



**exprodat**

integrating information

# ESDA, Data Modelling & Evaluation of Uncertainty

ArcGIS Geostatistical Analyst™ 10 (beta) and G&G software packages

Paola Peroni - Exprodat Consulting Ltd

July 2010

# Presentation plan

- Project background & scope
- Proposed workflow & tools
- Workflow in Geostatistical Analyst™ 10 (GA 10)
- G&G applications
- Results
- Conclusions and recommendations

# Background

- Derive 2D surfaces from sparsely, discrete data
- Variety of tools and functionality
  - Most GIS applications
  - G&G packages
- E&P workflows often miss data analysis and evaluation of uncertainty
- Limited knowledge of quality of the results and reliability of the models

# Project scope

- General workflow and tools for
  - Exploratory spatial data analysis (ESDA)
  - 2D interpolation
  - Evaluation of uncertainty
- Explore functionality and tools available in
  - ESRI ArcGIS GA 10
  - G&G applications widely used in the E&P industry
    - Application 1
    - Application 2

# Workflow & tools



**Univariate**

- Histogram
- Summary statistics
- Normal QQ plot
- Cumulative frequency plot

**Multivariate**

- QQ plot
- Scatterplot, conditional expectation curve
- Covariation of multiple variables

**Anomalies & multiple populations**

- Moving windows statistics
- Data transformation
- Voronoi diagram

**Spatial continuity & directional influences**

- Trend analysis
- Semivariogram/covariance surface
- Anisotropy

**Global estimators**

- Declustering tools
- Global polynomials/Trend surface

**Local estimators**

<b>Deterministic</b>	<b>Geostatistics</b>
• Search strategy tools	• Search strategy & semivariogram

<b>Without barriers</b>	<b>With barriers</b>	<b>Without barriers</b>	<b>With barriers</b>
• Smooth	• Automatic handle	• Kriging	• Automatic handle
• Exact		• Cokriging	

**Magnitude**

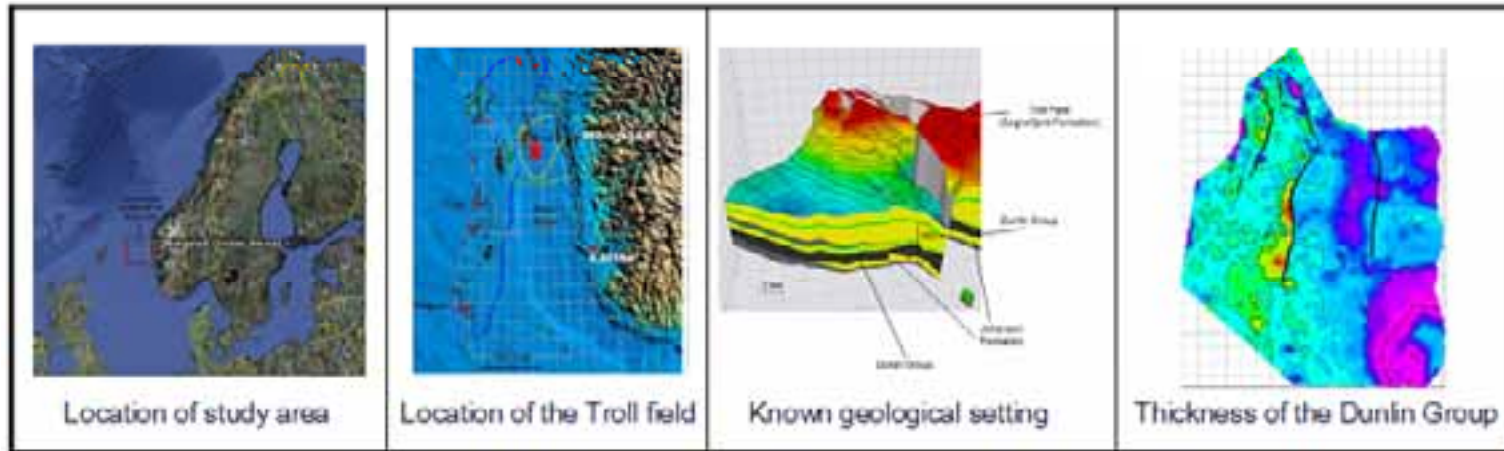
- Validation
- Cross-Validation

**Spatial distribution**

- Spatial distribution of residuals
- Simulation

# Test dataset

- Dataset of geological variables derived from published data (\*)



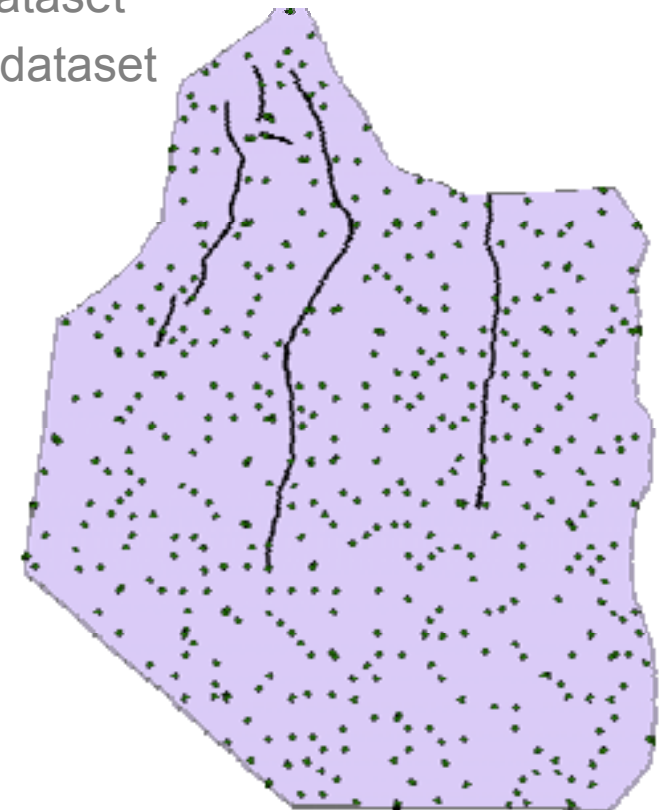
- Advantages
  - Focus on E&P variables
  - Horizon's geological model validated and published by a trusted source
  - Geological setting includes barriers (faults)

(\*) From SINTEF CO<sub>2</sub> storage project website: <http://www.sintef.no/Projectweb/MatMorA/Downloads/Johansen/>

# Derived continuous variables

- Two continuous variables derived from the thickness of the Dunlin Group
  - V1: variance of thickness within the sampling cell (\*)
  - V2: function of variance and mean within the sampling cell
- Stratified random sampling strategy to have:
  - Uneven spatial distribution of points in input dataset
  - All value classes equally represented in input dataset
  - A proxy for field-scale & regional studies

ID	Stage	X	Y	Variable1	Variable2
297	Point 250	981222.499959	9895482.12382	1.949999	0.919999
124	Point 250	922226.232088	8729888.81878	1.849272	0.819981
396	Point 250	984017.848177	9889781.74824	3.144818	0.839488
440	Point 250	959732.790918	8729898.28478	1.918984	0.83441
68	Point 250	989882.572182	8879222.87188	2.973488	0.827222
86	Point 250	973946.229482	8897138.47981	1.889188	0.821881
464	Point 250	979887.789836	9889228.87474	1.473548	0.998286
47	Point 250	988888.873828	8887828.18881	0.849182	0.108438
384	Point 250	988221.888278	8879282.72887	1.288148	0.102211
128	Point 250	888884.888488	8889222.88124	4.888424	0.128481
232	Point 250	871817.342481	8887882.82284	0.918887	0.142828
388	Point 250	988881.818881	8718888.78884	0.387881	0.147818
138	Point 250	984222.881888	8888732.81882	7.127888	0.107847
481	Point 250	988881.478884	8874848.81128	0.188878	0.288843
121	Point 250	982488.138274	8722274.81488	4.882822	0.288274
488	Point 250	981821.138827	8888818.88878	11.881848	0.334771
148	Point 250	987828.881878	8888481.31888	4.821382	0.43888
482	Point 250	882228.188747	8871871.28884	7.88848	0.438821
38	Point 250	948222.888818	8728272.47884	17.348872	0.478884
228	Point 250	888488.881187	8888814.88874	16.288482	0.51888
134	Point 250	881878.88818	8888782.88882	16.287827	0.548182
84	Point 250	982788.888878	8734882.82778	18.881874	0.828821
448	Point 250	982221.778818	8877112.48884	14.822888	0.887732
488	Point 250	978881.877888	8881782.48181	7.328818	0.887517
388	Point 250	984788.27248	8732228.88848	14.478848	0.874182
52	Point 250	981188.877881	8888888.87472	11.281182	0.888842
488	Point 250	888888.88821	8888828.88184	6.148882	0.72888
38	Point 250	888882.718481	8878882.1881	17.428811	0.754728
227	Point 250	888488.24814	8888842.88888	7.888282	0.78882
288	Point 250	888881.241881	8722288.81181	16.228222	0.778812
42	Point 250	972488.877911	8884272.48182	4.878882	0.842182
121	Point 250	948848.148821	8718841.44837	18.983884	0.888882



(\*) Sampling cell: 2500x2500m

# The workflow in GA 10



**Univariate**

- Histogram
- Summary statistics
- Normal QQ plot
- Cumulative frequency plot

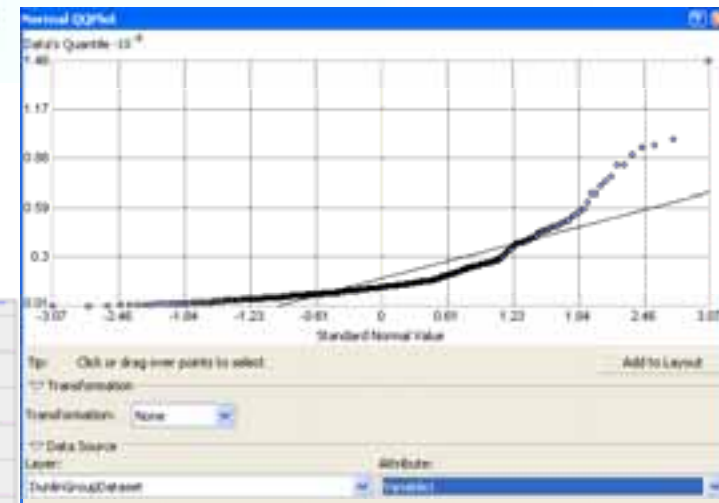
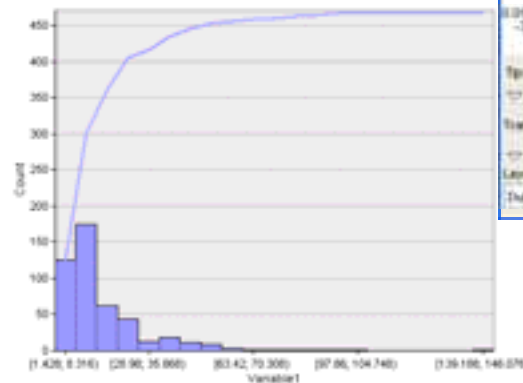
**Multivariate**

- QQ plot
- Scatterplot, conditional expectation curve
- Covariation of multiple variables



Histogram & summary statistics

Cumulative frequency plot (\*)

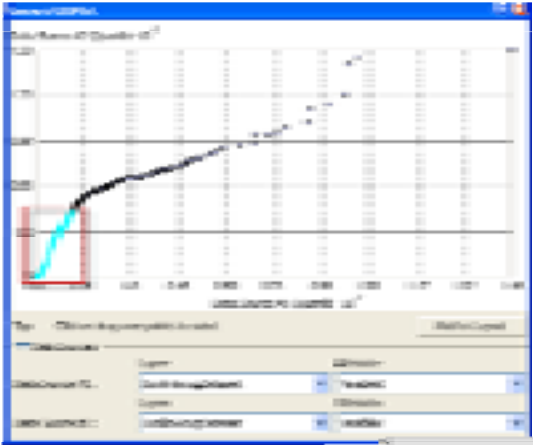
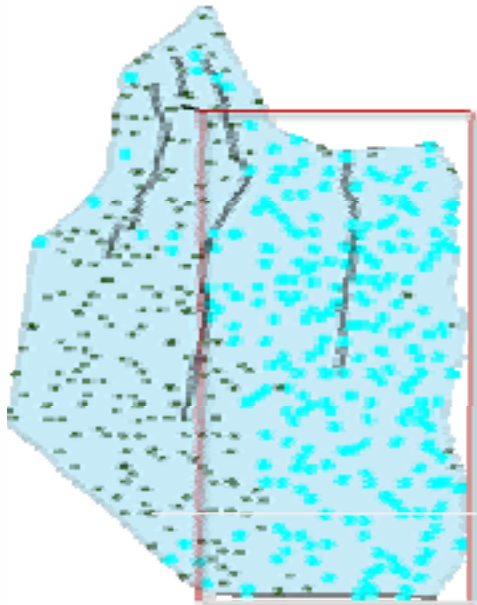


Normal QQ plot

(\*) Graph tool in ArcMap

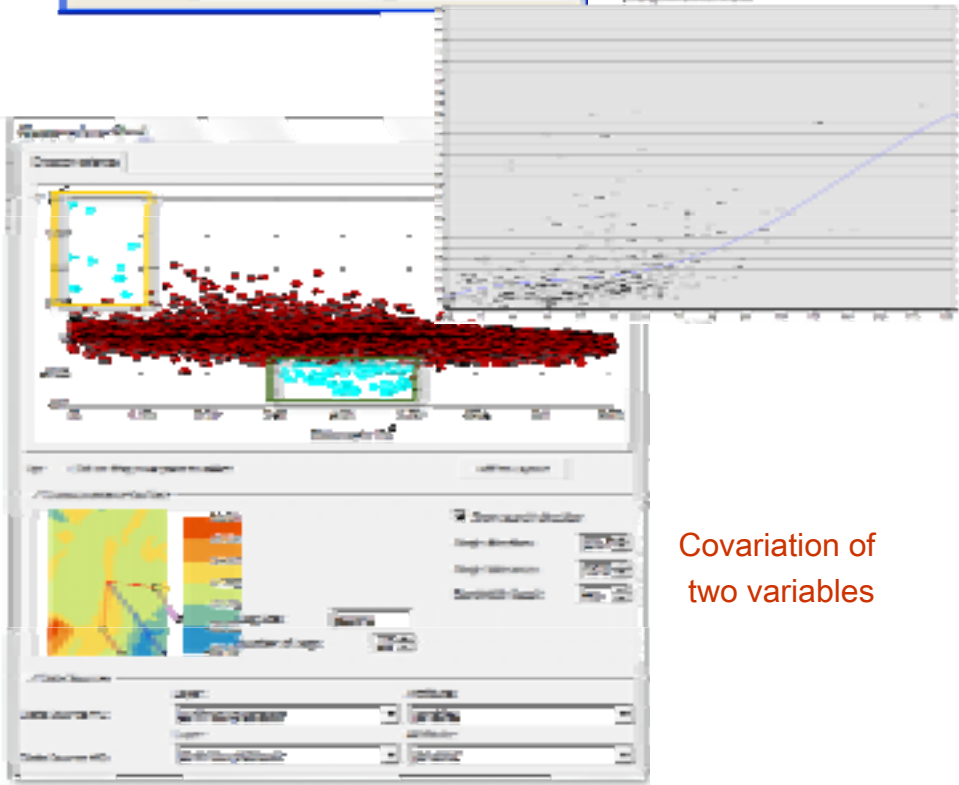
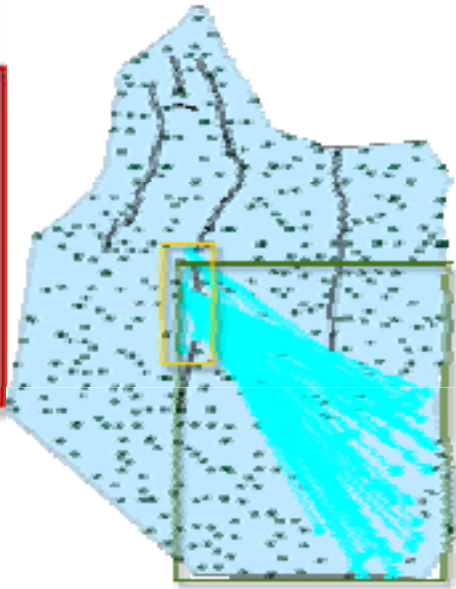


# The workflow in GA 10



QQ plot

- Univariate**
  - Histogram
  - Summary statistics
  - Normal QQ plot
  - Cumulative frequency plot
- Multivariate**
  - QQ plot
  - Scatterplot, conditional expectation curve
  - Covariation of multiple variables



Scatterplot

Covariation of two variables

# The workflow in GA 10



What we've learnt:

## Univariate

- Histogram
- Summary statistics
- Normal QQ plot
- Cumulative frequency plot

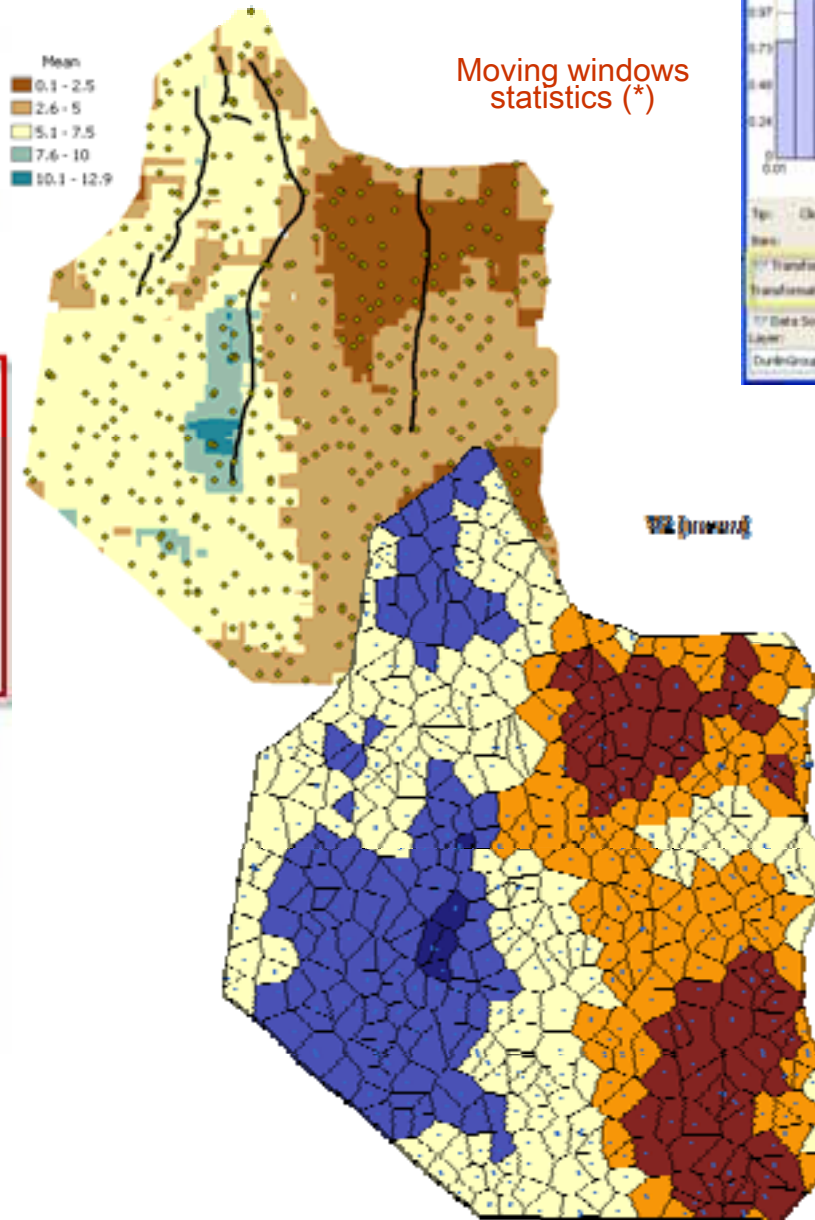
## Multivariate

- QQ plot
- Scatterplot, conditional expectation curve
- Covariation of multiple variables

- Univariate description:
  - Highly skewed distribution for V1, far from normal (Gaussian) and close to log-normal
  - Distribution of V2 much closer to normal
- Multivariate description:
  - Non-linearity in joint variation of V1 and V2
  - Inter-variable spatial correlation is maximum in  $\sim$  NS direction

# The workflow in GA 10

Data transformation



Moving windows statistics (\*)



Voronoi diagram

**Anomalies & multiple populations**

- Moving windows statistics
- Data transformation
- Voronoi diagram

**Spatial continuity & directional influences**

- Trend analysis
- Semivariogram/covariance surface
- Anisotropy

(\*) Done in ArcGIS Spatial Analyst™

# The workflow in GA 10

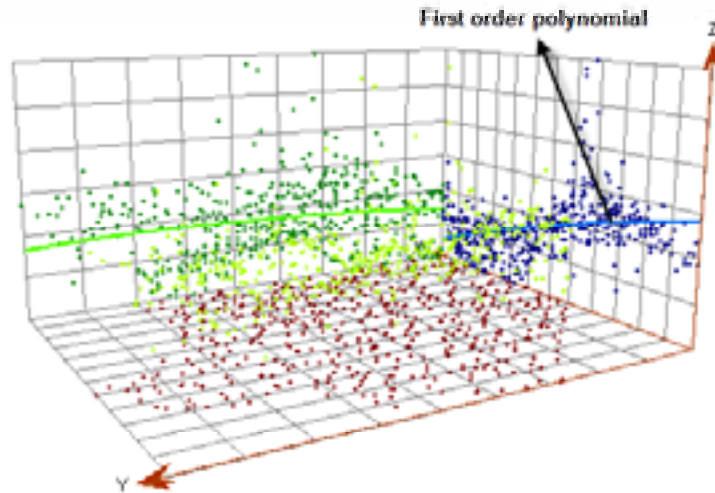


## Anomalies & multiple populations

- Moving windows statistics
- Data transformation
- Voronoi diagram

## Spatial continuity & directional influences

- Trend analysis
- Semivariogram/covariance surface
- Anisotropy



Rotation Angles

Location: 270°

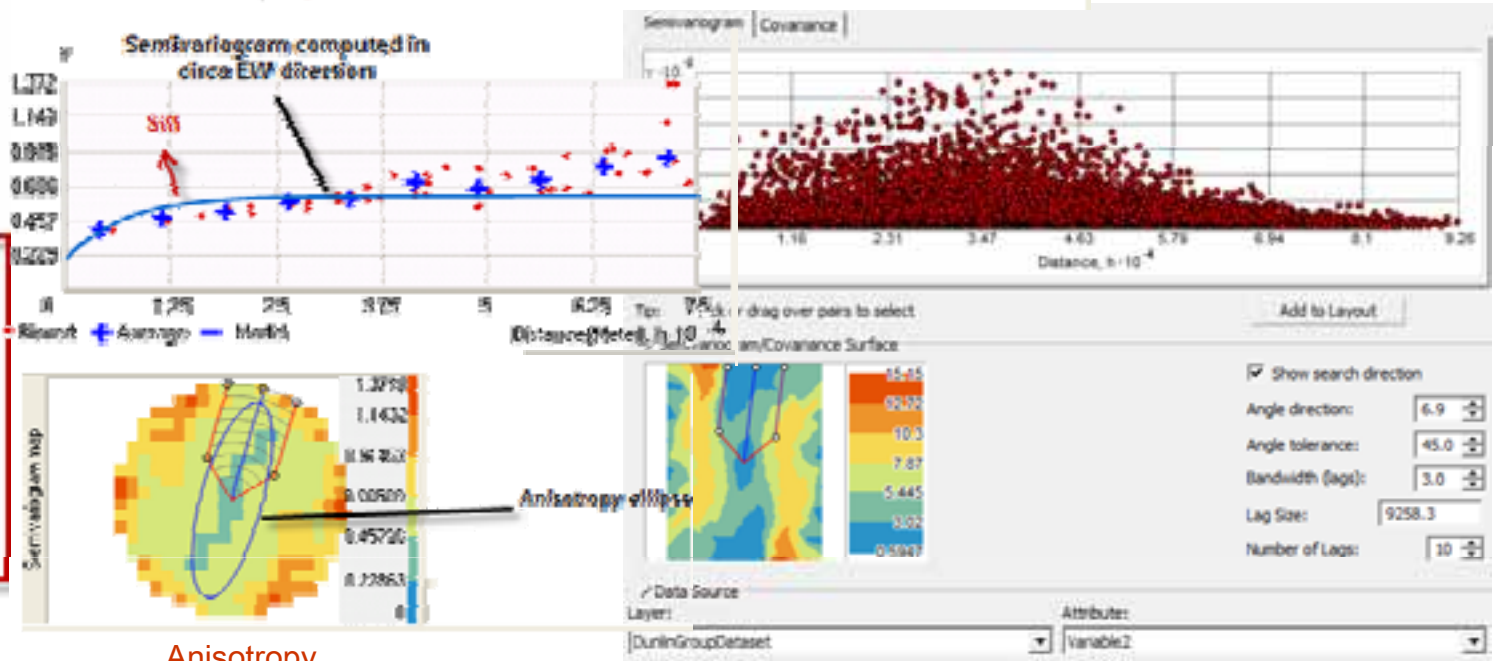
Horizontal: 120°

Vertical: -11.5°

Trend direction circa EW

Trend analysis


Semivariogram surface



Anisotropy




# The workflow in GA 10



## Data Analysis

What we've learnt:

- **Anomalies & multiple populations**
  - Stationarity assumption not met (V1). Source to be investigated further
  - Possibly two populations for V2
- **Spatial continuity & directional influences**
  - First order trend ~ EW affecting V2
  - Autocorrelation is directional-dependent for V1 (anisotropy)



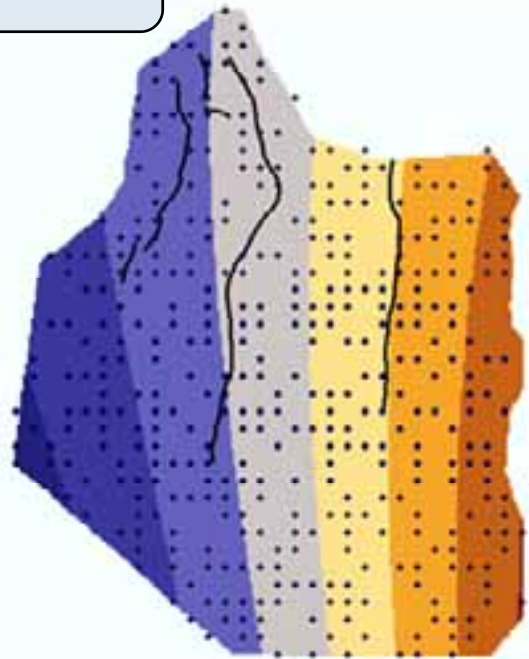
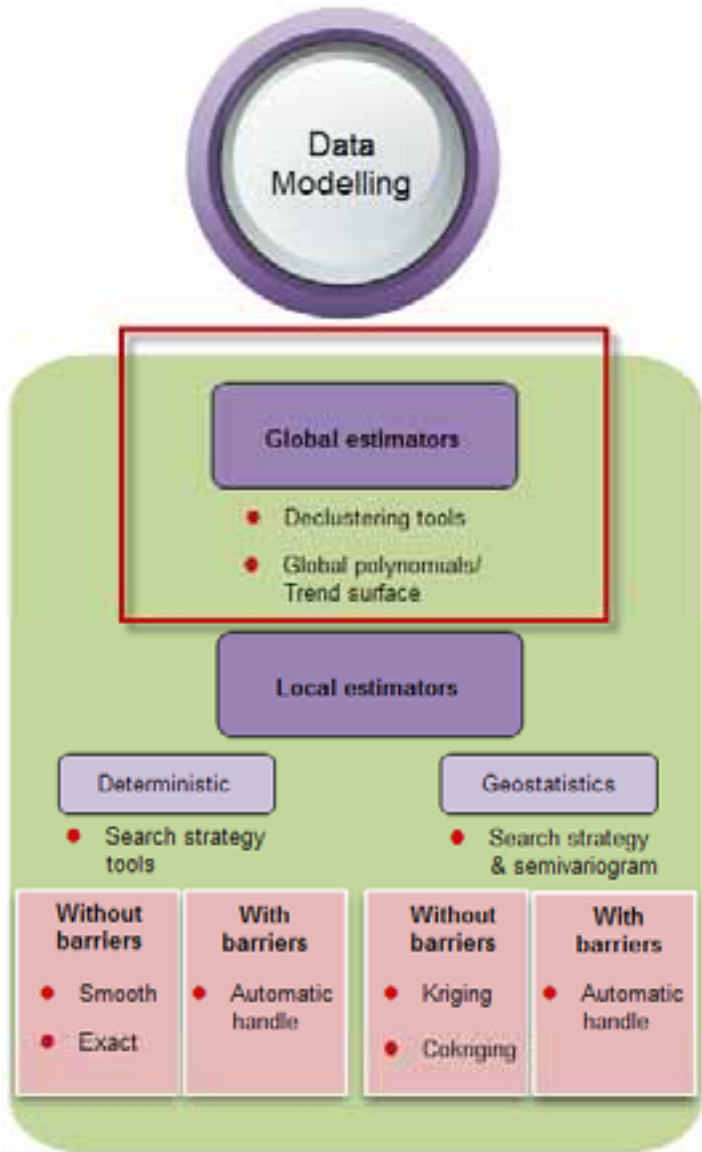
### Anomalies & multiple populations

- Moving windows statistics
- Data transformation
- Voronoi diagram

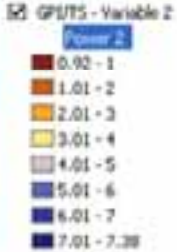
### Spatial continuity & directional influences

- Trend analysis
- Semivariogram/covariance surface
- Anisotropy

# The workflow in GA 10

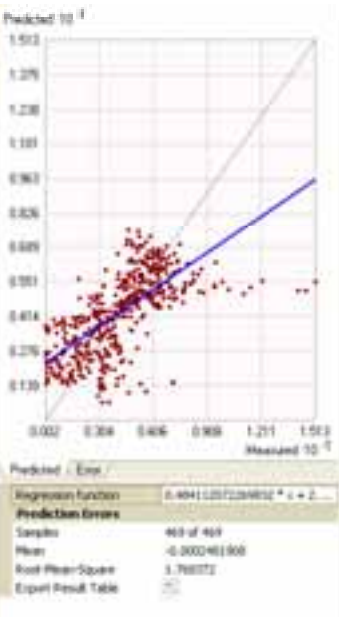


Declustered dataset (\*) & Trend surface



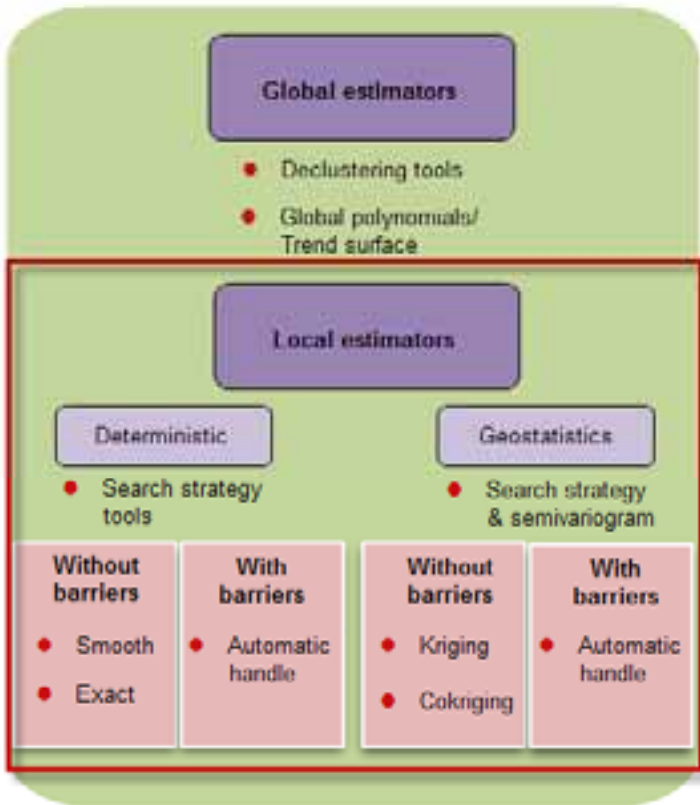
ID	Included	Observed	Predicted	Error
0	Yes	4.891870	7.1771	-2.285186660003008
1	Yes	4.077340	4.8384	-0.7611333122017688
2	Yes	5.607026	4.0513	1.555721150647100
3	Yes	5.930771	4.8112	1.11951689179470184
4	Yes	5.446166	4.2357	1.2108041094484395
5	Yes	5.143227	4.7004	0.4428230000000000
6	Yes	5.140129	4.2107	0.92940823779281995
7	Yes	4.450880	5.0999	-0.6491207700436664
8	Yes	4.124316	4.2296	-0.1052949667246360
9	Yes	5.001062	5.8425	-0.8414200968179001
10	Yes	5.011523	5.5636	-0.5520732151100000
11	Yes	4.576162	4.0960	0.480152008417140
12	Yes	4.550018	5.3771	-0.82707380702037
13	Yes	5.075790	5.8381	-0.762324710296021
14	Yes	5.261266	5.7281	-0.4668247600566307
15	Yes	4.947201	5.4993	-0.5521343944077135
16	Yes	5.88640	5.3007	0.585703809727008
17	Yes	4.273617	4.0773	0.1962630747000812
18	Yes	4.750041	4.0128	0.737810901499660
19	Yes	4.903061	5.2465	-0.3434930000000000
20	Yes	4.442226	5.1224	-0.6801757622888
21	Yes	7.184959	5.0517	2.133237657396024
22	Yes	4.60252	5.2664	-0.663873196207037
23	Yes	5.330946	5.1608	0.1703321110000000
24	Yes	5.078020	4.3878	0.6901930000000000
25	Yes	1.921446	4.0908	-2.1693731315601106
26	Yes	5.972226	4.9414	1.0308200000000000
27	Yes	5.172216	4.8811	0.2911240000000000
28	Yes	2.862140	4.9695	-2.107362401943047
29	Yes	5.231531	4.7054	0.52646215415021
30	Yes	4.054177	4.3004	-0.2462232000000000
31	Yes	2.740004	4.3896	-1.649596215415021
32	Yes	2.852049	4.3140	-1.4619772000000000
33	Yes	5.541089	4.2214	1.3191300000000000
34	Yes	2.924020	4.2324	-1.3084040000000000
35	Yes	2.717096	4.1577	-1.4406340000000000
36	Yes	2.127441	4.1402	-2.0127660000000000
37	Yes	3.064521	4.0094	-0.9448750000000000
38	Yes	2.410021	3.8781	-1.4680790000000000

Cross-validation statistics & plots

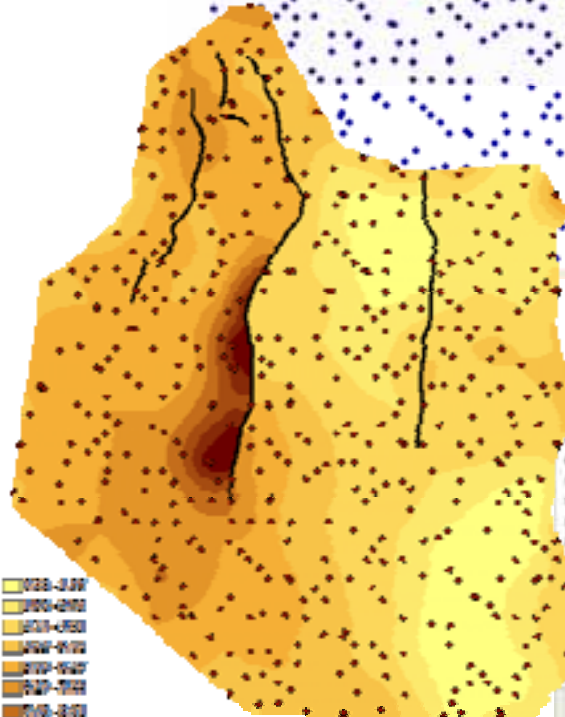
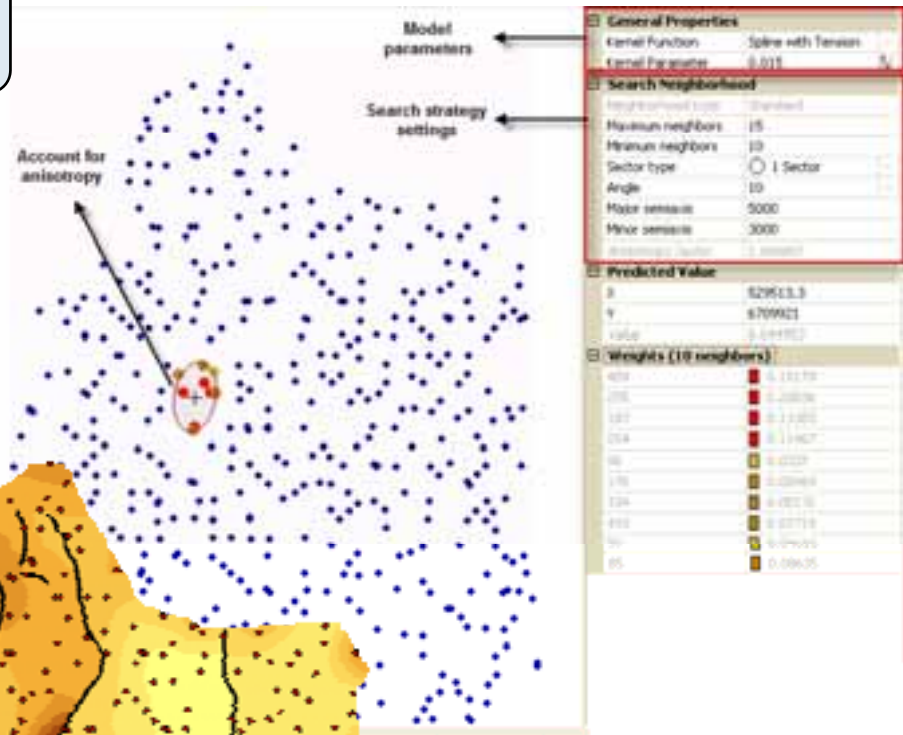


(\*) Done in ArcGIS Spatial Analyst™

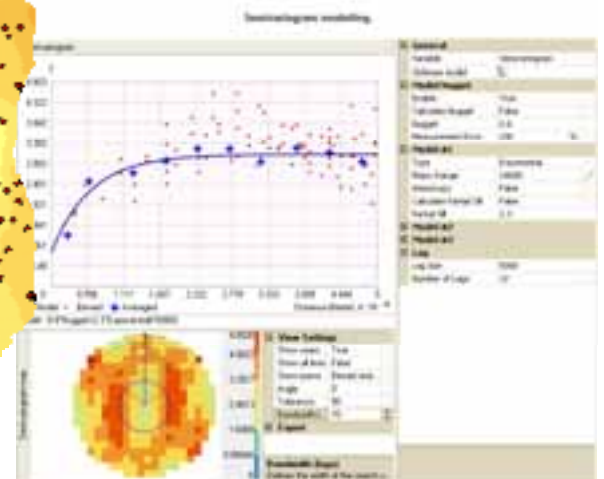
# The workflow in GA 10



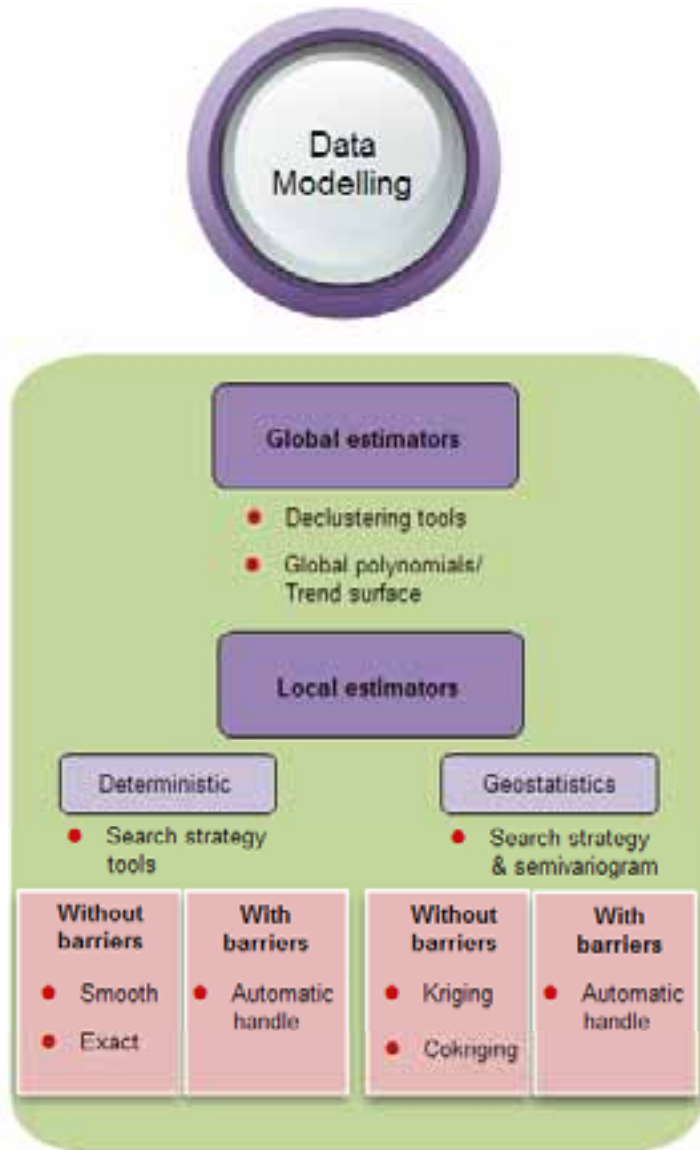
Search strategy



Interpolation with barriers & search strategy/semivariogram



# The workflow in GA 10



What we've done:

- Generated models for V1 & V2
  - Global estimators
  - Local estimators (deterministic & geostatistical)
- Two modelling approaches
  - Quick & Dirty
  - Customisation of modelling parameters



# The workflow in GA 10

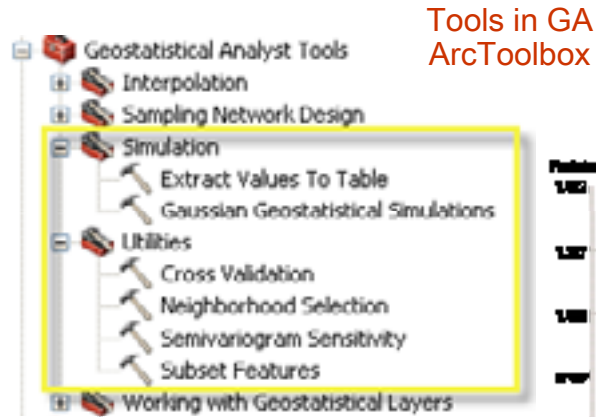
## Evaluation of Uncertainty

### Magnitude

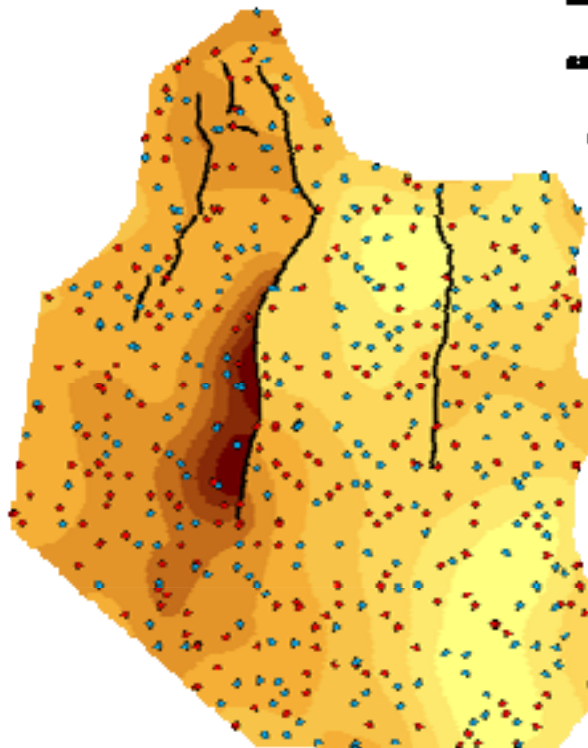
- Validation
- Cross-Validation

### Spatial distribution

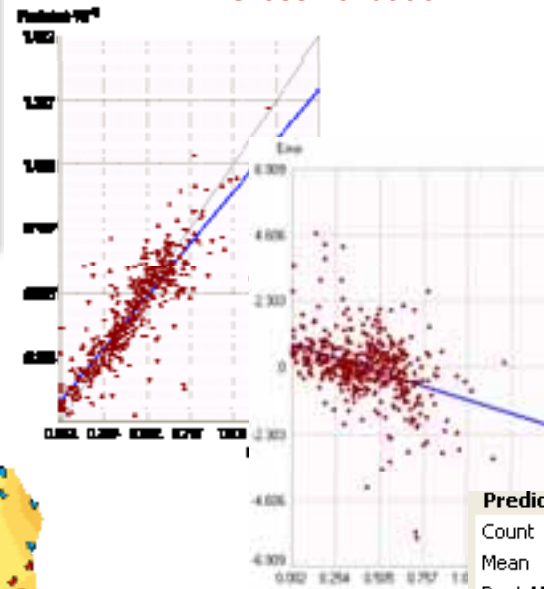
- Spatial distribution of residuals
- Simulation



Tools in GA  
ArcToolbox



Cross-validation



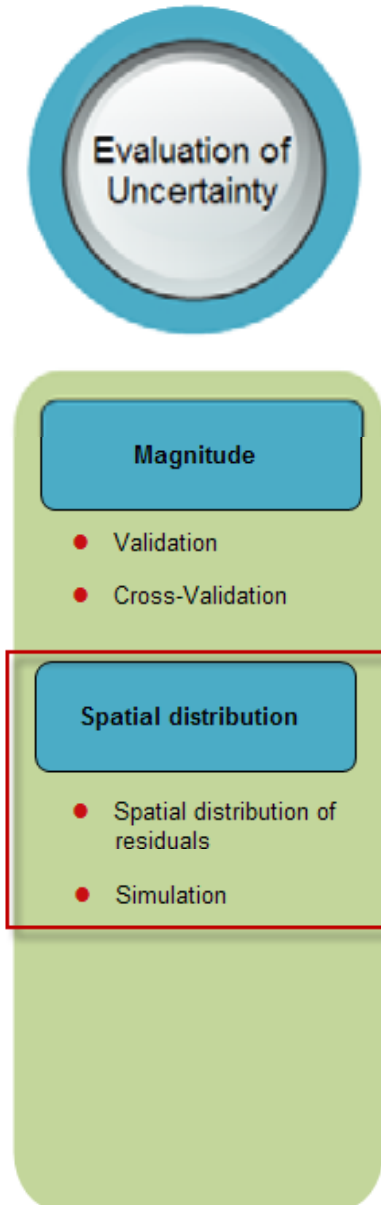
Scatterplot  
&  
error statistics

Prediction Errors	
Count	465
Mean	-0.002663062
Root-Mean-Square	1.216424
Mean Standardized	0.01094384
Root-Mean-Square Standar...	1.005046
Average Standard Error	1.164198

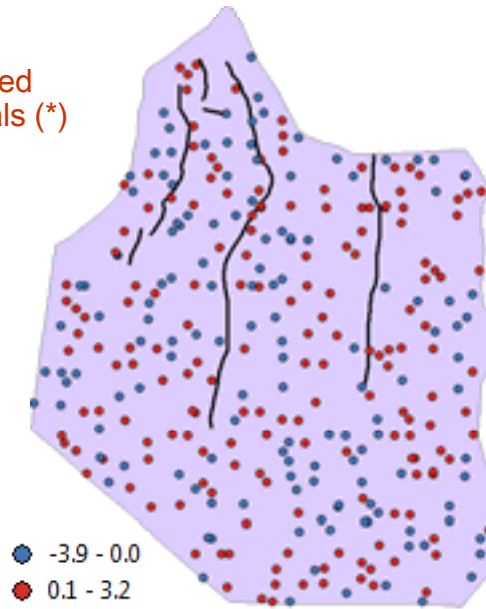
Validation

- Training dataset
- Test dataset

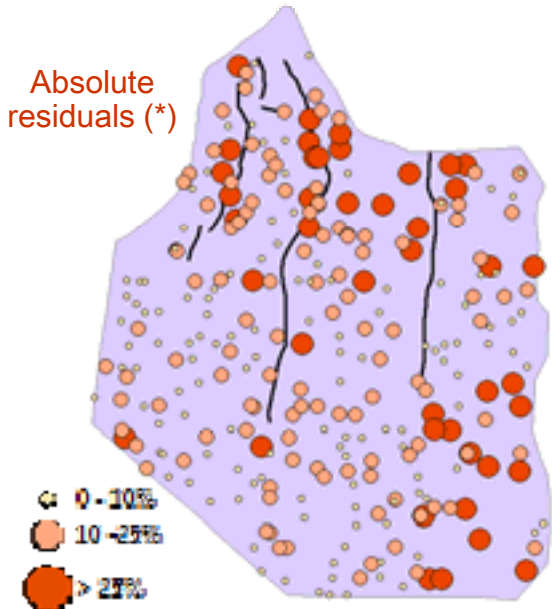
# The workflow in GA 10



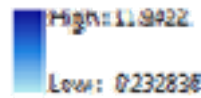
Signed residuals (\*)



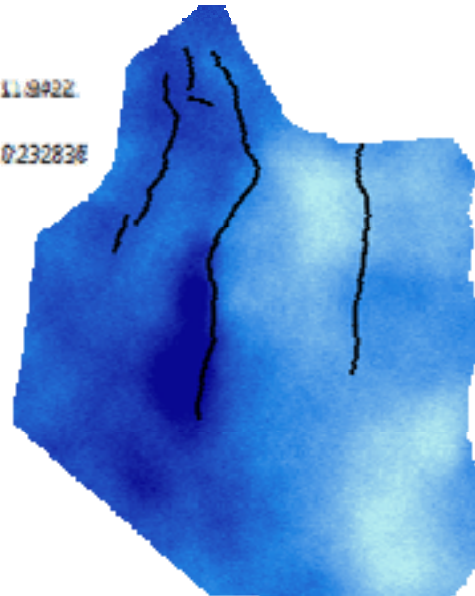
Absolute residuals (\*)



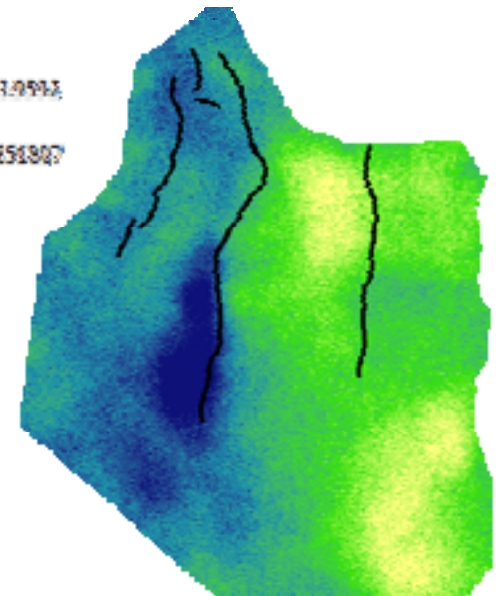
Mean



Gaussian simulation

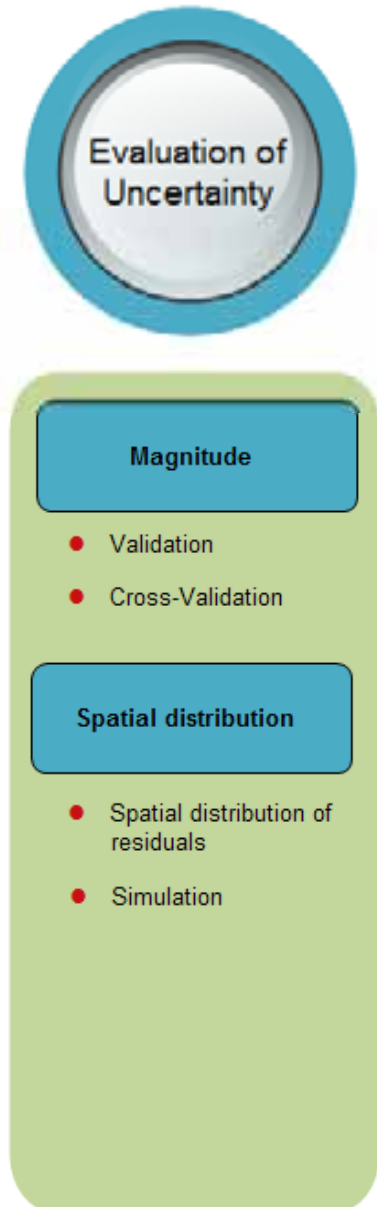


P 90



(\*) Generated in ArcMap™ 10

# The workflow in GA 10



What we've learnt:

- Identified models for which residuals are less biased
- For each model, identified areas with highest values of uncertainty
- Include stochastic concepts to deal with the uncertainty of the modelling process through simulation

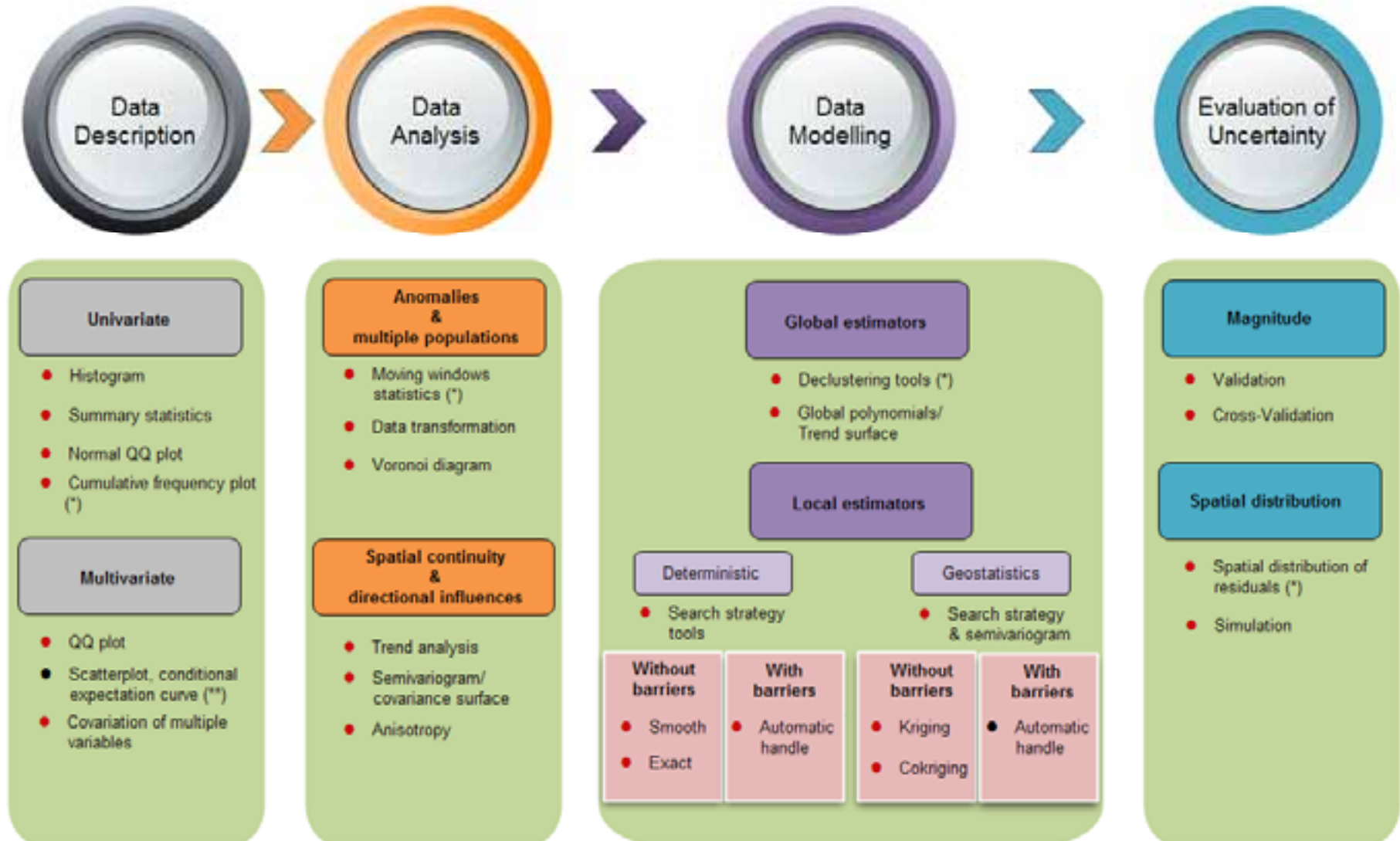
## Summary: GA 10

- Rich set of tools throughout the whole workflow
- Capability of building **advanced spatial models** to account for complex spatial behaviours
- Quality of the output
  - Interaction of tools
  - Excellent graphic
- Good “Help” and description of the scientific approach to modelling
- However:
  - Limited capability of including faults

# Summary: GA 10

Key:

- Available in GA 10
- Not available in GA 10
- (\*) Available in ArcGIS (ArcMap)
- (\*\*) Scatterplot available as Graph tool



Without barriers	With barriers	Without barriers	With barriers
● Smooth	● Automatic handle	● Kriging	● Automatic handle
● Exact		● Cokriging	

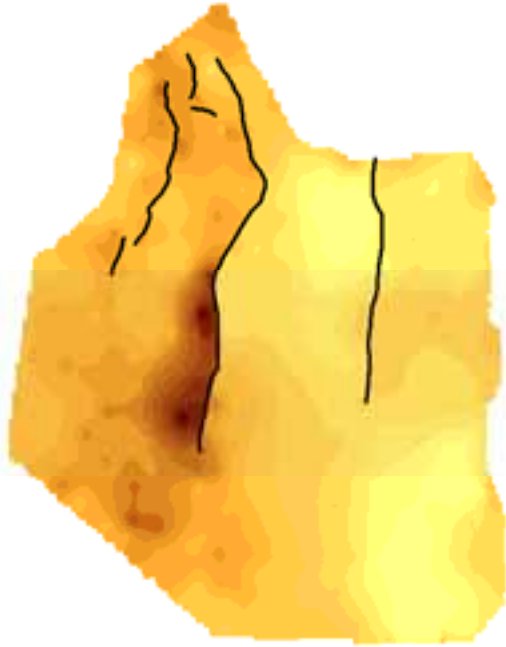
## G & G applications: general

- Follow the proposed workflow by using tools available in tested application
- Derive the same complex spatial models as done in ArcGIS GA by using variety of available algorithms
  - ‘Quick & Dirty’
    - Uses mostly default settings
    - Little or no customisation
  - Common custom parameters
    - Customising model parameters
- Comparison of model uncertainty with results obtained in GA 10

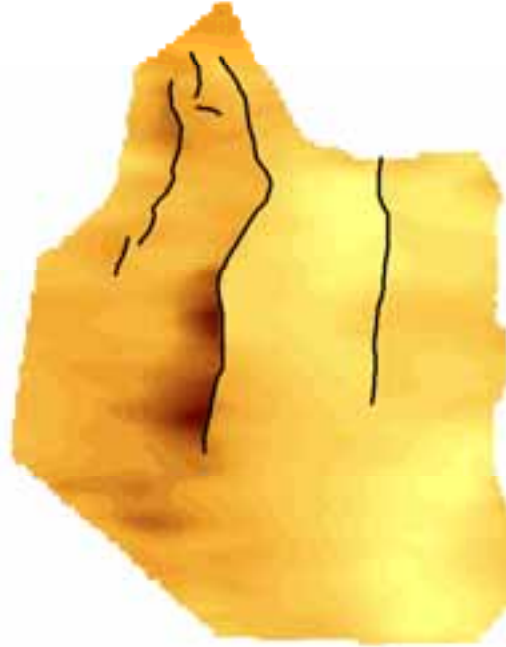
## Application 1 & 2: overview

- Good number of algorithms available (deterministic and geostatistics)
- Can automatically handle faults with any local interpolator
- Very few tools for data exploration and analysis (essentially Histogram)
- Limitations in the number of customisable modelling parameters
- Kriging error maps are the only evaluation of uncertainty tool

# Models: Application 1 & 2



Local polynomial  
with barriers



Ordinary Kriging with barriers  
and trend in ~ EW

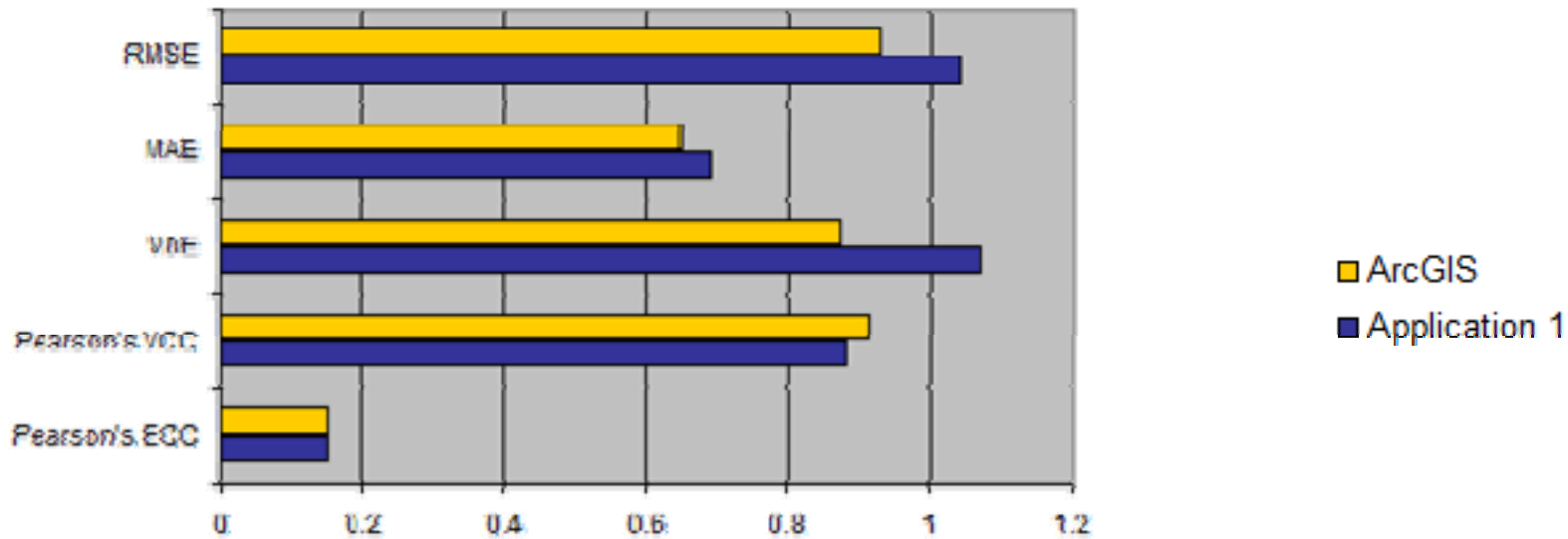


Spline  
with barriers

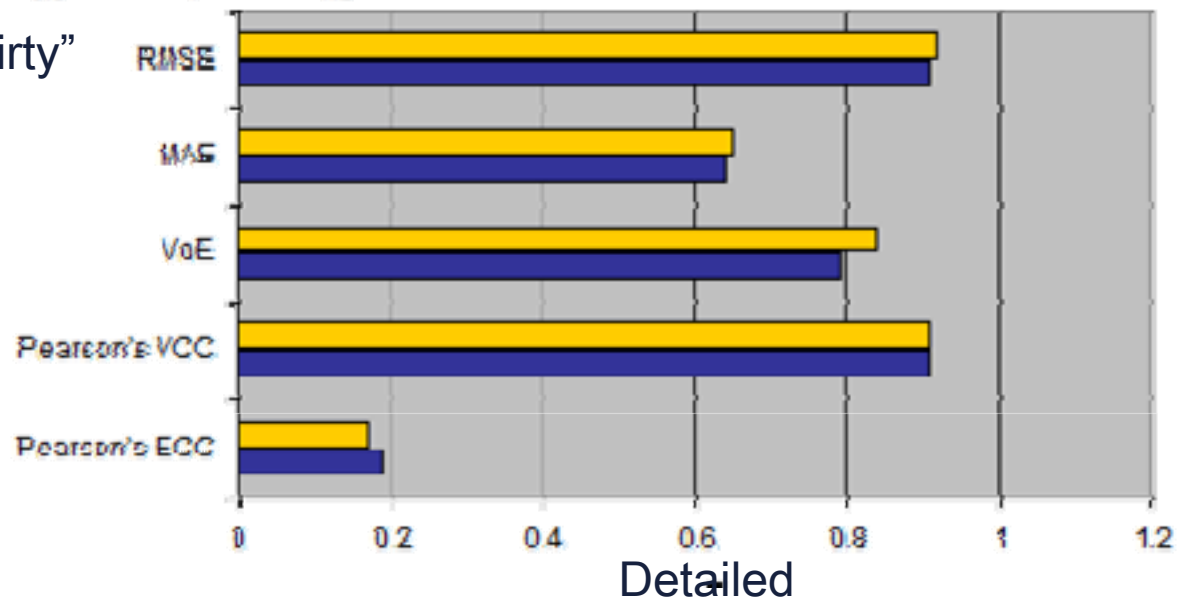


# Models comparison: Application 1 & GA

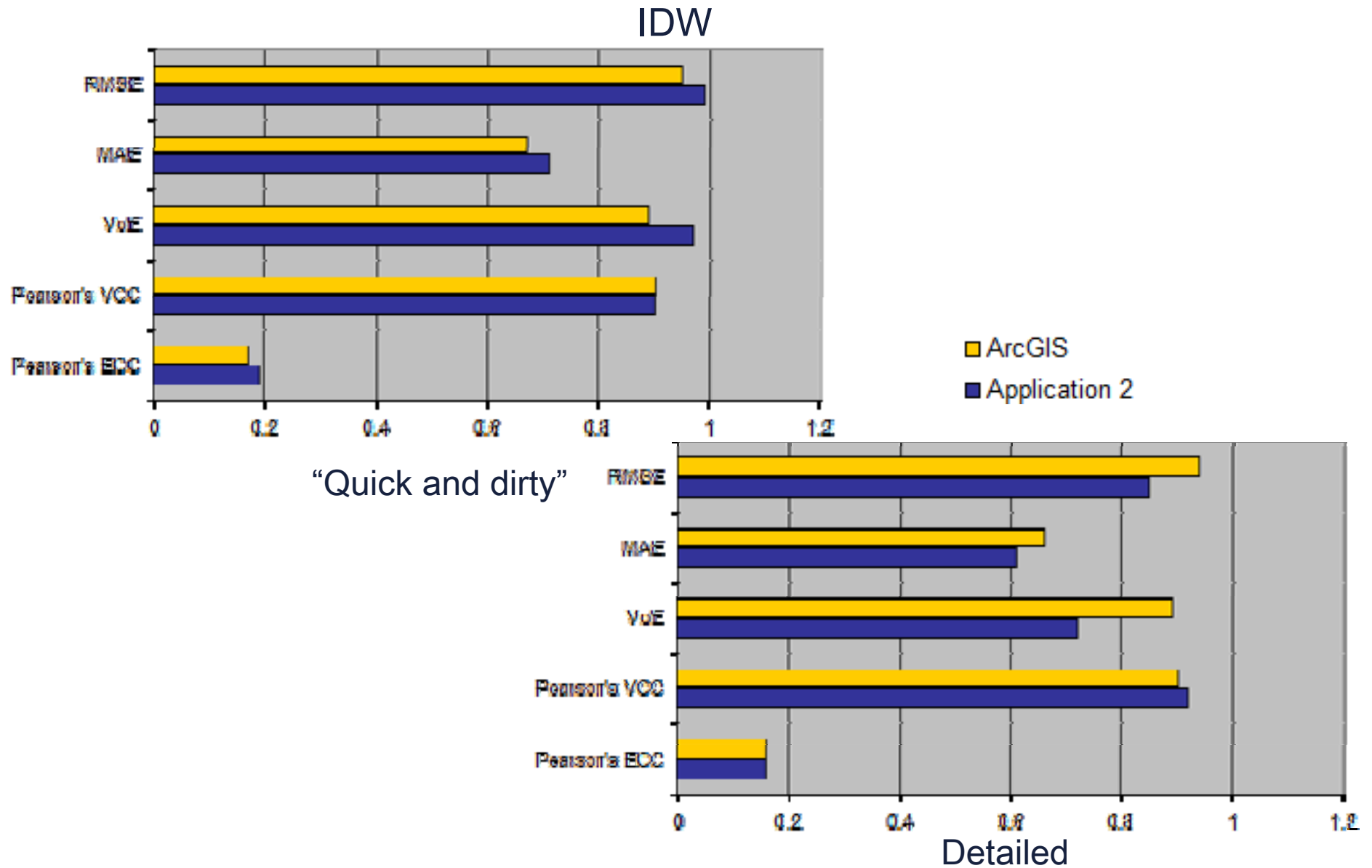
## Ordinary Kriging



"Quick and dirty"



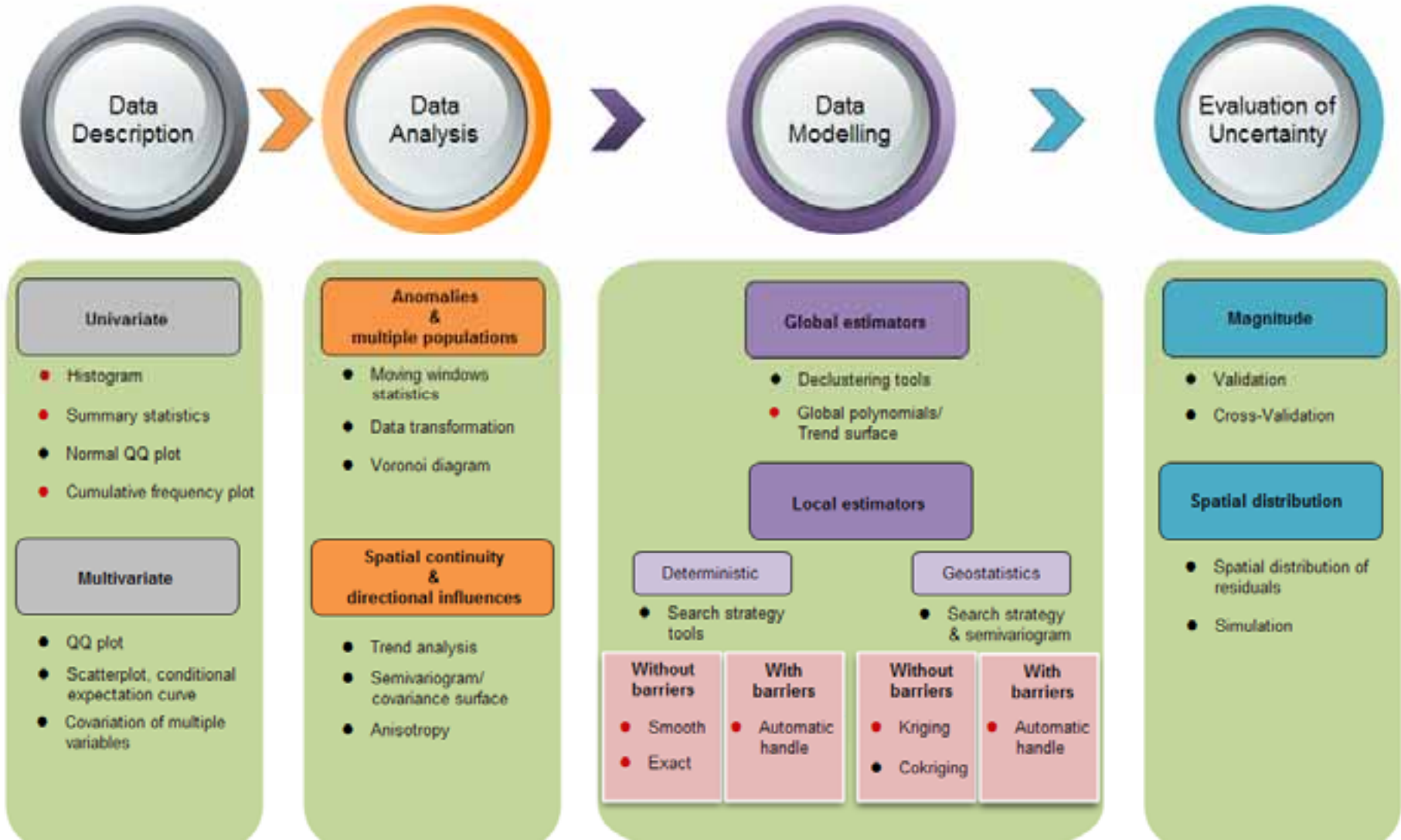
# Models comparison: Application 2 & GA



# Summary: Application 1 & 2

Key:

- Available in Application 1 & 2
- Not available in Application 1 & 2



# Summary

- GA 10 compares very favourably with established mapping applications in the E&P industry
- ArcGIS viable alternative to other mapping applications
  - Data-dependent
  - Goal-dependent
  - User knowledge-dependent
- Integrated workflow
  - GA 10 for advanced parameter customisation
  - Application 1/Application 2 for modelling (heavily faulted datasets)

## Some additional thoughts

- Invest time in data analysis before attempting interpolation
- Evaluation of uncertainty as a key step within the modelling process
- Investigate performance of software packages considering alternative “data scenarios”:
  - More sparse datasets
  - Well & 2D seismic data
  - More heavily faulted datasets



**exprodat**  
integrating information

**Thank you!**

**Paola Peroni**

**email:** [pperoni@exprodat.com](mailto:pperoni@exprodat.com)

**web:** [www.exprodat.com](http://www.exprodat.com)