Improving the Accuracy of Geo-Processing through a Combinative Computation Method

Zhiwei Cao[1], Mordechai (Muki) Haklay[2], Tao Cheng[3]

[1] Civil, Environmental and Geomatic Engineering Department,
University College London (UCL),
Gower Street, London, United Kingdom. WC1E 6BT
Email: z.cao@ucl.ac.uk

[2] Civil, Environmental and Geomatic Engineering Department,
University College London (UCL),
Gower Street, London, United Kingdom. WC1E 6BT
Telephone: 0044-20-7679-2745
Fax: 0044-20-7380-0453
Email: m.haklay@ucl.ac.uk

[3] Civil, Environmental and Geomatic Engineering Department,
University College London (UCL),
Gower Street, London, United Kingdom. WC1E 6BT
Telephone: 0044-20-7679-2738
Email: tao.cheng@ucl.ac.uk

1. Introduction

This paper describes how to build on previous work on Combinative Geo-processing (Cao et al., 2010). The primary objective of Combinative Geo-processing is to reduce the influence of data uncertainties of traditional geo-processing that are caused by error propagation due to processing of geographic information.

Combinative Geo-processing is inspired by the concept of Map Calculus (Haklay 2004), in which Geographic Information Science (GIS) layers are stored as functions and new layers are created through a combination of existing functions. For example, instead of generating a high-resolution grid from a point dataset by using some interpolation function, all functions and their related data are stored in a 'suspended' status in Map Calculus, and the resulting grid is only generated when the user requires it for visualisation or as output.

The purpose of geo-processing is to solve complex spatial problems through a range of computation tasks. For instance, to acquire slope values of a land surface , users may need two computation tasks: the first task is to generate elevations from a set of raw Light Detection And Ranging (LiDAR) points, and the second task is to evaluate slope values from the elevations. In fact, the traditional geo-processing method allows GIS users to chain together tasks based on a specific workflow, which loads the output of one task into a subsequent task. As a result, GIS users can use this ability to create different geo-processing models in order to solve complex spatial problems.

Although the traditional geo-processing method offers a powerful mechanism for GIS modelling and decision-making purposes, it has been repeatedly acknowledged in the literature that there are many problems associated with the traditional geo-processing method (Heuvelink 1998; Krivoruchko and Gotway-Crawford, 2003; Devillers and Jeansoulin, 2006) .

2. Combinative Geo-processing

With the aim of improving data quality in geo-processing, Combinative Geo-processing is introduced in this paper. Instead of the traditional geo-processing method, we propose the use of function-based layers to deal with geo-processing. The function-based layer has three characteristics: (a) it can store a single function, e.g. a function to calculate distance, or a set of functions (a template), e.g. the geo-processing model to analyse complicated objects; (b) the input of a function-based layer could be a dataset or a function (or functions), (e.g. an IDW function can be directly input to a SLOPE function to produce a function-based layer); (c) the output of a function-based layer could be a dataset or a function (or functions).

After applying the function-based layers, Combinative Geo-processing produces the output on time, together with symbolic computation. Figure 1 illustrates the generic frameworks of the traditional geo-processing and Combinative Geo-processing. The upper flow chart represents a common way in the traditional geo-processing framwork, where the initial step loads 'Input' into a function 'Fun1', and then the function 'Fun1' gives an output 'Out1'. Then, the 'Out1' is re-used as an 'Input' in the following function 'Fun2' and returns a result 'Out2'. As it can be also seen a sequential computation strategy is used until to a final result. In contrast, the bottom chart describes a completely different way which can be used to implement a set of functions in Combinative Geo-processing, where a function can take other functions as parameters and returns functions as results.

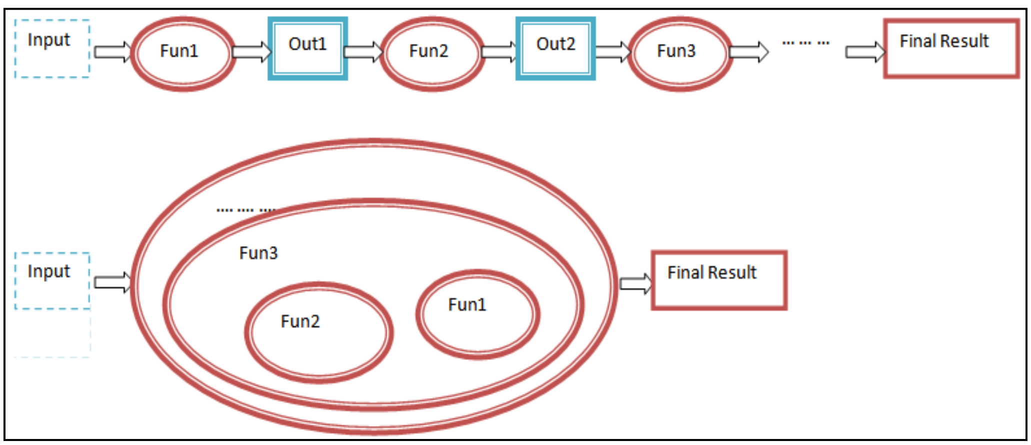

Figure 1. The generic frameworks of the traditional geo-processing method and Combinative Geo-processing.

The first attempt to develop Combinative Geo-processing is described in a previous paper (Cao et al., 2010), and the purpose is to examine the basic ability of our implementations to deal with the single layer and to compare its performance to raster-based implementations. This was done through the comparison of the implementation of the Inverse Distance Weight (IDW) function between our method and GIS packages, including ArcGIS, MapInfo, Manifold GIS and R. The results provide confidence regarding the development of the new method and the accuracy of our IDW implementation. Based on the results from the single function analysis, the aim of this paper is to focus on the multi-functions operation.

3. Case Study

To illustrate how to use Combinative Geo-processing in processing multi-functions, a simplified GIS model is used. The model involves two computation tasks or functions, such as 'Elevation' and 'Slope'. The objective of the model is to to understand the slope features in sampling locations, so there are 2,000 specific locations which are randomly selected.

This case study calculates the slope using two different methods, which involve the traditional geo-processing method (ModelBuilder tool in ArcGIS) and Combinative Geo-processing (using Scheme programming tool). Additionally, a Monte Carlo simulation method is adopted to investigate data quality problem.

Figure 2 and Figure 3 show the computation models in these two methods. Notably, we used one metre grid size in ArcGIS to produce elevation and slope data. In ArcGIS (Figure 2), there are three main steps: (a) Loading the raw LiDAR points to IDW function; (b) Input the IDW's result (output1) to slope function, then get slope result (output2) on raster representation; (c) Mach between Output2 and Specific Locations to capture the slope value on specific locations. In contrast, the major steps in Combinative Geo-processing method include (Figure 3): (a) Loading the raw LiDAR points to the combination of IDW and Slope functions; (b) Input a pair of X, Y coordinate value to generate the slope value of the requested location.

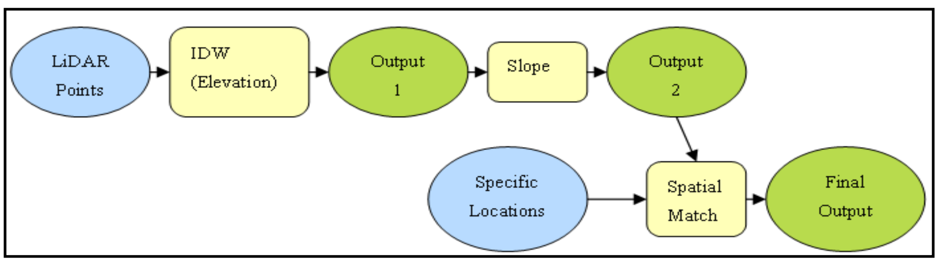

Figure 2. The computation model (in ArcGIS).

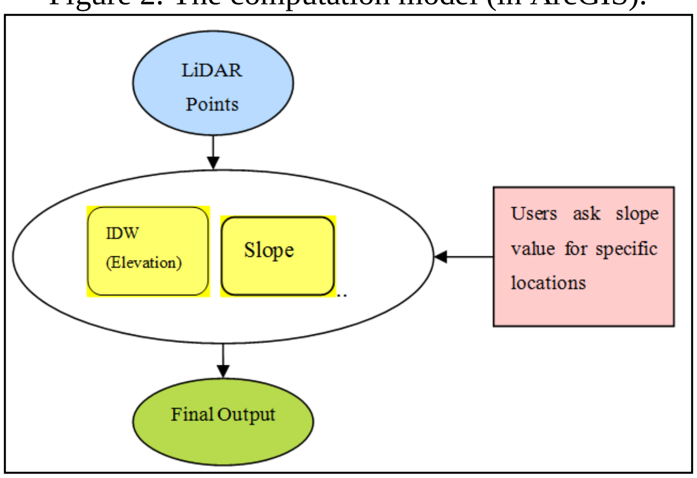

Figure 3. The computation model (using Combinative Geo-processing).

After running the models in ArcGIS and Combinative Geo-processing, Figure 4 displays the slope results according to randomly selected sampling locations. In this figure, the point value represents the slope features from lower degree (white) to higher degrees (black). The results show that there are many different values crossing the study region.

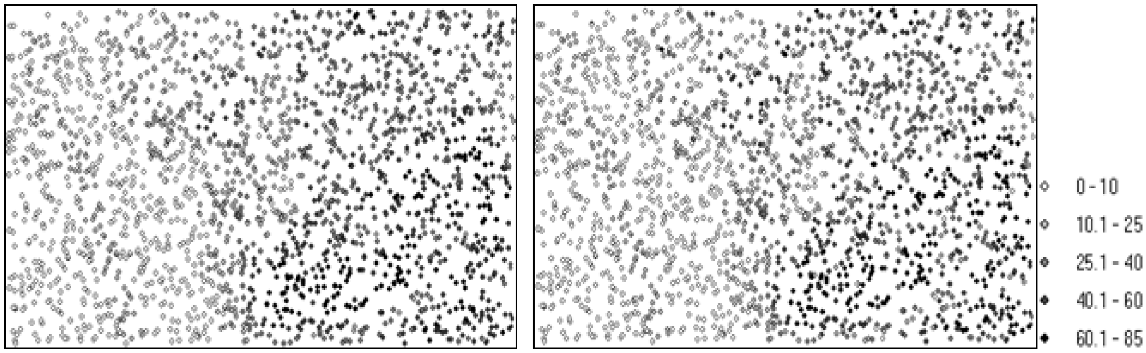

Figure 4 (a). Combinative Geo-processing result (slope values). (b) ArcGIS result (slope values).

In order to visualize the difference between the two outputs, Figure 5 is created by the subtracting operation (Slope$_{\text{Combinative Geo-processing}}$ - Slope$_{\text{ArcGIS}}$). From the legend of this figure, it is clear that there is a significant difference between Slope$_{\text{Combinative Geo-processing}}$ and Slope$_{\text{ArcGIS}}$; for example, the maximum difference value is 61 degrees (the highest value in the legend) and around 30% of specific locations (more than 600 sample location points) differ in more than 10 degrees.

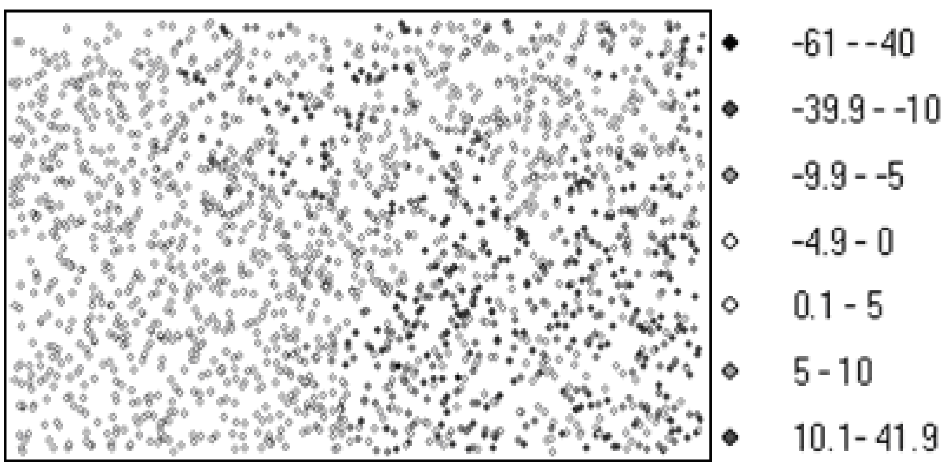

Figure 5. Differences in the slope values (after subtracting operation).

The diagram of Monte Carlo Simulation is given in Figure 6. This figure is based on the computation in Monte Carlo Simulation, which includes the distribution and variance of sample mean, maximum and minimum values. Parameters of the figure are fitted with the index of specific locations and their values. Figure 6 (a) shows that ArcGIS has a significant distribution of the values, which indicates that the slope results from ArcGIS have more noises and larger variances than the Combinative Geo-processing results.

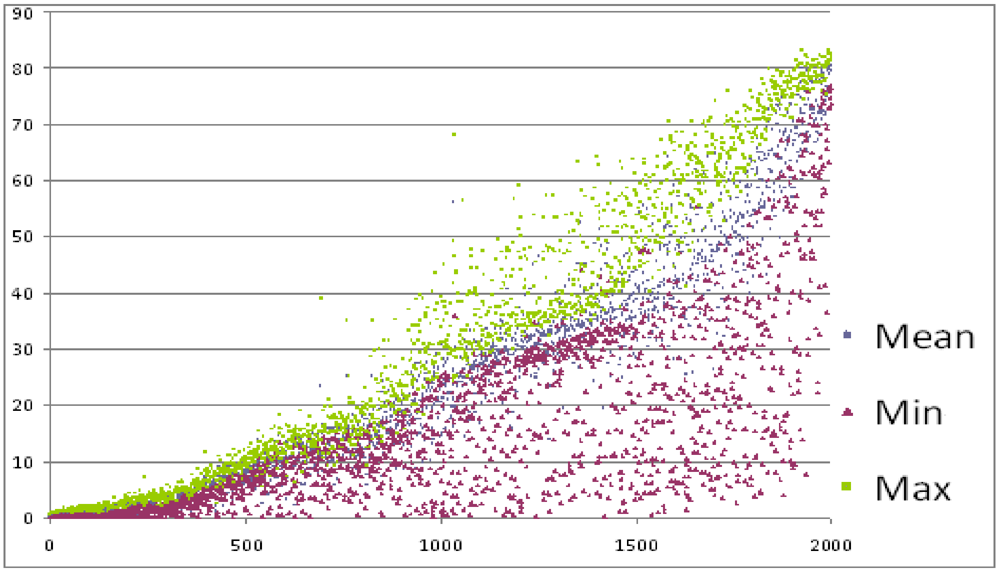

Figure 6 (a). ArcGIS result (Monte Carlo test).

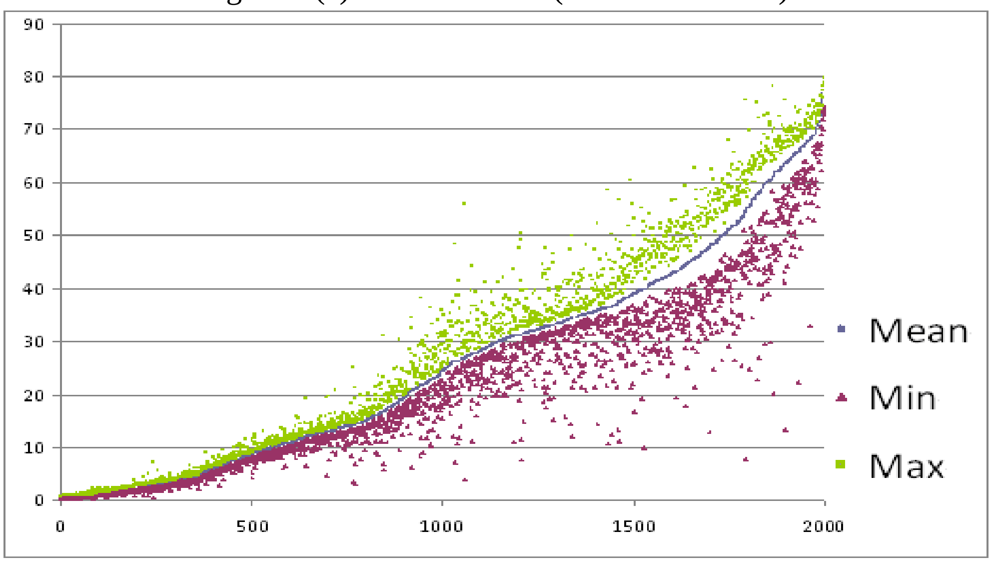

Figure 6 (b). Combinative Geo-processing result (Monte Carlo test).

In addition, the map of the resulting standard deviations in the randomly selected sampling locations is given in Figure 7. In statistics, standard deviation is widely used to measure variability and probability. The results appear that standard deviations can be quite large when near the highest elevation area (south-east), especially for the ArcGIS's results (Figure 7 a).

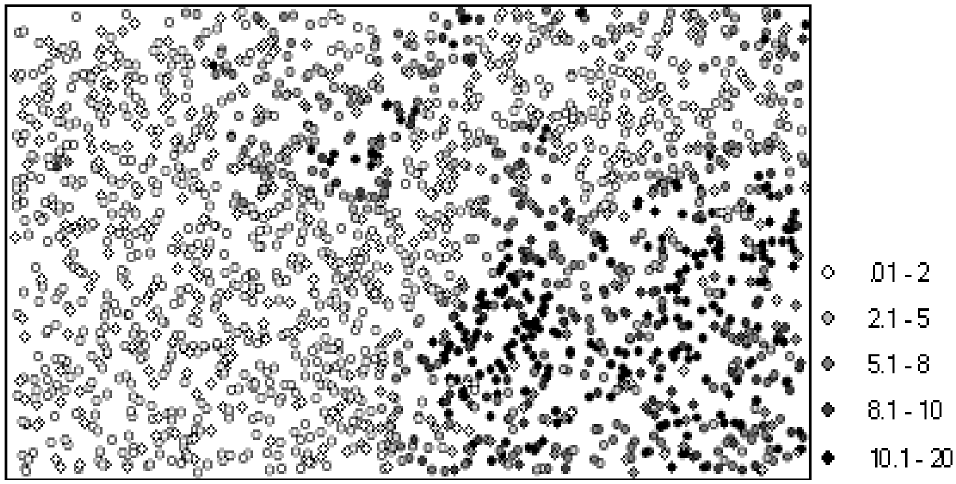

Figure 7 (a). Standard deviation result (ArcGIS).

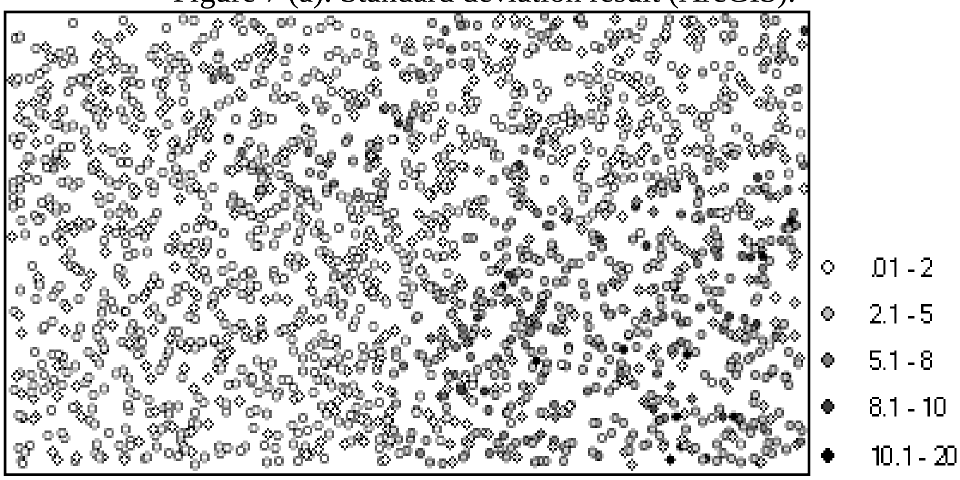

Figure 7 (b). Standard deviation result (Combinative Geo-processing).

4. Summary and Conclusion

In this research, using the comparison of slope implementations introduce the idea of Combinative Geo-processing and demonstrate its advantages over traditional geo-processing method, especially on multi-functions operation. The case study shows that the process in ArcGIS is a common computational strategy which executes functions sequentially and based on numerical computation. In contrast, Combinative Geo-processing offers an improved way to compute slope and provides better performance in Monte Carlo simulation.

Finally, we are aware that Combinative Geo-processing method implemented in this research is still at an early development stage. There are a variety of issues which will directly influence the development. For instance, the current limitation of our method is that it is a time intensive operation when huge datasets and complex models are involved. Therefore, future work on the 'integration' and 'evaluation' rules will be undertaken with the aim to improve the computation efficiency, such as implementing the intelligent simplification and priority algorithms on the function-based layers.

5. References

Cao, Z. Haklay, M. and Cheng, T. (2010) Evaluating the accuracy of IDW implementation in Map Calculus and commercial GIS products. GIS Accuracy Conference 2010, Leicester.

Devillers, R. and Jeansoulin, R., (2006). Fundamentals of Spatial Data Quality. Wiley-ISTE.

Haklay, M., (2004). "Map Calculus in GIS: a proposal and demonstration." International Journal of Geographical Information Science, 18(2004), pp.107-125. Available at: http://www.tandfonline.com/doi/abs/10.1080/13658810310001629583.

Heuvelink, G.B.M., (1998). Error Propagation in Environmental Modeling with GIS. London: Taylor & Francis.

Hunter, G.J. and Goodchild, M.F., (1997). "Estimates Derived from Spatial Databases." Geographical Analysis, 29(1).

Krivoruchko, K. and Crawford, C.A.G., (2003). "Assessing the Uncertainty Resulting from Geoprocessing Operations." In D. Maguire, M. Batty, & M. Goodchild, eds. GIS,Spatial Analysis,and Modeling. ESRI Press, pp. 62-67.