An ArcGIS ModelBuilder Application for Comparing Interpolation Methods

Kruk, Lukasz; Limbu, Minu Kumar; Okyay, Unal; Cabral, Pedro; Costa, Ana Cristina; Painho, Marco
{m2010157@isegi.unl.pt, m2010162@isegi.unl.pt, m2010164@isegi.unl.pt, ccosta@isegi.unl.pt, pcabral@isegi.unl.pt, painho@isegi.unl.pt}
Instituto Superior de Estatística e Gestão de Informação, Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa

INTRODUCTION

The utilization of mathematical and statistical functions for advanced surface modelling using deterministic geostatistical interpolation; techniques have been regularly put into test in different fields of sciences. Studies have also proved that the performance of interpolation techniques heavily depends on the type of modelled phenomena, geometrical configuration of the samples, spatial resolution, world region and others (Martínez-Cob, 1996; Goovaerts, 2000; Haberlandt, 2007). Therefore, the process of choosing the right interpolation technique for a given set of data points (samples) is not always easy – especially because it directly or indirectly influence the generation and visualization of new interpolated surfaces. This study focuses on the use of ModelBuilder, an ArcGIS application, as an approach to compare and understand how different interpolation methods perform or behave for different input data sets.

The main objective of this study is to develop a structured, user friendly and re-usable framework to facilitate easy comparison of different interpolation methods as well as provide a quick exploratory analysis tool. The aim of the model is to support quick evidence based evaluation of the datasets without having to go through the tedious process of producing detailed analysis. After running the model, the user is presented with a number of different views of the dataset allowing for quick grasp of the nature of modelled phenomena.

THE MODEL

The model (Fig. 1) uses a point shapefile with point attributes (e.g., elevation, temperature, etc.) as an input file. The user can further divide the input data into a training dataset for interpolating surfaces and a validation dataset for error estimation. There is a provision where the user should specify different environment settings on how exactly the output should look like. Once the parameters are set, the model automatically generates series of interpolated surfaces, difference maps and a point density map.

Based on the input data points, the interpolation model automatically generates a set of interpolated surfaces (Inverse Distance Weighted (IDW), Inverse Square Distance (ISQ – IDW with power of two), Spline and Natural Neighbourhood (NN)), difference maps (ISQ-IDW, ISQ-NN and ISQ-Spline) and error tables (Root Mean Square Error (RMSE) and Mean Error (ME)) that can be used as objective methods of determining the interpolation quality. Furthermore, performance of different techniques can also be compared through the point density maps – where the user can individually specify output raster parameters and identify how point density influences different interpolators.

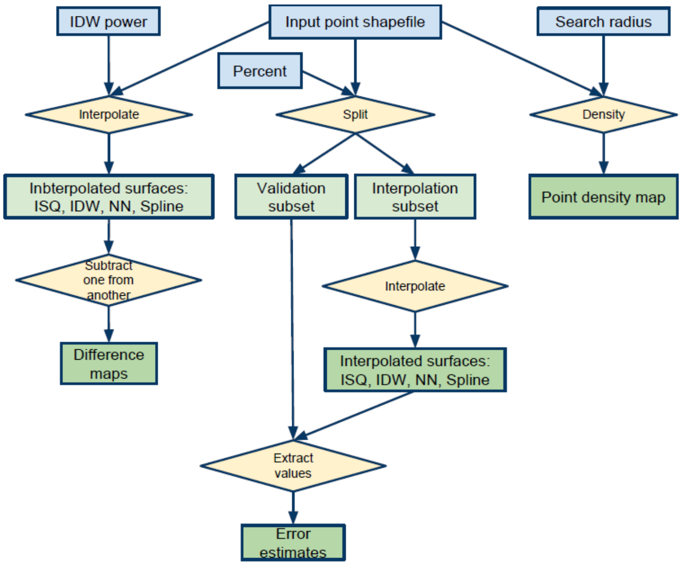

Figure 1: Model Flowchart

Primary outputs of the model are the interpolated surfaces produced with deterministic techniques such as ISQ, IDW, Spline and NN. It also generates three difference maps through raster algebra by subtracting the interpolated surfaces from one another: ISQ minus IDW, ISQ minus NN and ISQ minus Spline. When two compared interpolated surfaces are similar, the resulting value is close to zero, while for bigger differences between the two surfaces result in significantly higher or lower values based on the positivity or negativity of the differences. The difference is given in the same unit as the unit of the interpolated phenomena. Additionally, a point density map is generated for the analyst to be able to see if areas with higher point concentration result in more accurate estimations. The model also uses the separated validation datasets and generates error estimates. Tables of RMSE and ME are created using summary statistics. Hence, it can be further used in making informed decision about different interpolations. Some of the observations made during the testing of the model are documented in the following paragraph.

ILLUSTRATION OF THE MODEL APPLICATION

When tested with three different elevation sample patterns: regular grid - dense (consists of 3740 points 250 m apart) and sparse (consists of 242 points 1000 m apart) and randomly distributed points (consist of 203 points), the model outputs differ significantly. For the regular grid - dense samples, given the high degree of the input data, all the interpolators performed rather similarly and similar interpolated surfaces are produced. However, when the input datasets are sparser, the differences between interpolators are more apparent upon selection of correct interpolators and its parameters. From the difference maps, Spline produced more exaggerated (pronounced) local extremes.

The model, however, does not implement Kriging approaches as the model is designed for quick assessment of interpolated surfaces based on the input datasets whereas Kriging requires a thorough semivariogram analysis prior to the interpolation. The model cannot automatically apply custom symbology to raster maps – it is a manual process. A logarithmic scale symbology for difference maps files are provided as a starting point, but this is not an ideal solution. Also, the point density tool used doesn't allow the user to specify custom search radius. Apart from these, to help get the best out of the model, 'how to' user instructions are provided under the help menu of model's dialog box.

CONCLUSION

This paper describes the main features of an ArcGIS ModelBuilder application for comparing interpolation techniques. Furthermore, a first iterative model for comparing different interpolation methods is presented. There are many studies conducted to compare different interpolation methods. Just to compare three different interpolators - IDW, Spline and NN, the model generates eight different raster maps including interpolated surfaces, three difference maps and a density map. The model, however, is not designed to produce detailed investigative analysis of the input datasets. Even when the model is on its initial/first iteration development phase, it can easily and accurately provide quick assessment of the characteristics of input data through interpolated surfaces.

ACKNOLEDGEMENT

The model was developed as a part of GIS Application end semester project at the Universidade Nova de Lisboa, Erasmus Mundus Master of Geospatial Technologies. Presented example dataset analysis was carried out using the model as the final project assignment for Geostatistics course.

The authors would like to thank their colleagues at the university for helping in testing the model. The authors are also indebted to 'AI geostats' website for hosting the model which will be freely available for download at a later date*[currently under consideration, URL to be added before AGILE 2011 conference]*.