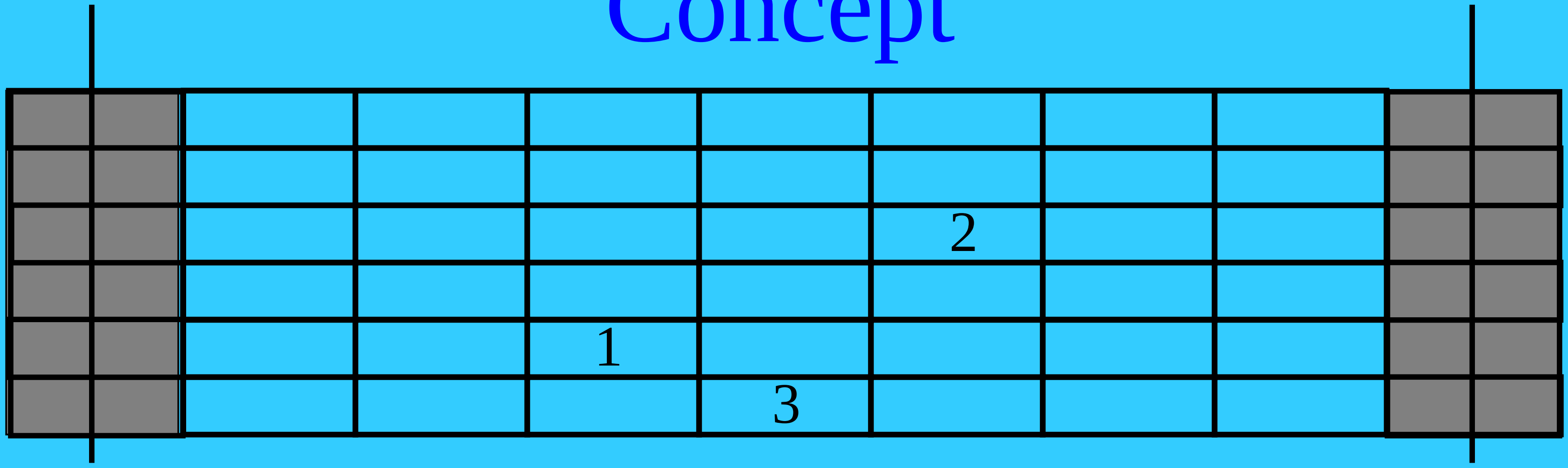
Indicator Simulation for Categorical Variables

- Sequential Simulation: the Concept
- Steps in Sequential Simulation
- SISIM Program

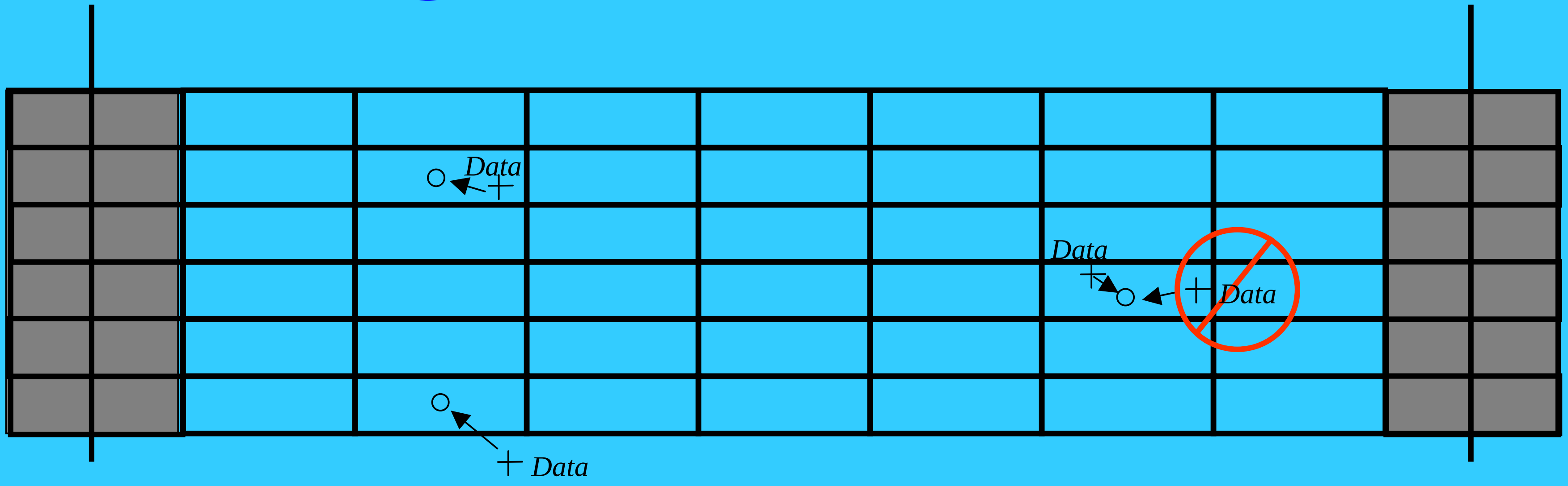
Sequential Simulation: the Concept



1. Assign data values to closest grid node
2. Establish a random path through all of the grid nodes
3. Visit each grid node:
   (a) find nearby data and previously simulated grid nodes
   (b) construct the conditional distribution by kriging (this is where the variogram comes in)
   (c) draw simulated value from conditional distribution
4. Check the results

STEP 1

Assign Data to Grid Nodes



Why?

- Explicitly honor data ⇒ data values will appear in final 3-D model
- Improves the CPU speed of the algorithm: searching for previously simulated nodes and original data is accomplished in one step

Considerations:

- Take the closest of multiple data within the same cell ⇒ could lose some information (see in figure ⊘)
- Not an option when simulating a cross-sectional or small-area model

STEP 2

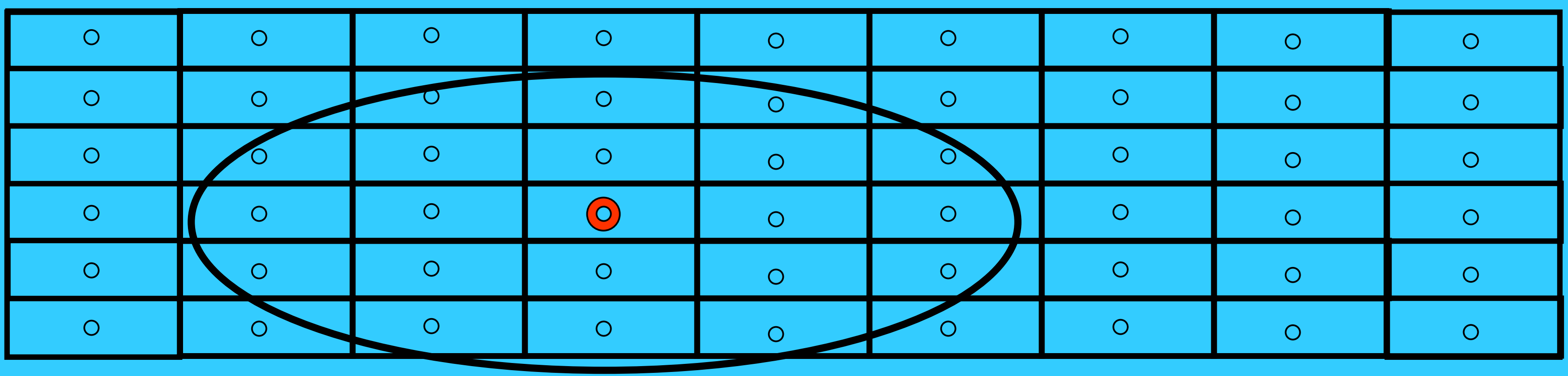
Establish a Random Path

		32	4	34	18	35	2	21		
		30	31	14	8	9	16	3		
		28	6	24	5	15	13	18		
		11	5	28	20	27	23	10		
		1	27	19	33	25	4	26		
		29	29	7	12	17	6	22		

- Visit each cell once and only once in random order
- Can do this in many ways:
  - draw a random number and multiply it by *N*
  - sort an array of random numbers while carrying an array of indices capitalize on the limited period length of linear congruential generators
- Skip over cells (actually grid nodes) that already have a value

STEP 3 - (a)

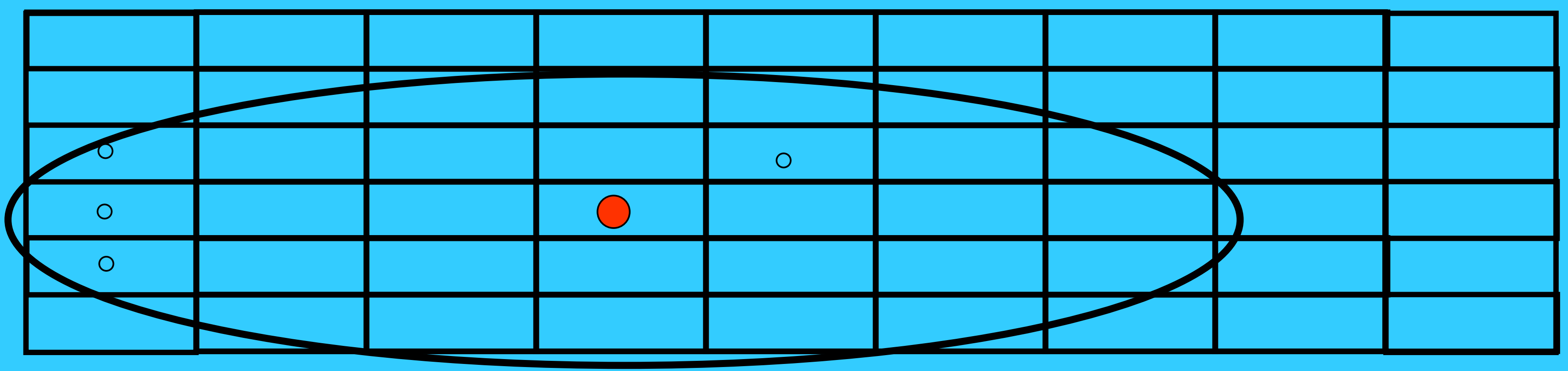
Find Nearby “Informed” Nodes



- “Informed” nodes refers to both data-nodes and nodes that have been informed earlier in the random path
- Typically use spiral search to identify the close nodes
- Limit the number of nodes actually considered:
  - octant search (?)
  - maximum per octant (say 4)
  - maximum number

STEP 3 - (b)

Construct Conditional Distribution



- Conditional distribution is constrained by:
  - global proportion of each lithology type
  - local data
  - “local” proportion from secondary data such as seismic
- Calculate by kriging the binary indicator transform for each rock type

Indicator Simulation (1)

- Define an indicator transform:

$$i(u_\alpha;k) = \begin{cases} 1, & if \quad lithosfacies \quad k \quad present \quad at \quad location \quad u_\alpha \\ 0, & if \quad not \end{cases}$$

- Average of an indicator is the global proportion:

$$proportion\ of\ k = \boldsymbol{E\{I(u_\alpha;k)\}}$$

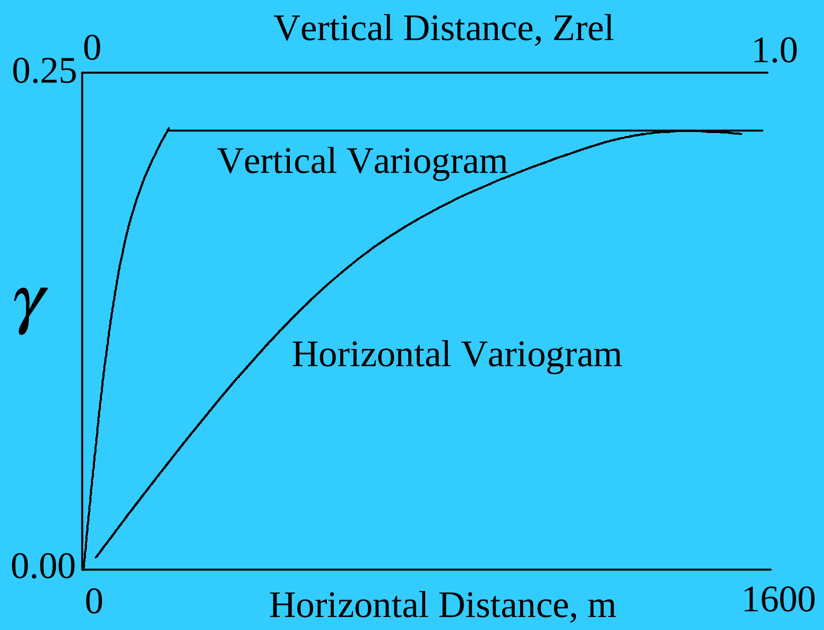
$$= \frac{\sum_{\alpha=1}^{n} w_\alpha \cdot i(u_{\alpha;k})}{n}$$

w_α values account for data clustering

Indicator Simulation (2)

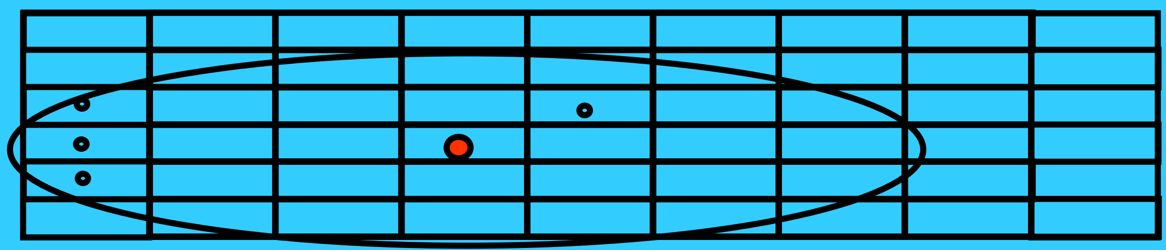
- The variogram of an indicator variable measures spatial correlation:

$$\gamma_I(h) = \frac{1}{2} E\{[I(u;k) - I(u+h;k)]^2\}$$



STEP 3 - (c)

Construct Conditional Distribution with Kriging



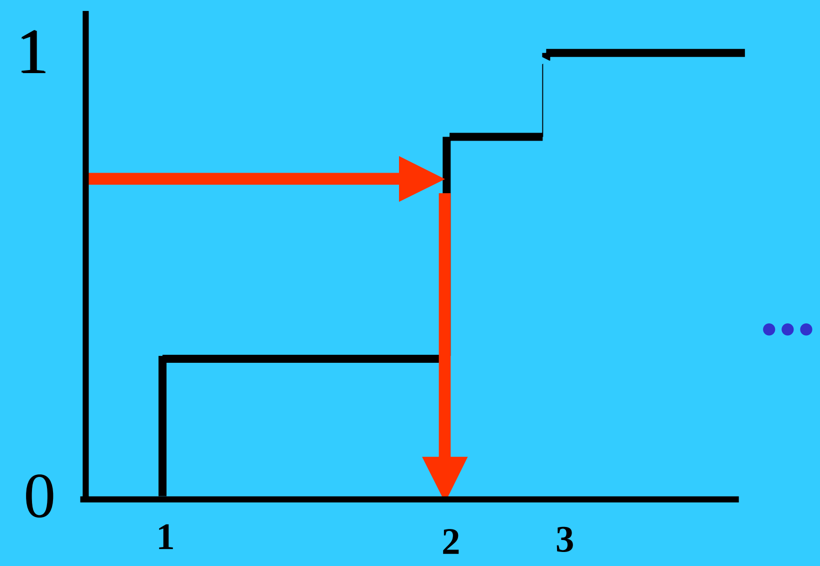
- Given n nearby data values $k(\mathbf{u}_i), i=1,\ldots,n$ how do we calculate the conditional distribution?
- Estimate conditional probabilities of each rock type $p_k^*(\mathbf{u})$, k=1,...,K by a linear combination of the nearby data:

$$p_k^*(u) = \sum_{\alpha=1}^{n} \lambda_\alpha(u) \cdot I(u_\alpha; k) + [1 - \sum_{\alpha=1}^{n} \lambda_\alpha(u)] \cdot m_k$$

- Determine weights $\lambda_\alpha(\mathbf{u})$, $\alpha=1,\ldots,n$ by the well known "normal system" or kriging.
- Kriging weights account for two things:
  - clustering of the data locations
  - closeness of the data to the location being considered

STEP 3 - (d)

Draw a Simulated Value



- probabilities $p^*_k(u)$, k=1,**...**,K are given by kriging
- Procedure:
  - draw a random number $\in [0,1]$
  - find which class *k* is specified by the random number
  - assign *k* to node
- Since the conditional probabilities were estimated by kriging with a given variogram $\gamma_k(\mathbf{h})$, k=1,**...**,K, the simulated values, taken all together, will reproduce those variograms ,$\gamma_k(\mathbf{h})$, k=1,**...**,K

Detailed Steps in SISIM

1. Establish grid network and coordinate system ($\boldsymbol{Z}_{rel}$-space)
2. Assign data to the nearest grid node (take the closest of multiple data assigned to same node)
3. Determine a random path through all of the grid nodes
   (a) find nearby data and previously simulated grid nodes
   (b) construct the conditional probabilities by kriging
   (c) draw simulated value from conditional distribution
4. Check results
   (a) honor data?
   (b) honor global proportions?
   (c) honor variogram?
   (d) look reasonable

Parameters for SISIM

****

```
START OF PARAMETERS:
1                                   \ 1=continuous(cdf), 0=categorical(p
5                                   \ number thresholds/categories
0.5   1.0   2.5   5.0   10.0        \ thresholds / categories
0.12  0.29  0.50  0.74  0.88        \ global cdf / pdf
../data/cluster.dat                 \ file with data
1   2   0   3                       \ columns for X,Y,Z, and variable
direct.ik                           \ file with soft indicator input
1   2   0   3 4 5 6 7               \ columns for X,Y,Z, and indicators
0                                   \ Markov-Bayes simulation (0=no,1=yes)
0.61  0.54  0.56  0.53  0.29        \ calibration B(z) values
-1.0e21    1.0e21                   \ trimming limits
0.0   30.0                          \ minimum and maximum data value
1      0.0                          \ lower tail option and parameter
1      1.0                          \ middle option and parameter
1     30.0                          \ upper tail option and parameter
cluster.dat                         \ file with tabulated values
3   0                               \ columns for variable, weight
0                                   \ debugging level: 0,1,2,3
sisim.dbg                           \ file for debugging output
sisim.out                           \ file for simulation output
```

```
1                                  \ number of realizations
50   0.5    1.0                    \ nx,xmn,xsiz
50   0.5    1.0                    \ ny,ymn,ysiz
1    1.0   10.0                    \ nz,zmn,zsiz
69069                              \ random number seed
12                                 \ maximum original data  for each kriging
12                                 \ maximum previous nodes for each kriging
1                                  \ maximum soft indicator nodes for kriging
0                                  \ assign data to nodes? (0=no,1=yes)
0     3                            \ multiple grid search? (0=no,1=yes),num
0                                  \ maximum per octant    (0=not used)
20.0  20.0  20.0                          \ maximum search radii
 0.0   0.0   0.0                   \ angles for search ellipsoid
0    2.5                           \ 0=full IK, 1=median approx. (cutoff)
0                                  \ 0=SK, 1=OK
1    0.15                          \ One   nst, nugget effect
1    0.85 0.0   0.0   0.0                 \ it,cc,ang1,ang2,ang3
         10.0  10.0  10.0                 \ a_hmax, a_hmin, a_vert
1    0.10                          \ Two   nst, nugget effect
1    0.90 0.0   0.0   0.0                 \ it,cc,ang1,ang2,ang3
         10.0  10.0  10.0                 \ a_hmax, a_hmin, a_vert
1    0.10                          \ Three nst, nugget effect
1    0.90 0.0   0.0   0.0                 \ it,cc,ang1,ang2,ang3
         10.0  10.0  10.0                 \ a_hmax, a_hmin, a_vert
```