New version of the Spatial Data Modeler tool: ArcSDM 5

ArcSDM 5
Final Seminar
May 4th 2018, Rovaniemi
Acknowledgements

• This presentation may include slides borrowed from the following persons with or without proper citation:
  – Gary Raines
  – Graeme Bonham-Carter
  – Carlos Roberto de Souza Filho
  – Stephen Gardoll

THANK YOU VERY MUCH GUYS!
ArcSDM 5
Background

- ArcSDM was originally developed by Gary Raines (USGS) and Graeme Bonham-Carter (GSC) and was coded by Don Sawatzky (USGS)
- ArcSDM has been maintained by Prof. Carlos Roberto de Souza Filho, University of Campinas, Sao Paolo, Brazil, http://www.ige.unicamp.br/sdm/
- MPM project has recoded the tools into ArcGIS 10.4 and ArcGIS Pro platforms and also implemented some new tools
ArcSDM5
New toolbox

- Runs on ArcGIS 10.4 and beyond + ArcGIS Pro
- Maintained by GTK + open source community
- Available from https://github.com/gtkfi/ArcSDM
GITHUB, ArcSDM 5 Installation

• https://github.com/gtkfi/ArcSDM
• Download ZIP file to your computer
• Add toolbox(es) into your ArcGIS map document
ArcToolbox

“Traditional” SDM Tools

Experimental SDM Tools
Spatial Data Modeller (SDM) toolbox

- Geoprocessing tools for integration of spatial data to predict the location to any features (i.e. mineral deposits, animal habitat, disease outbreaks ... etc).
- Fuzzy Logic
- Neural Network
- ROC Tool
- Utilities
- Weights of Evidence
- Logistic regression
Spatial Analyst toolbox

Fuzzy logic tools in Overlay toolset
- Fuzzy Membership
- Fuzzy Overlay
- Weighted Overlay
- Weighted Sum
Weights of Evidence

- Originally developed as a medical diagnosis system
  - relationships between symptoms and disease evaluated from a large patient database
  - each symptom either present/absent
  - weight for present/weight for absent \((W+/W-)\)
- Apply weighting scheme to new patient
  - add the weights together to get result
- Can be applied also to predict potential for mineral deposit based on exploration data (geophysical or geochemical anomalies, geology)
Weights of Evidence: Conditional Probability

- Data driven technique
  - Requires training sites
- Statistical calculations are used to derive the weights based upon training sites.
- Evidence (maps) are generally reclassified into binary patterns.

\[
\text{P(D | A)} = \frac{\text{P(D } \cap \text{ A)}}{\text{P(A)}} = \frac{\frac{\text{N(D } \cap \text{ A)}}{n}}{\frac{\text{N(A)}}{n}} = \frac{\frac{5}{1000}}{1000} = 0.005
\]

\[\text{N(D)} = 5 \quad \text{(in greenstone)}\]
\[\text{N(A)} = 1,000\]
\[\text{N(D } \cap \text{ A)} = 5\]
\[\text{N(total)} = n = 10,000\]
Traditional ’crisp’ logic vs. fuzzy logic

-> fuzzy sets

source: MathWorks
A geological example: Distance to Thrusts
A geological example: Distance to Thrusts
Problems with existing methods

Weights of Evidence

- statistically-based => can’t use in poorly-explored areas

Fuzzy Logic

- subjective judgement => difficult to reproduce
- deposit models but
  a) new deposit types found
  b) existing models revised
Why use neural networks?

pattern recognition

Advantages over statistical methods:
uncertain, noisy data
outliers
non-linear relationships
multiple interdependent parameters
multiple populations
mixed data sets - categorical & ratio
Converting GIS layers to feature vectors

GIS raster layers

- Rocktype
- EM
- Gravity Gradient
- Distance to Thrusts
- Training sites Deposits

Input feature vector

[3, 8, 33, 800]

target output

1
Model validation

- Efficiency of Classification
  - Training sites
- Efficiency of Prediction (Validation)
  - Sites not used for training
- ROC curves
- Jack-knife/Cross-validation
PRC: Efficiency of Prediction
SRC: Efficiency of Classification

• Intersect points with response grid.
• Plot Cumulative area versus cumulative number of points
• Calculate area under the curve.
  – Area under the curve for sites should be greater than 50% of total area, then have a positive association with points.
  – Area under the curve for “Not” sites should be less than 50% of total area, then have a positive association with points
  – If area under the curve, then have a random association with the evidence. Evidence provides no better information than guessing.
• Point in curve where goes from steep slope to flat slope is an optimal break between predicted sites and not sites.
The ROC curve

- The ROC curve is a graphical method for evaluating the performance of binary classifiers.
- The threshold value takes all possible threshold values.
- For each threshold value, the following rates are calculated:
  - False Positive Rate: $\text{FPR} = \frac{\text{number of false positives}}{\text{total number of negatives}}$
  - True Positive Rate: $\text{TPR} = \frac{\text{number of true positives}}{\text{total number of positives}}$
- The (FPR, TPR) points are plotted as a line graph.

A ROC curve which is equal to tossing a coin to determine the group (positive or negative).
ROC tool in ArcSDM

- ArcToolbox
  - 3D Analyst Tools
  - Analysis Tools
  - ArcSDM Tools
  - Fuzzy Logic
    - ROC Tool
      - Calculate ROC Curves and AUC Values
  - Utilities
  - Weights of Evidence
  - Cartography Tools
  - Conversion Tools
  - Data Interoperability Tools
  - Data Management Tools
  - Editing Tools
  - Geocoding Tools
  - Geostatistical Analyst Tools
  - Linear Referencing Tools
  - Multidimension Tools
  - Network Analyst Tools
  - Parcel Fabric Tools
  - Schematics Tools
  - Server Tools
  - Spatial Analyst Tools
  - Spatial Statistics Tools
  - Tracking Analyst Tools
Experimental SDM toolbox

• Created by ESRI Germany: Irvine Gabrera and Melanie Brandmeier.
• Based on ArcPy, Scikit-learn and Matplotlib Python modules
• Algorithms implemented
  – Adaboost
  – BrownBoost
  – Logistic regression
  – Random Forest
  – Support Vector Machine
Experimental SDM toolbox
BrownBoost model

• Steps in Experimental tools:
  1. First need to re-scale data values to 0 – 1
  2. Combine re-scaled grids using **Composite Bands**
  3. Create random points (not deposit sites, N = deposit sites)
  4. Enrich points -> extract grid values to points
  5. Train BrownBoost model
  6. Apply Brown Boost model
  7. Validate model
Experimental SDM toolbox
BrownBoos model

- Rescale data
- Enrich points
- Train model
- Apply model
Experimental SDM toolbox
BrownBoost
Experimental SDM toolbox
BrownBoost

• Validation done using ROC method
• AUC=0.847
ArcSDM5
Summary

• Runs on ArcGIS 10.4 and beyond + ArcGIS Pro
• Maintained by GTK + open source community
• Available from https://github.com/gtkfi/ArcSDM