معادله خط، مقادیر برای نقاط بدون اندازه‌گیری درون‌یابی می‌شود.



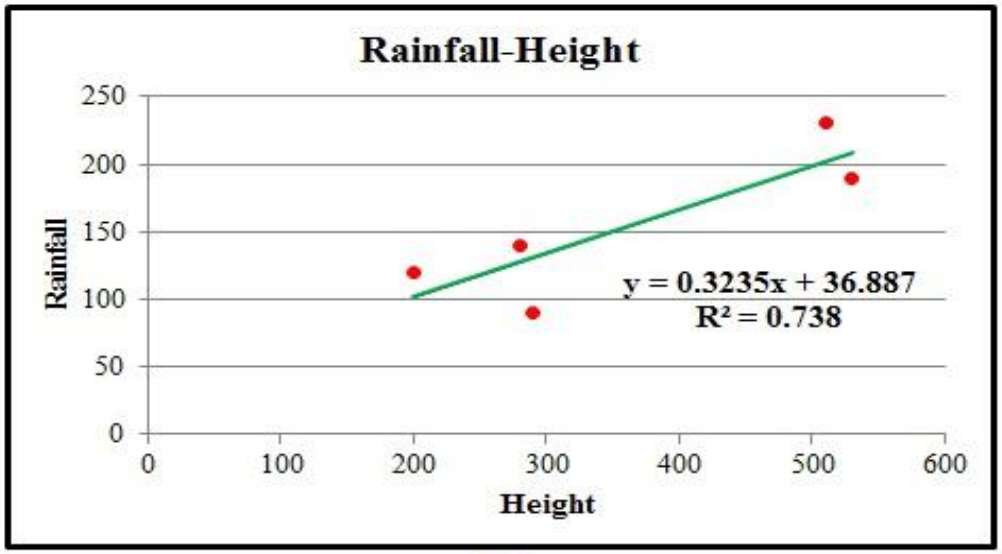
روش رگرسیون خطی (بارش با ارتفاع)

مثال: نقشه عمق بارندگی را با استفاده از DEM و ایستگاه‌های باران‌سنجی داده شده، به روش رگرسیون (بارش با ارتفاع) تعیین کنید. ابعاد سلول را ۱ کیلومتر در نظر بگیرید.

Height (m)								
500	440	430	660	430	220	400	300	210
450	480	430	550	530	380	240	240	320
220	300	430	420	210	190	170	410	190
380	280	330	350	520	480	105	510	230
190	120	230	280	170	190	80	270	120
120	400	160	70	220	430	400	310	200
290	110	100	150	180	160	140	160	300
440	70	340	170	130	190	120	320	400
70	100	85	170	520	480	200	220	110

Annual Rainfall Depth (mm)								
Point1	Point2	Point3	Point4	Point5	Point6	Point7	Point8	Point9
Point10	Point11	Point12	Point13	**190(C)**	Point15	Point16	Point17	Point18
Point19	Point20	Point21	Point22	Point23	Point24	Point25	Point26	Point27
Point28	**140(D)**	Point30	Point31	Point32	Point33	Point34	**230(A)**	Point36
Point37	Point38	Point39	Point40	Point41	Point42	Point43	Point44	Point45
Point46	Point47	Point48	Point49	Point50	Point51	Point52	Point53	Point54
90(E)	Point56	Point57	Point58	Point59	Point60	Point61	Point62	Point63
Point64	Point65	Point66	Point67	Point68	Point69	Point70	Point71	Point72
Point73	Point74	Point75	Point76	Point77	Point78	**120(B)**	Point80	Point81

روش رگرسیون خطی (بارش با ارتفاع)



Height	Rainfall	Interpolated Rainfall
530	190	208.342
280	140	127.467
510	230	201.872
290	90	130.702
200	120	101.587

y = 0.3235x + 36.887

حل: در جدول مقابل مقادیر درون‌یابی شده بارش در ایستگاه‌های باران‌سنجی مشاهده می‌شود. روش دقیق محسوب نمی‌شود.

روش رگرسیون خطی (بارش با ارتفاع)

حل: نتیجه نقشه برآورد شده از درون‌یابی به روش رگرسیون خطی بارش- ارتفاع مطابق با شکل زیر است:

Interpolated Annual Rainfall Depth In Other Pixel(mm)								
198.637	179.227	175.992	250.397	175.992	108.057	166.287	133.937	104.822
182.462	192.167	175.992	214.812	**190(C)**	159.817	114.527	114.527	140.407
108.057	133.937	175.992	172.757	104.822	98.352	91.882	169.522	98.352
159.817	**140(D)**	143.642	150.112	205.107	192.167	70.8545	**230(A)**	111.292
98.352	75.707	111.292	127.467	91.882	98.352	62.767	124.232	75.707
75.707	166.287	88.647	59.532	108.057	175.992	166.287	137.172	101.587
90(E)	72.472	69.237	85.412	95.117	88.647	82.177	88.647	133.937
179.227	59.532	146.877	91.882	78.942	98.352	75.707	140.407	166.287
59.532	69.237	64.3845	91.882	205.107	192.167	**120(B)**	108.057	72.472

تمرین ۳: روش رگرسیون

DEM و عمق بارندگی در ۵ ایستگاه باران‌سنجی مطابق شکل داده شده است. به کمک روش رگرسیون، مقدار بارش را در سایر سلول‌ها تعیین نمایید. ابعاد سلول را ۱ کیلومتر در نظر بگیرید.

Annual Rainfall Depth (mm)								
						140		
	200							
								80
		120						
						120		

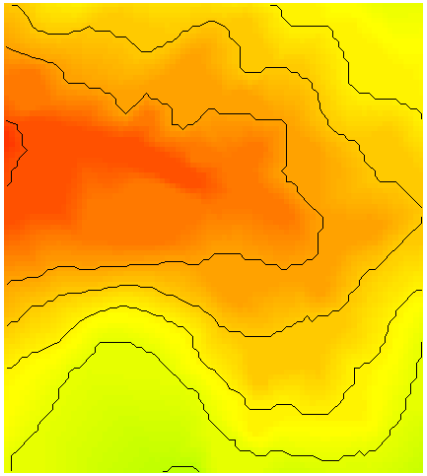
DEM (m)								
620	600	590	650	400	300	320	340	290
600	528	516	792	516	288	480	360	252
540	576	516	660	636	456	288	288	384
264	360	516	504	252	228	204	492	228
456	336	396	420	624	576	126	612	276
228	144	276	336	204	228	126	324	144
144	480	192	120	264	516	480	372	240
348	132	120	180	216	192	168	192	360
528	90	408	204	156	228	144	384	480
84	120	102	204	624	576	240	140	132

Common interpolation problems

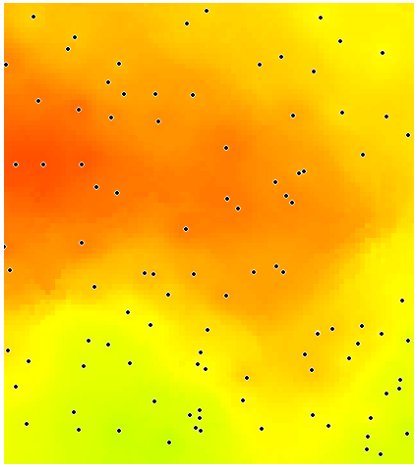
- Input data uncertainty
  - Too few data points
  - Limited or clustered spatial coverage
  - Uncertainty about location and/or value
- Edge effects
  - Need data points outside study area (extrapolation)
  - improve interpolation and avoid distortion at boundaries

Effects of data uncertainty

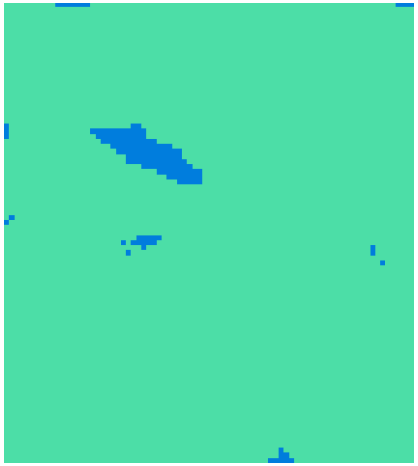
Original surface



Interpolation based on 100 points



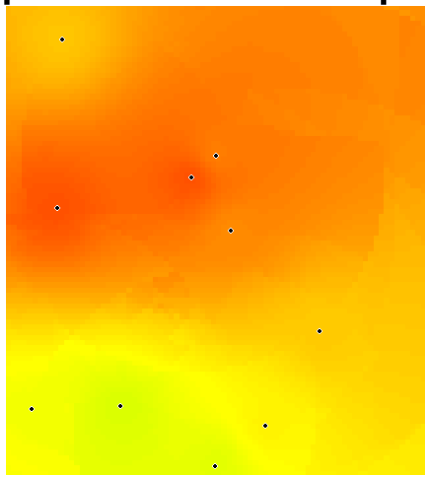
Error map



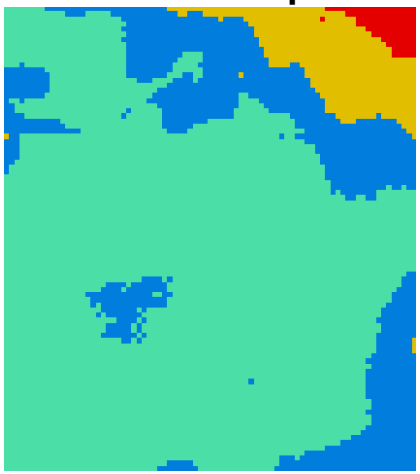
Low

High

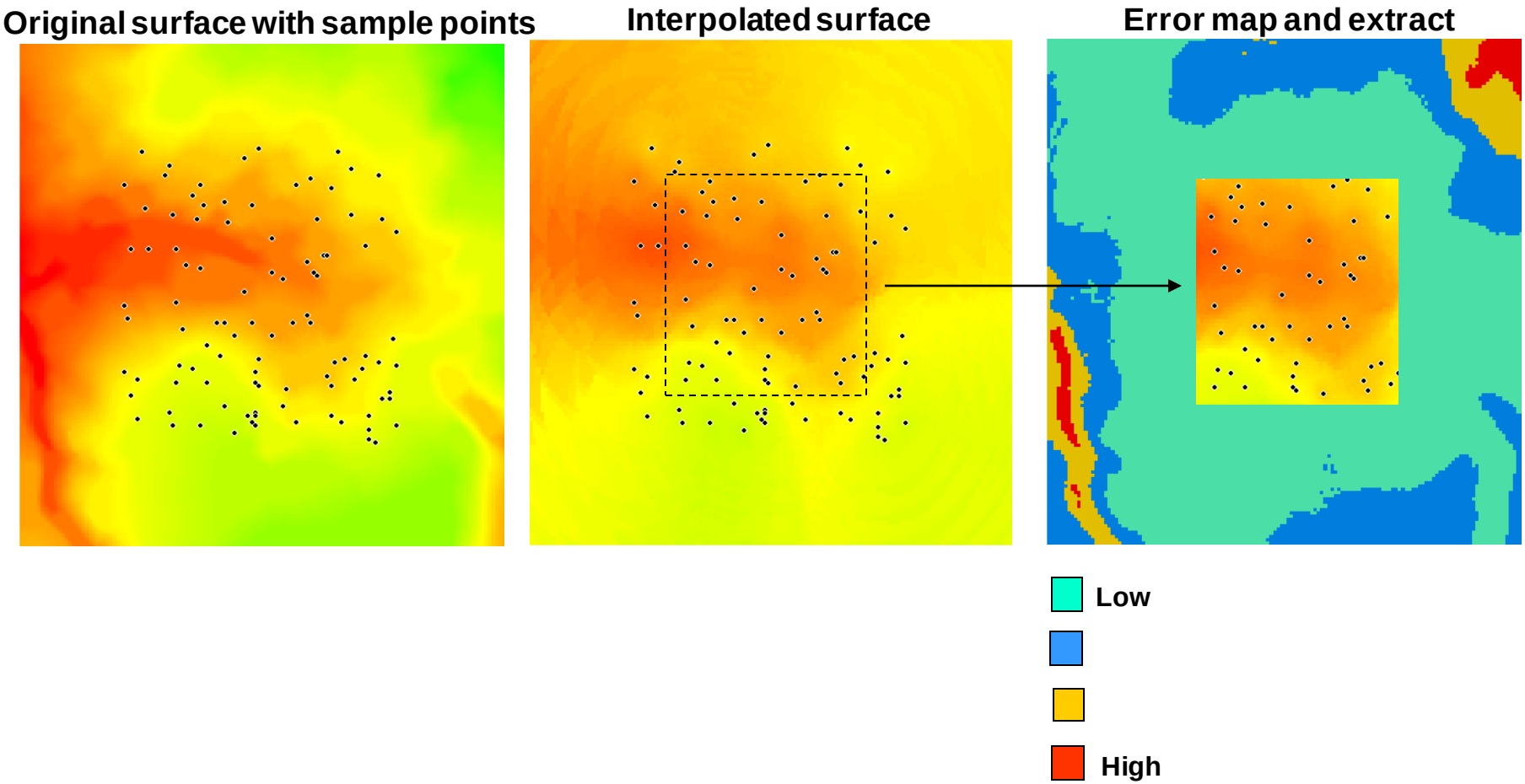
Interpolation based on 10 points



Error map

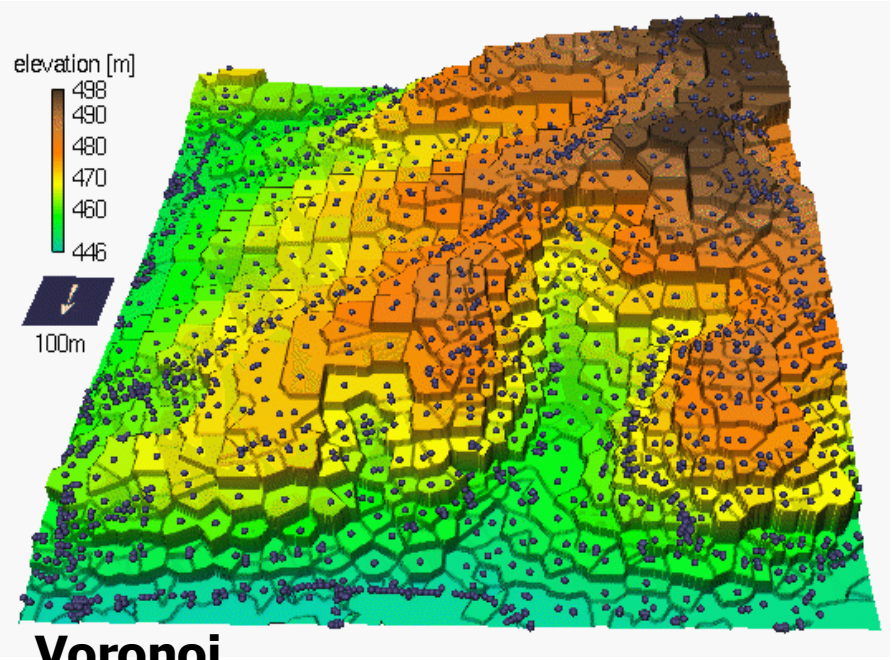


Edge effects

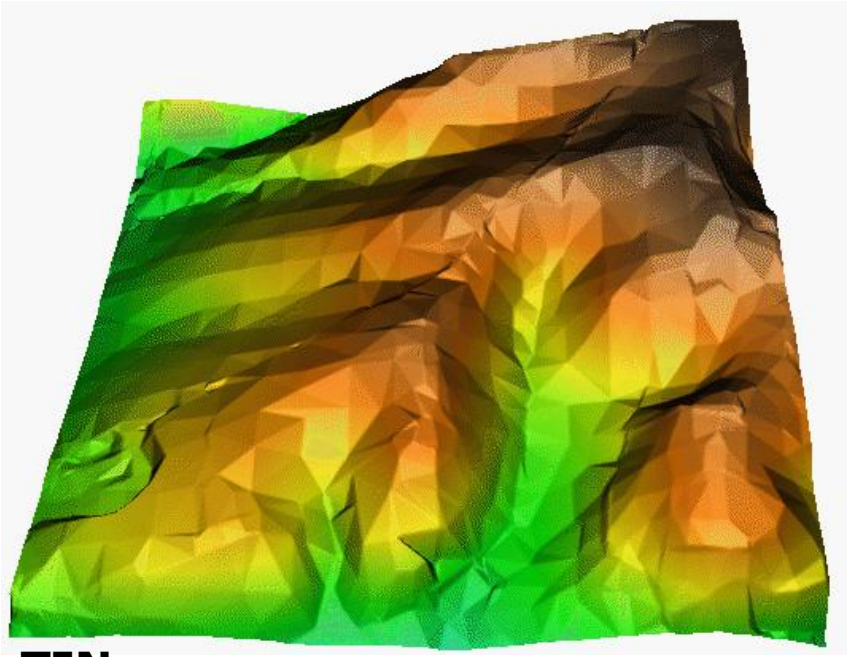


Summery

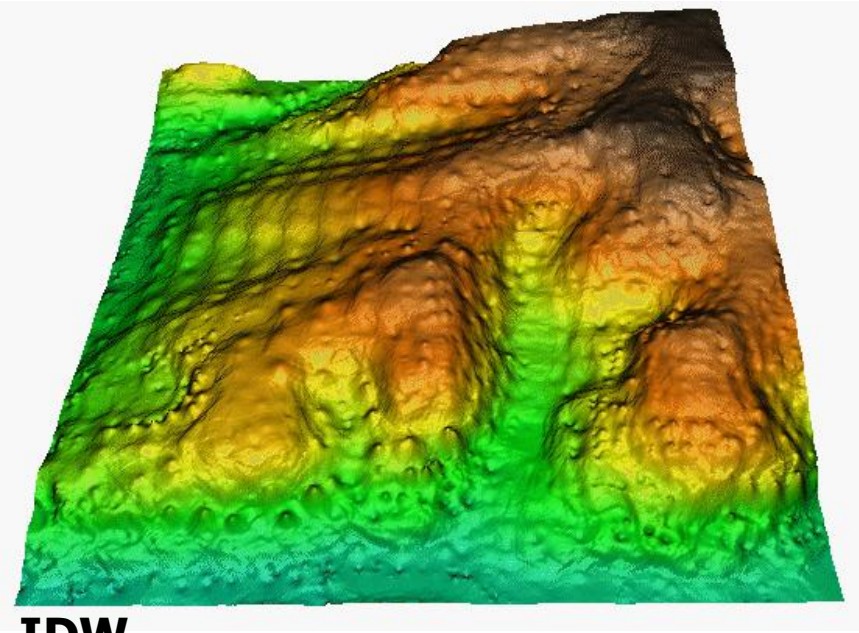
- Interpolation of point data is an important skill in of environmental and hydrological studies.
- Many methods classified by
  - local/global, approximate/exact, gradual/abrupt and deterministic/stochastic
  - choice of method is crucial to success
- Error and uncertainty
  - poor input data
  - poor choice/implementation of interpolation method



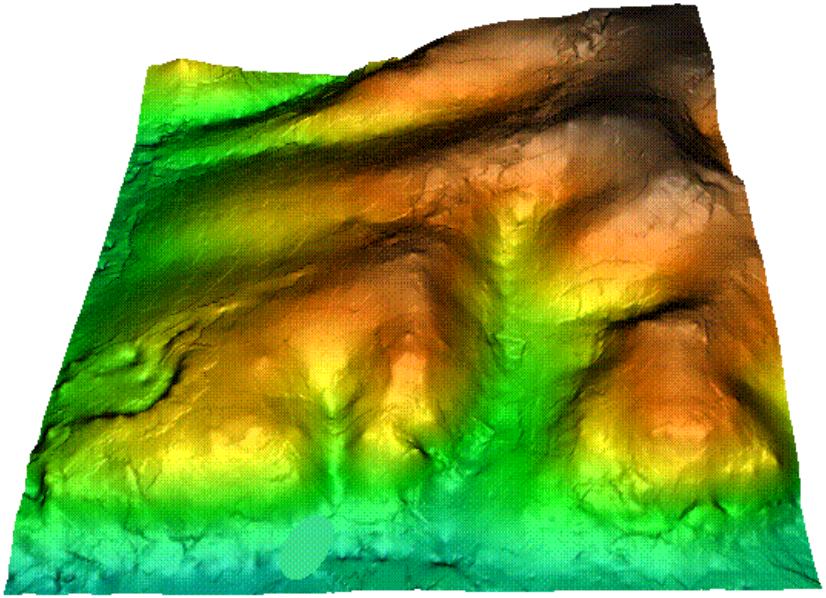
Voronoi



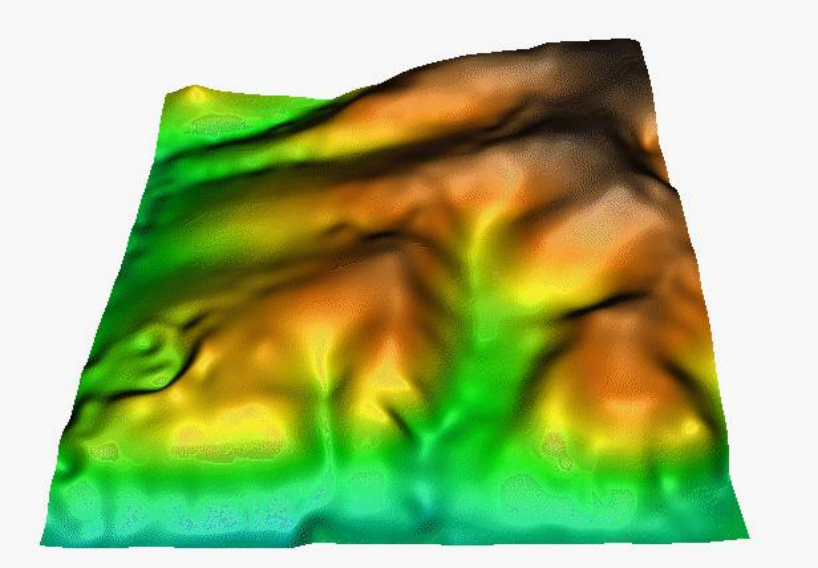
TIN



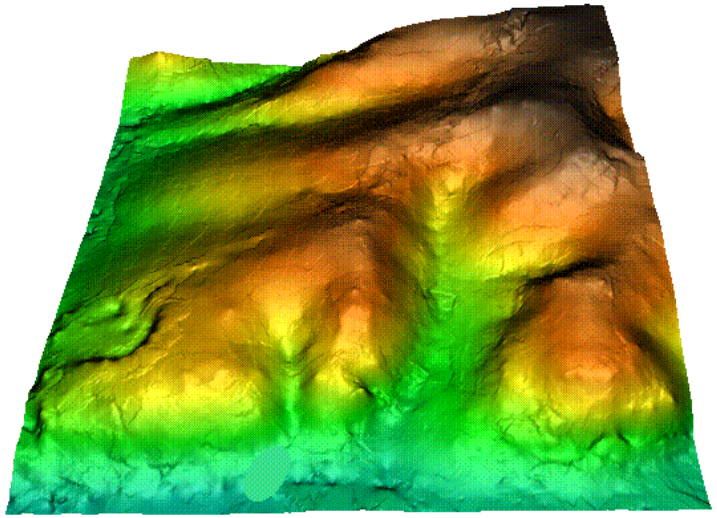
IDW



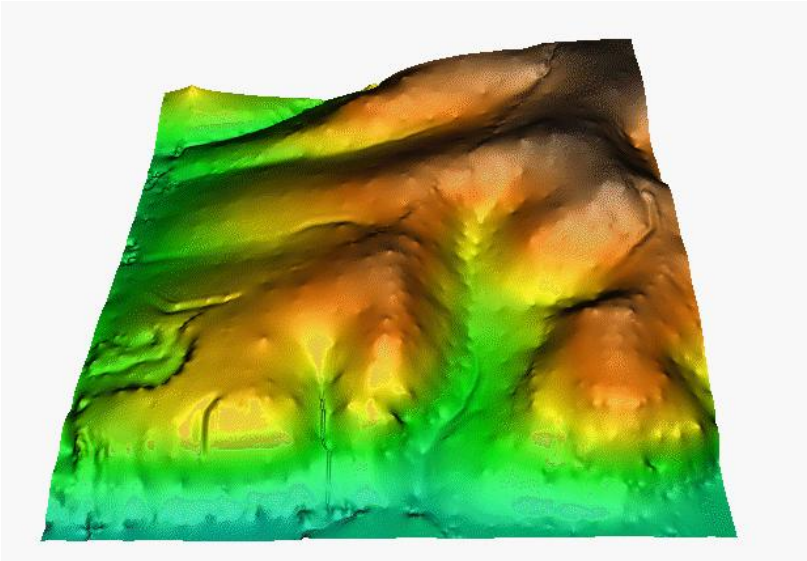
Kriging



Surf.tps (GRASS)



Thin plate spline



TOPOGRID

Interpolation Comparison

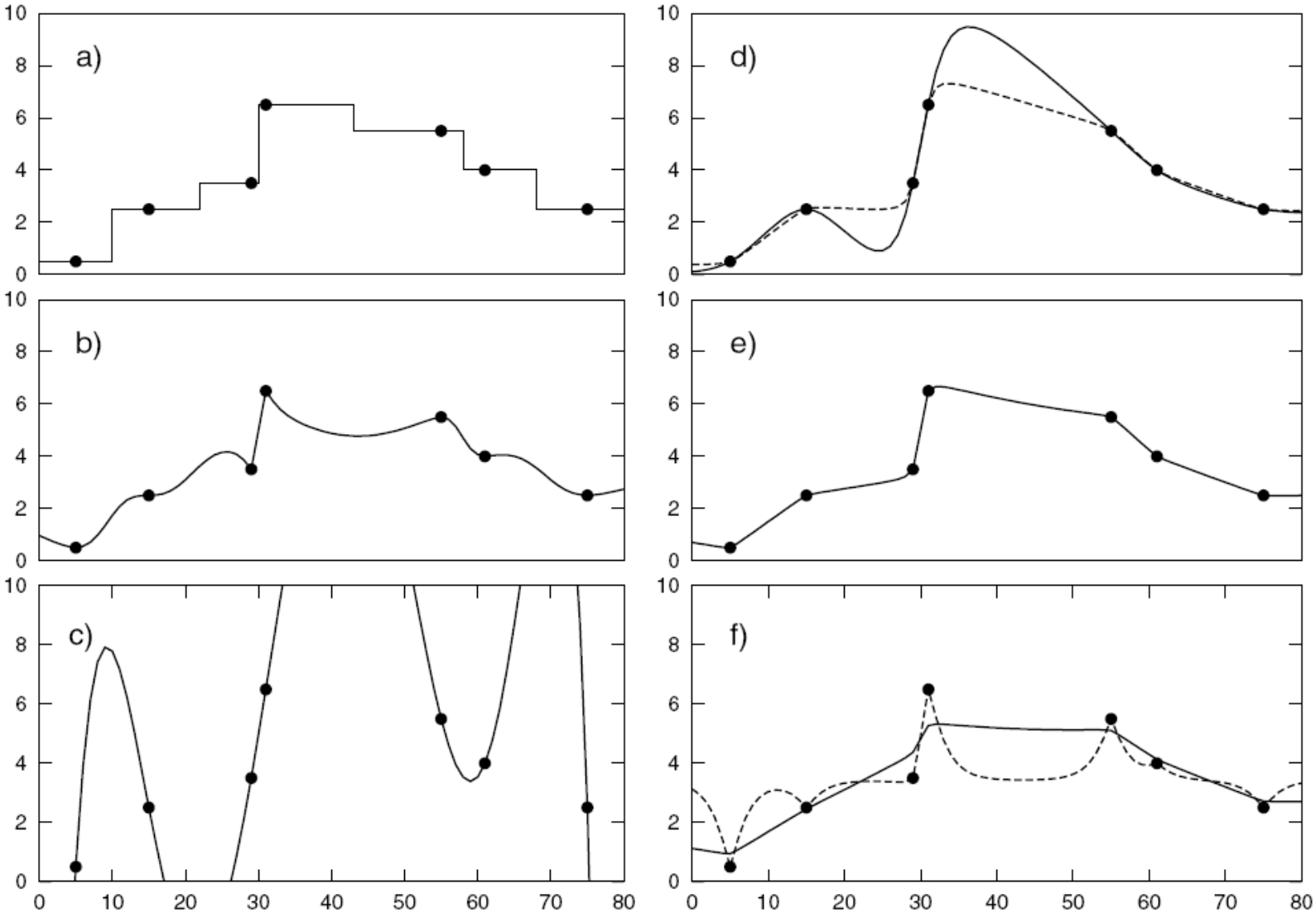


Figure 2.4. Comparison of interpolation methods for a one-dimensional example: (a) Thiessen method; (b) inverse distance squared; (c) example of overfitting using a sixth-order polynomial; (d) thin plate splines with different tension parameters; (e) kriging (zero nugget, large range); (f) kriging (solid line: large nugget, large range; dashed line: zero nugget, very short range).

Grayson, R. and G. Blöschl, ed. (2000)

Method	Assumption	Univariable / Multivariable	Deterministic/ stochastic	Local/ global	Exact/ inexact	Abrupt/ gradual	Limitation of the procedure	Computing load	Output data structure	Suitability
Nearest neighbours (NN)	Best local predictor is nearest data point	Univariable	Deterministic	Local	Exact	Abrupt	No error assessment, only one data point per polygon. Tessellation pattern depends on distribution of data.	Small	Polygons or gridded surface	Nominal data from point observations
Triangulation (TIN)	Best local predictor is data points on the surrounding triangle	Univariable	Deterministic	Local	Exact	Abrupt	No error assessment. TIN pattern depends on distribution of data and there a few ways to form triangulation and no one is better than any other.	Small	Triangles or Gridded surface	Quick interpolation from sparse data on regular or irregularly spaced samples.
Natural neighbours (NaN)	Best local predictor is data points in the surrounding polygons	Univariable	Deterministic	Local	Exact	Gradual or abrupt	No error assessment	Small	Gridded surface	Quick interpolation from sparse data on regular or irregularly spaced samples.
Inverse distance weighting (IDW)	underlying surface is smooth	Univariable	Deterministic	Local	Inexact (but can be forced to be exact)	Gradual	No error assessment. Results depend on size of search window and choice of weighting parameter. Poor choice of window can give artefacts when used with high data densities such as digitised contours.	Small	Gridded surface, contours	Quick interpolation from sparse data on regular grid or irregularly spaced samples.
Regression models (LM)	Samples are independent, normal and homogeneous in variance	Univariable/ Multivariable	Stochastic	Global	Inexact	Gradual if inputs have gradual variation	Results depend on the fit of the regression model and the quality and detail of the input data surfaces. Error assessment possible if input errors are known.	Small	Polygons or continuous, gridded surface	Simple numerical modelling of expensive data when better methods are not available or budgets are limited
Trend surface analysis (TSA)	Phenomenological explanation of trend, normally distributed data	Multivariable	Stochastic	Global	Inexact	Gradual	Physical meaning of trend may be unclear. Outliers and edge effects may distort surface. Error assessment limited to goodness of fit.	Small	continuous, gridded surface	Quick assessments and removal of spatial trend

Method	Assumption	Univariable / Multivariable	Deterministic/ stochastic	Local/ global	Exact/ inexact	Abrupt/ gradual	Limitation of the procedure	Computing load	Output data structure	Suitability
Splines & Local trend surfaces (LTS)	Best local predictor is the nearest data point and data normality	Multivariable	Stochastic	Local	Inexact	Gradual	Results depend on span parameter and detail of the input data surfaces.	Moderate	continuous, gridded surface	Quick interpolation from sparse data on regular grid or irregularly spaced samples.
Classification (Cl)	Homogeneity within boundaries	Univariable	Deterministic "soft" information	Global	Inexact	Abrupt	Delineation of areas and classes may be subjective. Error assessment limited to within-class standard deviations.	Small	Classified polygons	Quick assessments when data are sparse Removing systematic differences before continuous interpolation from data points
Regression tree (CART)	Phenomenological explanation of variance	Multivariable	Stochastic	Global	Inexact	?	?	Small	Gridded surface	
Thin plat splines (TPS)	Underlying surface is smooth everywhere	Univariable/ Multivariable	Deterministic	Local	Exact	Gradual	Goodness of fit possible, but within the assumptions that the fitted surface is perfectly smooth.	Small	Gridded surface, contours	Quick interpolation (univariate or multivariate) of digital elevation data and related attributes to create digital elevation models (DEM) from moderately detailed data
Kriging*	Interpolated surface is smooth. Statistical stationarity and the intrinsic hypothesis.	Univariable/ Multivariable	Stochastic	Local	Exact	Gradual	Error assessment depends on variogram and distribution of data points and size of interpolated blocks. Requires care when modelling spatial correlation structures.	Moderate	Gridded surface	When data are sufficient to compute variograms, kriging provides a good interpolator for sparse data. Binary and nominal data can be interpolated with Indicator kriging. Soft information can also be incorporated as trends or stratification. Multivariate data can be interpolated with co-kriging.

* Of the kriging methods, BK is an inexact interpolator.

مقایسه: روش اعتباریابی حذفی (Cross Validation)

- به منظور ارزیابی روش‌های مختلف درون‌یابی و بررسی میزان خطای روش‌های مختلف و مقایسه آن‌ها با یکدیگر، از روش اعتباریابی حذفی استفاده می‌شود.
- در روش اعتباریابی حذفی، هر یک از مشاهدات به نوبت حذف شده و بر مبنای سایر مشاهدات، مقدار متغیر موردنظر در نقطه حذف شده به کمک روش‌های مورد مقایسه برآورد می گردد.
- این عمل را برای کلیه ایستگاه‌ها انجام شده و مقدار خطای هر روش تعیین می‌شود.
- احتمالا شعاع تاثیر داده شده برای ارزیابی کافی نیست.

Common Error Criteria

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Z^*(x_i) - Z(x_i)\right|$$

$$MBE = \frac{1}{n}\sum_{i=1}^{n}\left(Z^*(x_i) - Z(x_i)\right)$$

مثال ارزیابی روش‌های رگرسیون و IDW

مثال: روش‌های رگرسیون و IDW را برای ایستگاه‌های باران‌سنجی زیر به وسیله روش Cross Validation و معیارهای MAE و MBE ارزیابی کنید.

Height (m)								
500	440	430	660	430	220	400	300	210
450	480	430	550	530	380	240	240	320
220	300	430	420	210	190	170	410	190
380	280	330	350	520	480	105	510	230
190	120	230	280	170	190	80	270	120
120	400	160	70	220	430	400	310	200
290	110	100	150	180	160	140	160	300
440	70	340	170	130	190	120	320	400
70	100	85	170	520	480	200	220	110

Annual Rainfall Depth (mm)								
Point1	Point2	Point3	Point4	Point5	Point6	Point7	Point8	Point9
Point10	Point11	Point12	Point13	**190(C)**	Point15	Point16	Point17	Point18
Point19	Point20	Point21	Point22	Point23	Point24	Point25	Point26	Point27
Point28	**140(D)**	Point30	Point31	Point32	Point33	Point34	**230(A)**	Point36
Point37	Point38	Point39	Point40	Point41	Point42	Point43	Point44	Point45
Point46	Point47	Point48	Point49	Point50	Point51	Point52	Point53	Point54
90(E)	Point56	Point57	Point58	Point59	Point60	Point61	Point62	Point63
Point64	Point65	Point66	Point67	Point68	Point69	Point70	Point71	Point72
Point73	Point74	Point75	Point76	Point77	Point78	**120(B)**	Point80	Point81

ابعاد سلول را ۱ کیلومتر در نظر بگیرید.

در روش IDW فاصله تاثیر را ۶ کیلومتر و توان را ۱ درنظر بگیرید.

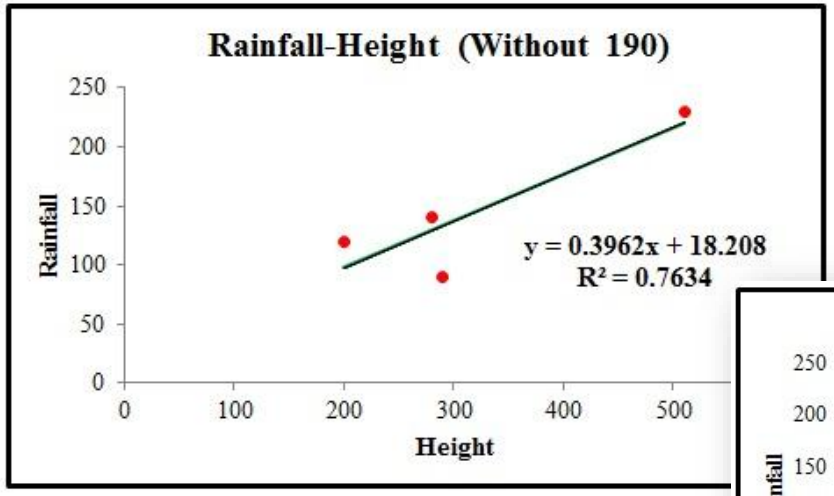
مثال ارزیابی روش‌های رگرسیون و IDW

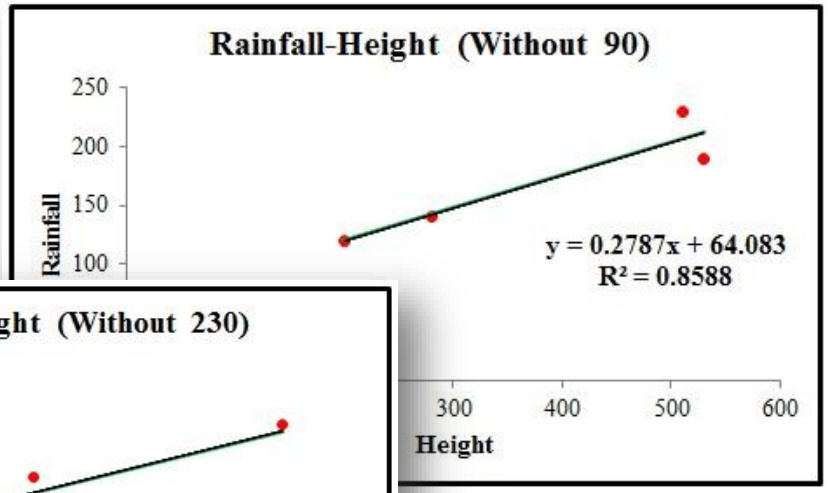
حل:

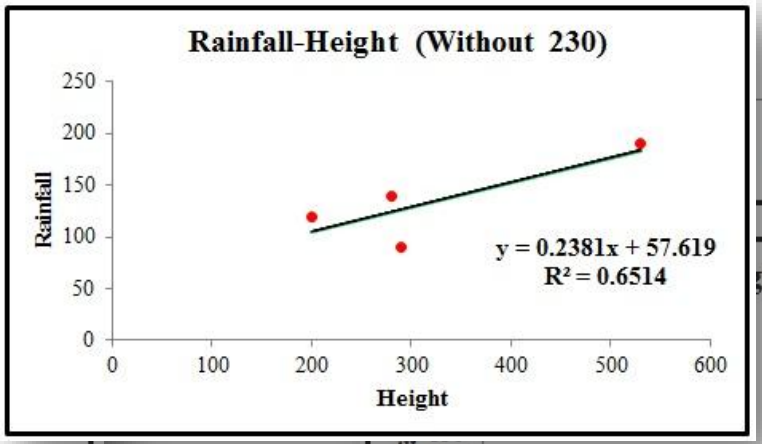
- اعتباریابی حذفی در روش رگرسیون – در این روش ابتدا بایستی مشاهدات موجود (ایستگاه‌های بارش) را حذف نموده و با استفاده از سایر ایستگاه‌ها (چهار ایستگاه باقیمانده) رگرسیون تکرار می‌شود.
- با معادله حاصل، مقدار عمق بارش در نقطه حذف شده را به دست می‌آوریم.
- این عمل را برای کلیه ایستگاه‌ها انجام داده و مقدار خطاهای MAE و MBE تعیین می شود.

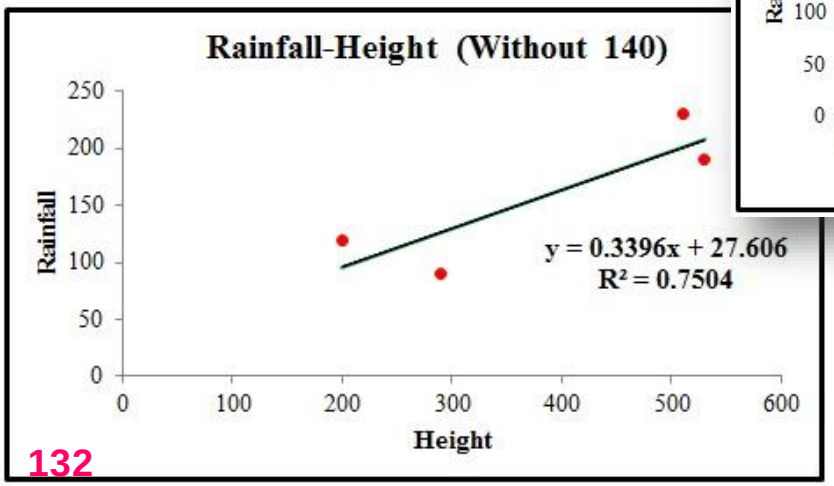
مثال ارزیابی روش های رگرسیون و IDW

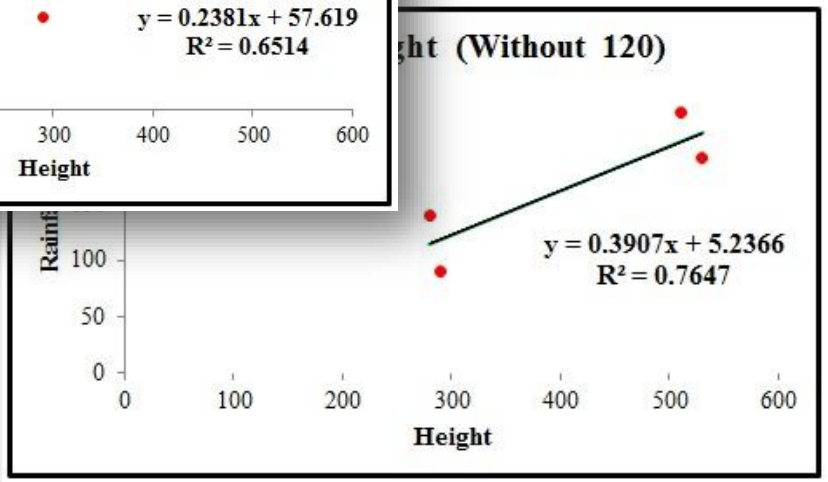
حل: اعتباریابی حذفی در روش رگرسیون











مثال ارزیابی روش‌های رگرسیون و IDW

حل: نتایج حاصل از روش رگرسیون به صورت زیر است:

Height	Rainfall	Predicted Rainfall	Z*-Z	|Z*-Z|	MAE	MBE
530	190	228.19	38.19	38.19	39.60	-2.36
280	140	122.69	-17.31	17.31		
510	230	179.05	-50.95	50.95		
290	90	144.91	54.91	54.91		
200	120	83.38	-36.62	36.62		

مثال ارزیابی روش‌های رگرسیون و IDW

حل:

- اعتباریابی حذفی در روش IDW- در این روش ابتدا بایستی عمق بارش مشاهداتی موجود را در هر یک از ایستگاه‌های بارش حذف نموده و با استفاده از اعماق بارش مشاهداتی چهار ایستگاه باقیمانده، به کمک روش درون‌یابی IDW، مقدار بارش نقطه حذف شده را به دست آورد.
- این عمل برای تمامی ایستگاه‌ها انجام می‌شود.

مثال ارزیابی روش‌های رگرسیون و IDW

حل: نمونه‌ای از محاسبات اعتباریابی حذفی با روش درون‌یابی IDW

نقطه حذفی	Control Points	Rain Depth	Distance	Inverse Distance	Weight	Weighted Value
190	1	230	3.61	0.277	0.39	89.74
	2	140	3.61	0.277	0.39	54.62
	3	90	6.40	0.156	0.22	19.77
	4					
	5					
	Σ			0.711	1	**164.13**

نقطه حذفی	Control Points	Rain Depth	Distance	Inverse Distance	Weight	Weighted Value
140	1	190	3.61	0.277	0.31	58.44
	2	90	3.16	0.316	0.35	31.56
	3	230	6.00	0.167	0.18	42.51
	4	120	7.07	0.141	0.16	18.82
	5					
	Σ			0.902	1	**151.34**

نقطه حذفی	Control Points	Rain Depth	Distance	Inverse Distance	Weight	Weighted Value
230	1	190	3.61	0.277	0.43	82.32
	2	140	6.00	0.167	0.26	36.45
	3	120	5.10	0.196	0.31	36.76
	4					
	5					
	Σ			0.640	1	**155.54**

مثال ارزیابی روش‌های رگرسیون و IDW

حل: اعتباریابی حذفی در روش IDW– محاسبه خطاهای MAE وMBE

IDW D=6km						
Height	Rainfall	Predicted Rainfall	Z*-Z	|Z*-Z|	MAE	MBE
530	190	164.13	-25.87	25.87	50.10	9.97
280	140	151.34	11.34	11.34		
510	230	155.54	-74.46	74.46		
290	90	156.53	66.53	66.53		
200	120	192.29	72.29	72.29		

مثال ارزیابی روش های رگرسیون و IDW

حل مثال: مقایسه دو روش رگرسیون و IDW

Height	Rainfall	Predicted Rainfall	Z*-Z	|Z*-Z|	MAE	MBE
530	190	228.19	38.19	38.19	39.60	-2.36
280	140	122.69	-17.31	17.31		
510	230	179.05	-50.95	50.95		
290	90	144.91	54.91	54.91		
200	120	83.38	-36.62	36.62		

روش رگرسیون

IDW D=6km						
Height	Rainfall	Predicted Rainfall	Z*-Z	|Z*-Z|	MAE	MBE
530	190	164.13	-25.87	25.87	50.10	9.97
280	140	151.34	11.34	11.34		
510	230	155.54	-74.46	74.46		
290	90	156.53	66.53	66.53		
200	120	192.29	72.29	72.29		

روش IDW

مقادیر خطاهای MAE و MBE بدست آمده از روش رگرسیون، کمتر از روش IDW بوده و بنابراین روش رگرسیون در این مسئله، روشی مناسب‌تر نسبت به روش IDW است.

روش اعتباریابی حذفی (Cross Validation)

راهنمایی:

به تراکم ایستگاه‌های اندازه‌گیری توجه نمائید.

سوال: با ذکر علت، توضیح دهید که روش اعتباریابی حذفی در کدامیک از روش‌های درون‌یابی که تا کنون ذکر شد، قابل استفاده و در کدامیک قابل استفاده نمی‌باشد؟

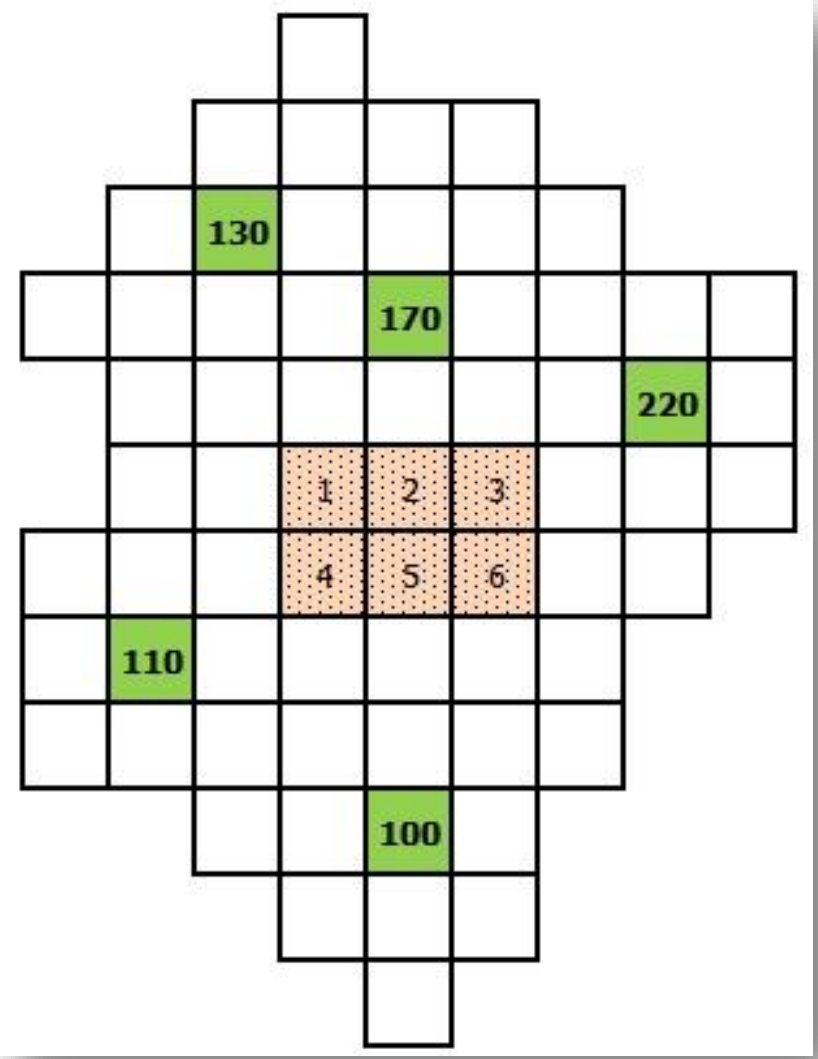
Regression

TIN

IDW

Thiessen

تمرین ۴: روش IDW و ارزیابی روش



عمق بارندگی در ۵ ایستگاه داده شده است. نقشه بارش را به روش IDW با در نظر گرفتن فاصله تاثیر ۵ کیلومتر و توان ۲ بدست آورید. ابعاد سلول را ۱ کیلومتر در نظر بگیرید. هم چنین روش IDW را به وسیله روش Cross Validation و معیارهای MAE و MBE ارزیابی کنید.

تمرین ۵: عمومی

با استفاده از ۵ ایستگاه باران‌سنجی و DEM داده شده مطلوبست:

الف) عمق باران به روش‌های زیر در سلول‌های شماره A تا X

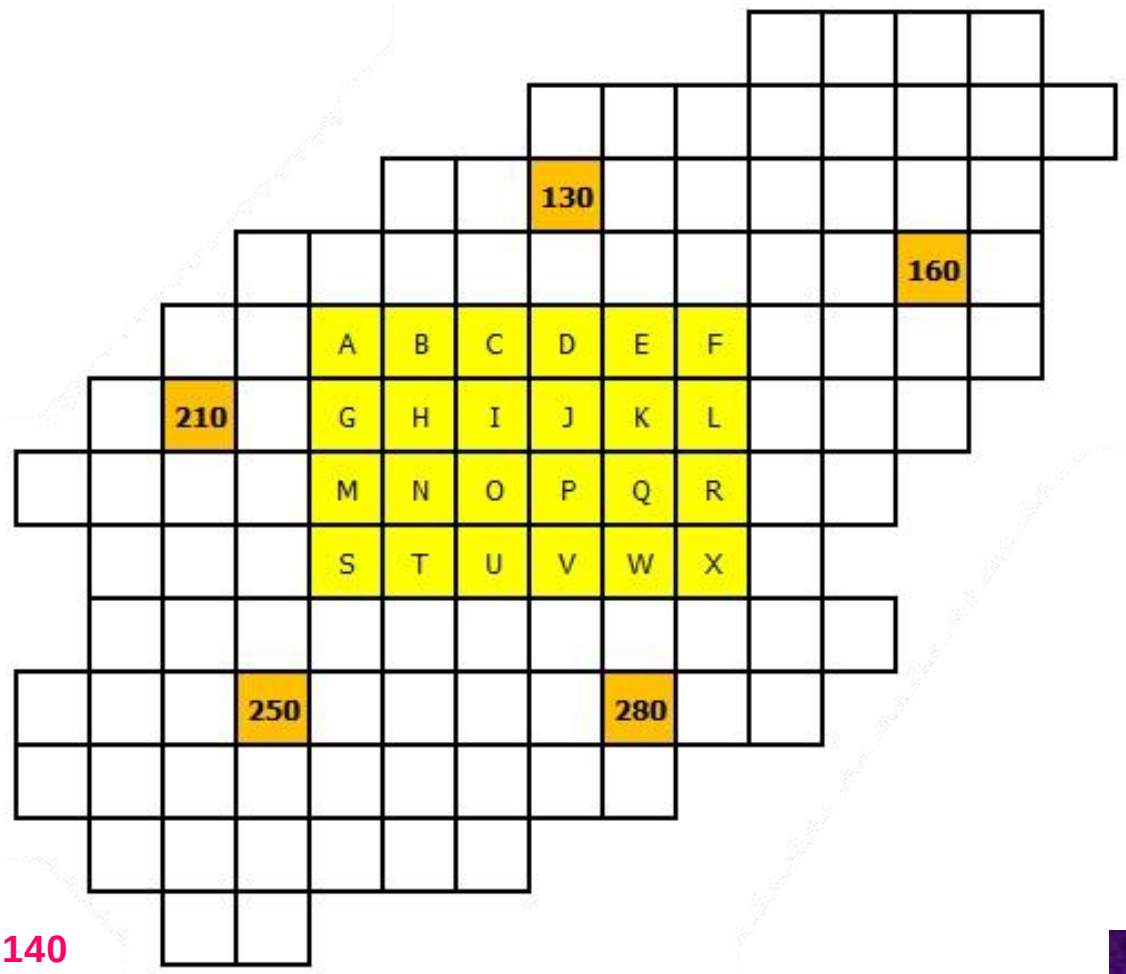
- رگرسیون (بارش با ارتفاع)
- مثلث‌بندی
- IDW
- (D= 10 km , exponent = 1)
- (D= 5 km , exponent = 2)

ب) ارزیابی و مقایسه روش‌های فوق به وسیله Cross Validation

ج) مقایسه روش‌های فوق با محاسبه شاخص‌های آماری زیر

- میانگین حوزه
- Max , Min حوزه
- SD

طول و عرض سلول‌ها را ۱ کیلومتر در نظر بگیرید.



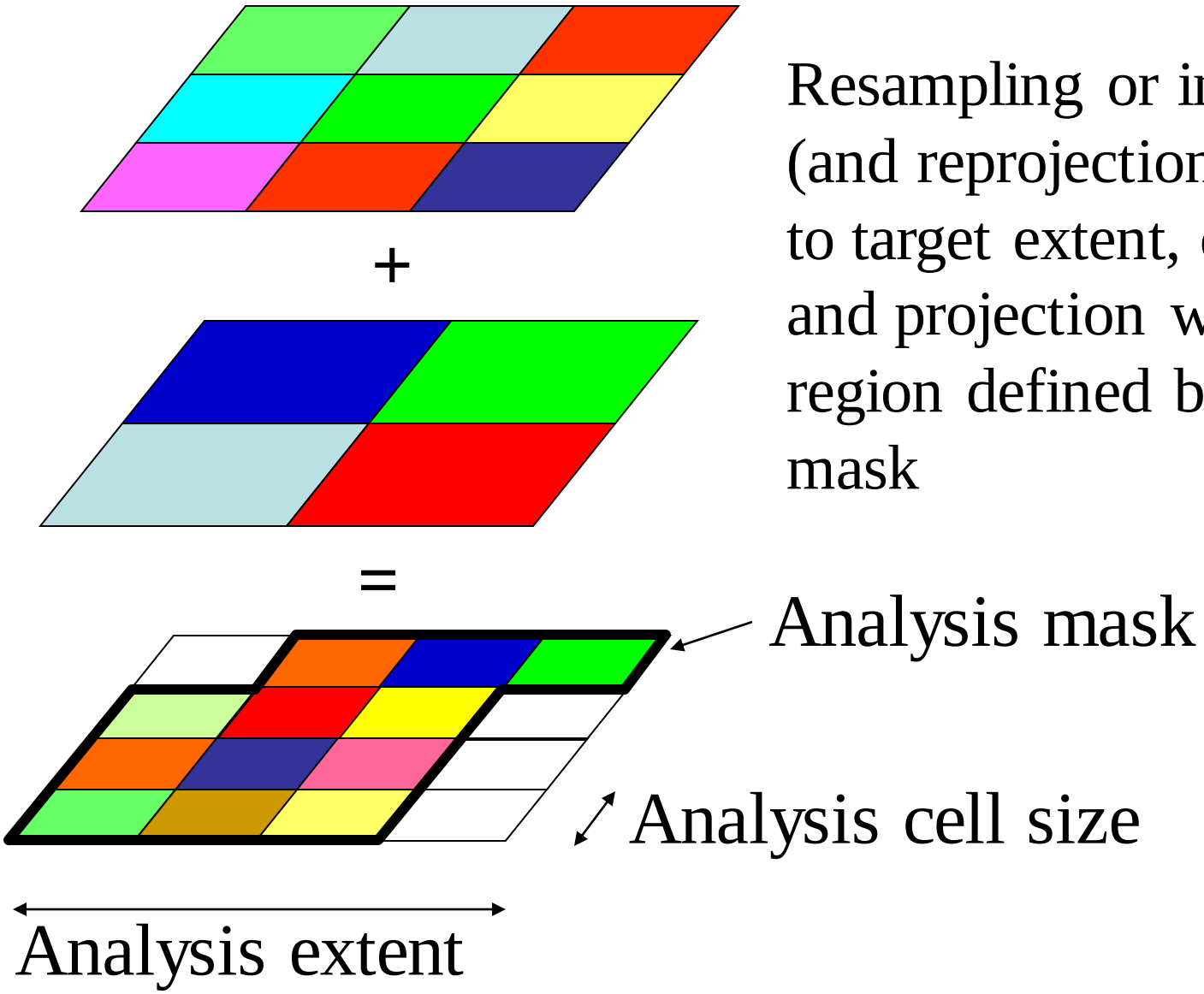
تمرین ۵: عمومی

اطلاعات DEM

									250	240	230	220	
						420	410	400	210	310	320	290	280
				480	450	430	600	440	220	400	300	210	
			440	470	420	560	540	400	250	240	320	310	200
		210	290	400	420	220	200	190	380	210	170	140	120
	370	260	320	340	520	480	105	510	230	220	180	150	
170	130	210	260	200	190	200	270	120	120	100	90		
	130	380	160	110	220	430	400	310	200	110			
	280	110	100	150	180	160	140	160	300	130	100		
420	110	300	180	130	190	120	320	400	200	160			
90	100	120	180	520	480	200	220	110					
	80	90	100	200	250	190							
		80	70										

HEIGHT (m)

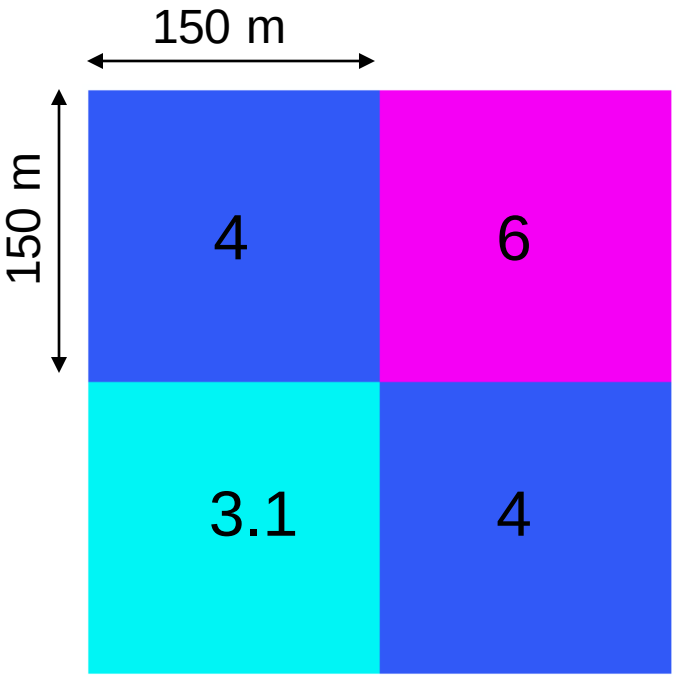
Raster calculation – Resample



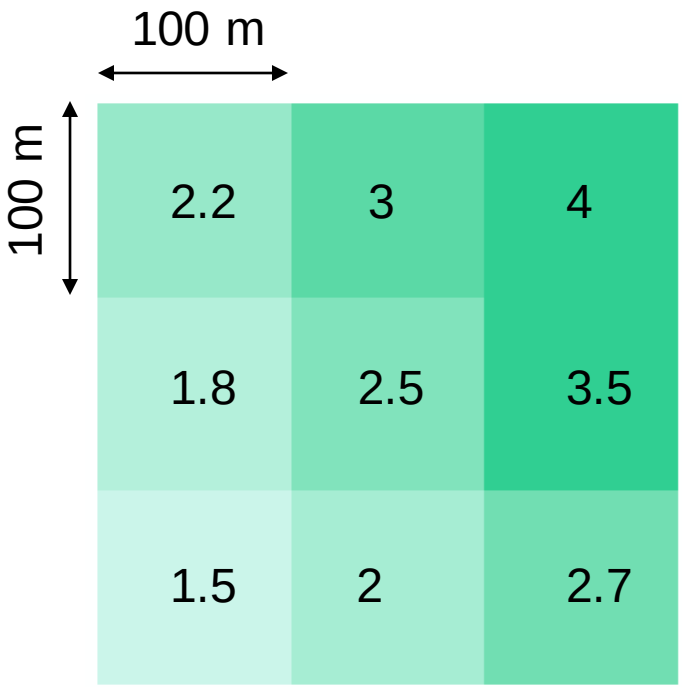
Resampling or interpolation (and reprojection) of inputs to target extent, cell size, and projection within region defined by analysis mask

Runoff Generation Raster Calculation Example

Precipitation, P cm/hr



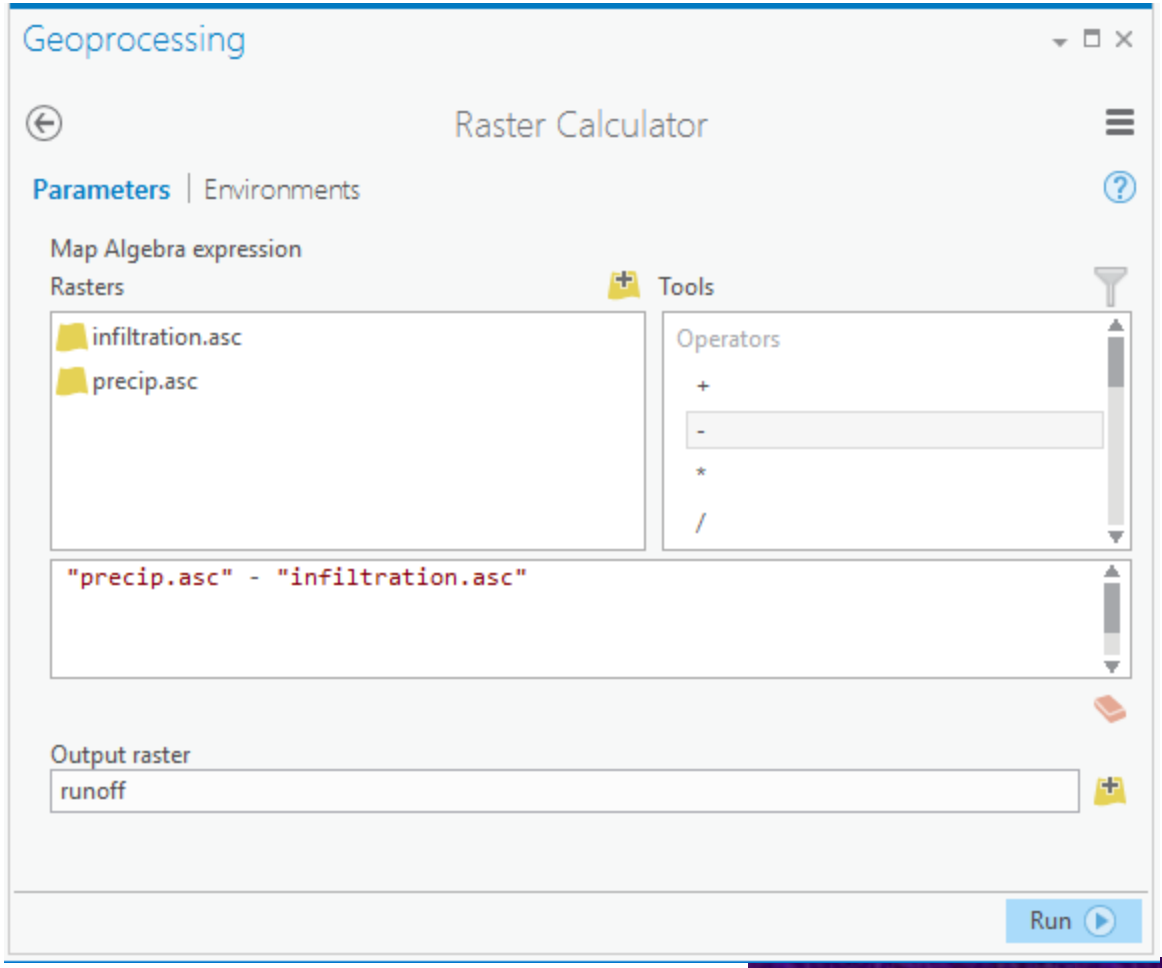
Infiltration Rate, f cm/hr



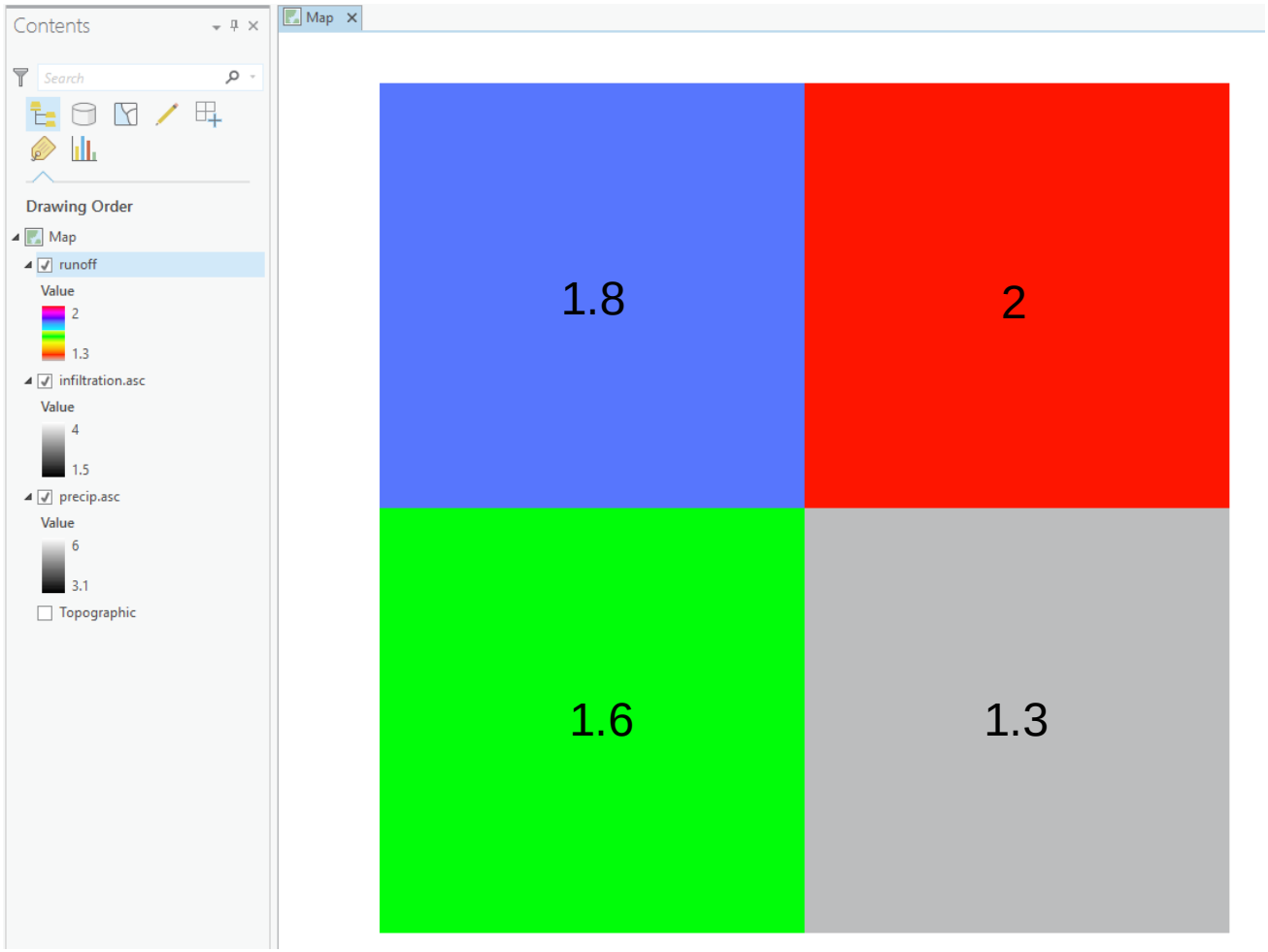
Runoff R = P-f

Runoff calculation using Raster Calculator

"precip.asc" - "infiltration.asc"

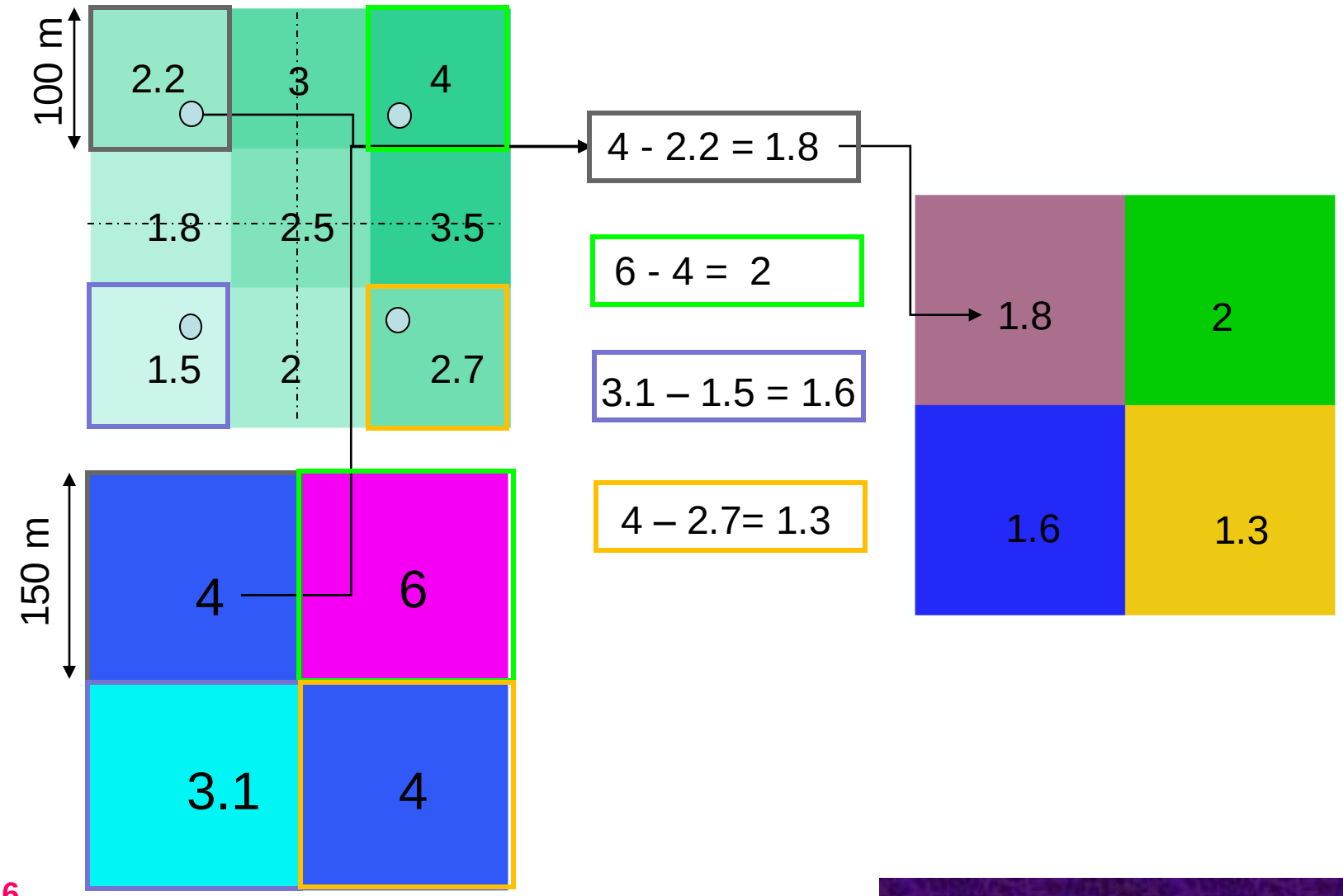


The Result

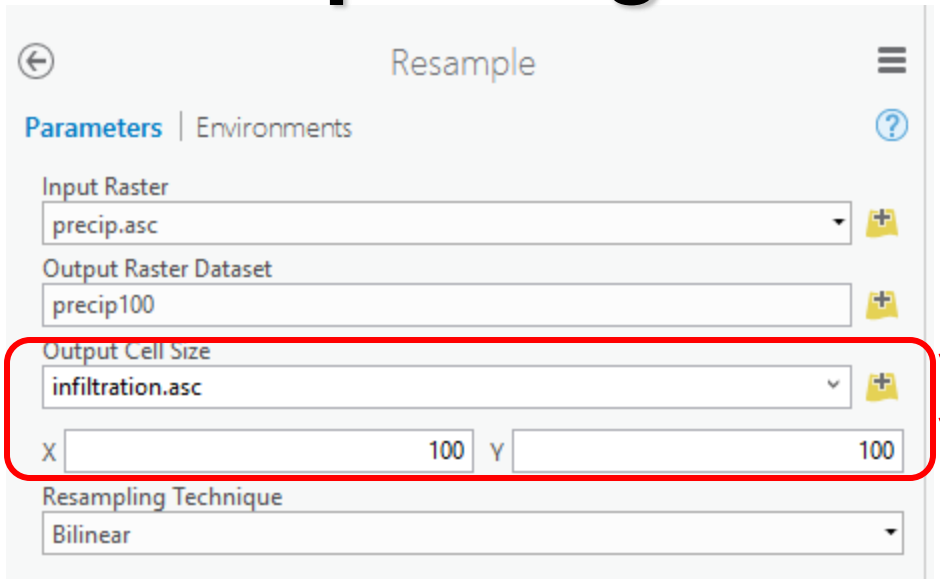


- Outputs are on 150 m grid.
- How were values obtained ?

Nearest Neighbor Resampling with Cellsize Maximum of Inputs



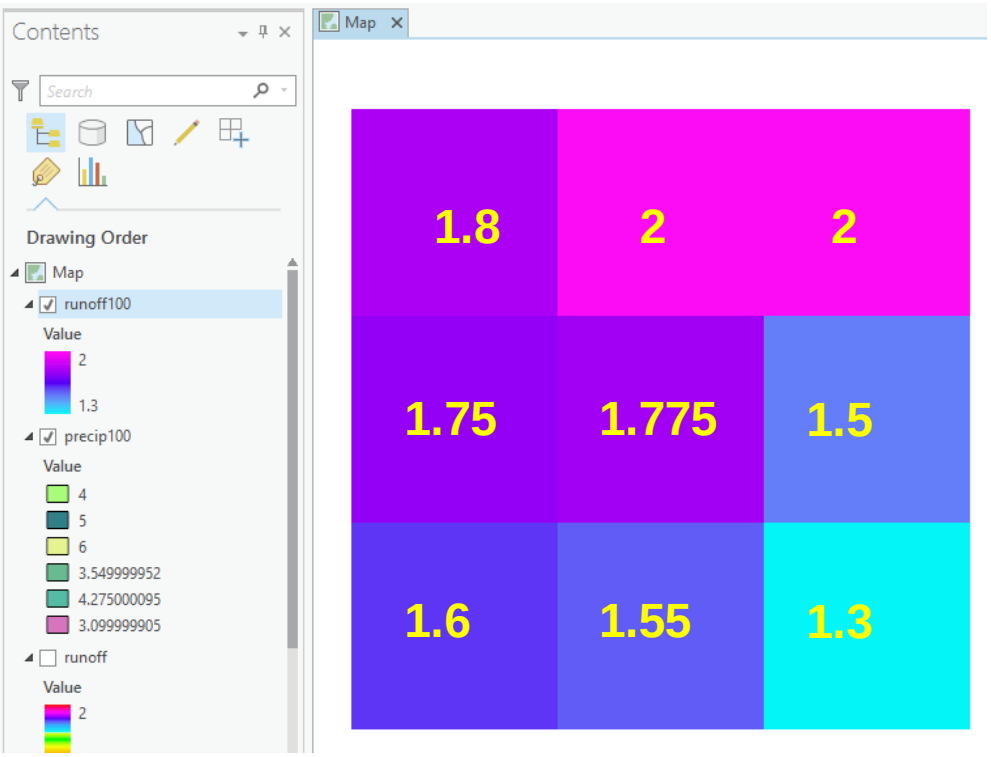
Resample to get consistent cell size



Spacing & Support

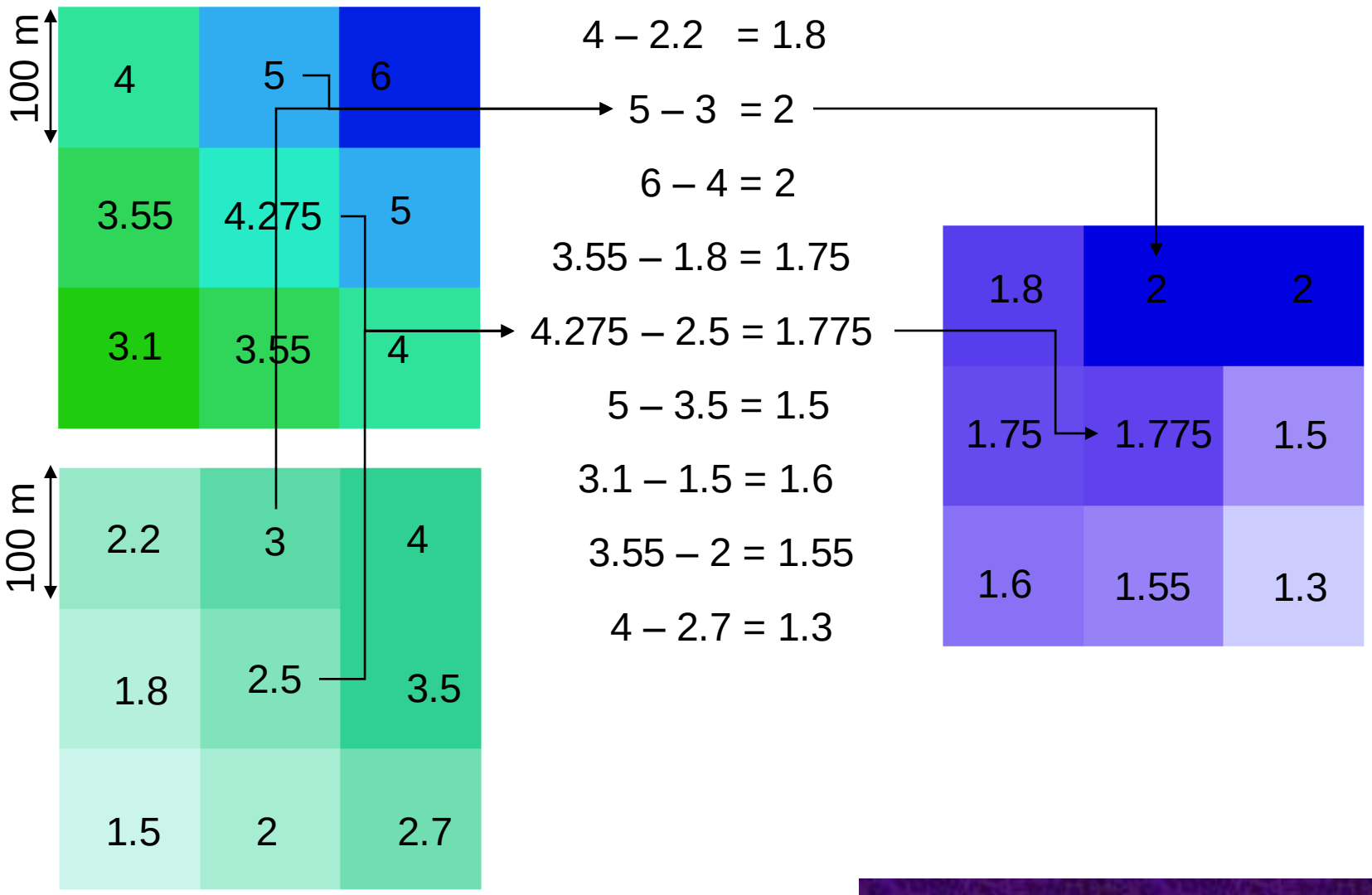
Calculation with consistent 100 m cell size grid

“precip100” - “infiltration”

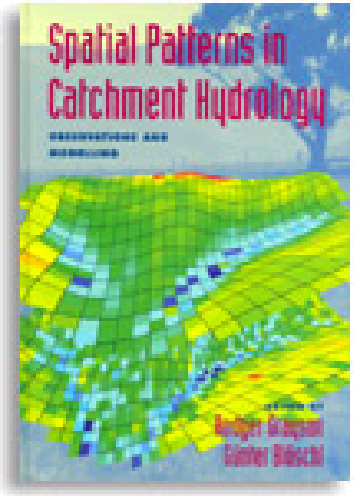


- Outputs are on 100 m grid as desired.
- How were values obtained ?

100 m cell size raster calculation



Further Reading



Grayson, R. and G. Blöschl, ed. (2000), Spatial Patterns in Catchment Hydrology: Observations and Modelling, Cambridge University Press, Cambridge, 432 p.

Chapter 2. Spatial Observations and Interpolation

Full text online at:

http://www.hydro.tuwien.ac.at/forschung/publikationen/engl-publikationen/spatial-patterns-in-catchment-hydrology.html