*Accuracy 2004: The 6th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Science*, Portland, Maine, June 28-July 1, 2004

# Multicriteria Line Simplification (MCLS) for AEM Groundwater Modeling

**Gaurav Sinha[1] , Warit Silavisesrith[1], James R. Craig[2] & Douglas M. Flewelling[1]**
1. Department of Geography
2. Department of Civil, Structural & Environmental Engineering

University at Buffalo, NY, USA

## Abstract
Some regional scale environmental simulation models require generalization of high resolution data to obtain an acceptable numerical solution within a reasonable interval of time. This paper evaluates the efficacy of geospatial data reduction for input to a polylinear groundwater model that relies on the vector-based Analytic Element Method (AEM) to represent and simulate a groundwater system. Methods for geospatial data generalization have been developed in the past but mainly for addressing cartographic concerns. We therefore suggest a new approach—Multicriteria Line Simplification (MCLS) that simplifies polylinear analytic elements under constraints determined both by cartographic and geophysical considerations. These constraints are derived from experts' domain knowledge and realized within the simplification system as multiple interactive criteria. In a case study, the tradeoff between computation time and the errors introduced in model predictions is analyzed at several simplification levels for different weighted combinations of MCLS criteria. The results are used to discuss future changes in this framework.

**Keywords**: Multicriteria Line Simplification, Analytic Element, Groundwater

## 1. Introduction
Regional scale environmental simulation models mathematically simulate real-world geophysical processes based on knowledge of both the dominant processes and the mathematical approximations appropriate for numerical modeling. Besides knowledge of the process, environmental input is also crucial for the success of the simulation. Geospatial databases are therefore queried for accessing environmental data essential for simulation and calibration purposes. In the last decade, due to technological advancements and a surfeit of application areas, these databases have become quite large. From an environmental modeling perspective, large high resolution datasets are not always desirable; in fact it may become an encumbrance if less detailed data can suffice.

One area therefore, where environmental modelers desire help from data modeling schemes is geospatial data reduction. Reducing high resolution geospatial data to a tractable simplified numerical model configuration ensures that simulations converge to sufficiently accurate results within a reasonable interval of time. The level of detail required for simulation models is a dynamic variable, in that the level of detail varies with the geographic sub-domain being resolved. A general modeling approach begins with relatively coarse data and demands higher resolution only for certain geographical areas and/or time intervals identified as interesting during the initial stages of analysis. This general paradigm is the underpinning of most computational environmental models, regardless of the data model in use or the geophysical phenomenon being simulated. The need for data reduction is important not only due to practical computational constraints, but from the standpoint of cognitive efficiency as well. Preservation of the essential characteristics of dataset facilitates conceptual modeling. Our ability to conceive processes within a limited range of spatio-temporal scales is reflected in the way simulation models are designed and discretized.

In this paper we use the Analytic Element Method (AEM) for groundwater modeling to explore the significance of geospatial data reduction in environmental modeling. AEM models are complex numerical models comprised of vector based hydrographic features that can benefit immensely from simplification based data reduction methods. These models often depend on hydrographic, geologic and hyposgraphic data from geographic databases. Since linear features tend to dominate the hydrographic database extensively, we provide in this paper a new methodology called Multicriteria Line Simplification to automate the simplification of linear hydrographic features based on spatial, topological and hydrogeologic constraints. Results show that simplification of polylinear elements reduces simulation time considerably. Model error incurred from numerous different simplification schemes is evaluated to show that not only the degree of simplification but the choice of criteria and constraints affect model prediction as well.

## 2. Analytic Element Simplification

Like many other models of geophysical phenomena, groundwater flow models often rely on numerical simulation. One method of simulation, the analytic element method (AEM) (Strack, 1989; Haitjema, 1995) is particularly appealing as a test subject for data reduction/simplification techniques because it uses a vector representational scheme. In the analytic element method, a numerical model is composed of polylinear, polygonal, and point "analytic elements" which represent hydrogeologic features, such as streams, lakes, pumping wells, and zones of different aquifer properties. These analytic elements often directly correspond to map features in geographic databases. Using the geometry and attributes of these hydrogeologic features as boundary conditions, AEM simulates the distribution and magnitude of hydraulic head and groundwater fluxes in the subsurface through solution of a physics-based partial differential equation.

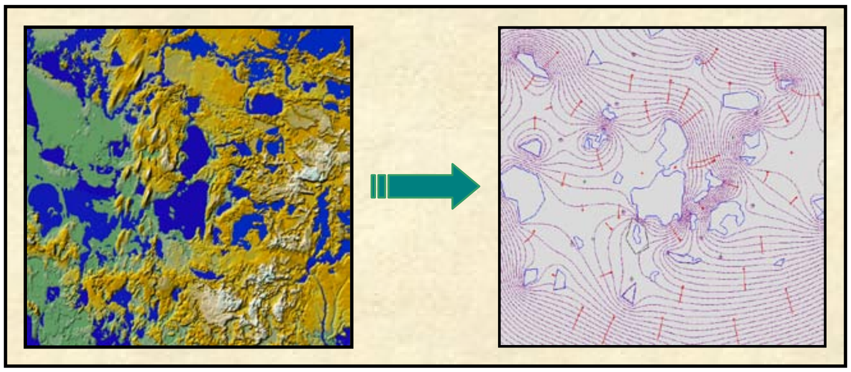


Figure 1. Hydrographic features and representative analytic elements

The complexity of an AEM model is a function of the geometric and parametric model configuration and the total number of elements in the system. Higher complexity typically translates into higher computational costs and undermines comprehension of groundwater system interactions; but it also tends to produce more accurate results. Highly detailed geometry (i.e., more polylinear element segments), however, does not necessarily translate into more accurate models, primarily due to the unknown makeup of the subsurface and the often minimal impact of minor geometric variations. Thus, there are dual benefits to AEM model simplification: the generalization process can boost model performance and improve the level of confidence in the generality of the results.

Because of the geometric model (points, lines, polygons) used for representing geographic features, geographic databases are predominantly populated with linear features (McMaster & Shea, 1992). Since there is a correspondence between geographic reality and the AEM groundwater model, polylinear elements which represent surface linear hydrographic features are the most important class of analytic elements. Hence, in this paper, we focus exclusively on the simplification of polylinear elements.

Each polylinear element is a composite of segments topologically connected through nodes. This topological connection is required to maintain the continuity of flow of water down "stream" elements or to ensure consistency of boundary conditions along "inhomogeneity" elements (Strack, 1989). Other inter-element and intra-element consistency constraints are also required for effective solution. This means that purely geometric method of line simplification (as in cartography and computer vision) will generally not suffice in the context of AEM model simplification.

## 3. Cartographic Line Simplification

Polylinear analytic elements can be derived from geographic databases by identifying the corresponding linear hydrographic features (e.g. rivers and streams) and linking their geometry to analytic elements. Geographic data are geometrically modeled in GIS through three spatial primitives: points, lines and polygons. Lines are ordered sets of points, and polygons are generally constructed as boundary enclosed areas. The problem of polylinear analytic element simplification therefore translates into the reduction of the arity of the set of geometric points that comprise a line in a GIS database.

The simplest way to select points from a line automatically is to treat all points as having equal probability of selection and choose random points or every $n^{th}$ point (Lang, 1969). However this is too simplistic an approach and assumes an equiprobable field of selection-potential for all points. Automated line simplification has progressed much beyond this simple concept in the digital age. The philosophy of line simplification and its automation has been discussed extensively by cartographers (Buttenfield, 1985; McMaster, 1987; Beard, 1991). Many algorithms have been proposed to automate lines; McMaster & Shea (1992) provide a good discussion and classification scheme for line simplification algorithms. Line simplification algorithms tend to concentrate on preserving the shape and positional accuracy of the original line. The line is treated as a perceptual phenomenon (Buttenfield, 1985) and characteristics (e.g. bends, sinuosity, position) that control the visual perception of the line are paid the most attention. In the past few years, the scope of line simplification philosophy and its algorithms has been broadened by considering the geometric context to guide the actual choice of simplification operators (Plazanet, 1995; Wang & Muller, 1998; Skopeliti & Tsoulos, 1999).

There are two reasons however, why cartographic line simplification operators (algorithms) should not be directly applied to polylinear element simplification. First, the cartographic view of the line as only a geometric entity is inconsiderate of the domain specific treatment of the line; for process modelers, the actual nature of the geographic feature as it exists and interacts with other features in its real-world setting is more important for modeling purposes. This is reflected in the context of AEM model simplification, as the position, spatial configuration and hydrographic properties of polylinear elements are all important input parameters in AEM modeling. Groundwater modelers have, besides shape and position, many other factors (e.g. hydraulic conductivity, proximity to pumping wells, topographic slope) to consider before deciding on the optimal geometry of the problem.

The second impediment arises from the criteria used to evaluate the success of the simplification process. Cartographic evaluations involve a consideration of only vector and coordinate based errors (McMaster, 1986; Skopeliti & Tsoulos, 2001), but for groundwater modeling, in general, these maybe treated as only partial indicators of accuracy and efficiency. The real indicators of success are simulation time and prediction errors (e.g. hydraulic head, flow direction, contaminant concentrations and migration directions). The accuracy of predictions made with numerical models, depends upon input data, the size of the spatio-temporal discretization scheme, and the numerical method used to solve the governing equations. The accuracy of model predictions and simulation time are the two most important criteria that AEM modelers use as indicators of success. In fact, positional accuracy and shape of line segments is of secondary importance, if other more significant constraints on the simplification system can make the model more accurate.

## 4. Spatial and Semantic Constraints on Line Simplification

There are numerous spatial and semantic constraints that can be imposed on a line simplification system. Beard (1990) used a rule based approach to include multiple constraints for map generalization (i.e. the process of reducing detail on maps to facilitate better comprehension). More recently Weibel and Dutton (1998) have also advocated the use of graphical, topological, structural, Gestalt and process kinds of constraints to reduce the number of acceptable solutions and as map design guidelines. All these constraints should play an important role in AEM model simplification as well, especially process based constraints that encompass all the domain semantics and will not be considered by cartographers normally. The idea of constraints based simplification is the basis of the method used to simplify line analytic elements in this paper.

Any process of line simplification begins with identification of relevant constraints. Some of the most important constraints for polylinear analytic element simplification are as follows: i) shape of the line and position of characteristic points; ii) all topological connections between elements have to be maintained, to ensure that water always flows downstream and that tributaries and distributaries connect to larger streams at exactly the same location; iii) pumping wells should remain on the same side of the line; iv) known geophysical laws (e.g. water only flows downhill) always have to be obeyed regardless of the simplification process chosen and v) higher resolution should be maintained near locations of 'semantic events' (e.g. sudden change in conductivity, significant change in temperature, etc.). There are many more specific constraints that will appear on a careful analysis of the hydrogeologic and numerical modeling aspects of and given groundwater system; however we will limit ourselves to discussing only a few in this paper for illustrative purposes. In the next section we describe a framework that relies on decision making techniques to solve problems with constraints such that mathematical programming is not feasible.

## 5. Spatial Multicriteria Analysis

It is desirable to have a process based on understanding (Richardson, 1996; Wang & Müller, 1998) while generalizing* spatial data. A careful listing of all applicable constraints and desired model characteristics fosters such an understanding. It is also true that modelers would prefer to automate whatever little they consistently repeat in their model development phase. Hence the knowledge required to generalize geospatial data in general has to be encoded into the automatic simplification process. Not only does it free up their time for analysis but it provides an element of repeatability so crucial to progress in science. Cartographers have tried to use rules and knowledge databases to solve this problem for map generalization (Buttenfield & McMaster, 1991), but the process has proved to be generally intractable till now. For now, the AEM modeler will have to be involved in the process of analytic element simplification because a number of conflicting constraints that cannot be resolved without expert intervention arise during data reduction.

Given this paradoxical status of geospatial data reduction, we chose to depend on the Multicriteria Decision Making (MCDM) (Zeleny, 1982; Voogd, 1983) paradigm for geospatial data reduction for environmental modeling. The basic philosophy of MCDM is to help a decision maker select the 'best' alternative from the number of feasible choice-alternatives when one is faced with multiple criteria and diverse criterion properties (Jankowski, 1995). This paradigm of resolving conflicts is well established and is widely used in spatial analysis and environmental planning as well. Malczweski (2000) provides a comprehensive introduction to spatial multicriteria analysis in GIS.

In the next section we describe the multicriteria line simplification system. Note that multicriteria analysis can be dichotomized into multiobjective and multiattribute analysis. Some authors (Cromley & Morse, 1988; Cromley & Campbell, 1991) have used mathematical programming to optimize line simplification under constraints, but the method is not easily adaptable for line element simplification. There is no way (and perhaps no need) to provide complex and untenable mathematical equations and constraints at a stage when we are merely exploring the nature of the simplification system. We rely on heuristics instead to guide us and will depend only on multiattribute decision making (MCDM) (Malczweski, 2000; Yoon & Hwang, 1995) to develop our methodology.

## 6. Multicriteria Line Simplification Methodology

Based on our discussion of the polylinear analytic element as a set of topologically ordered points, and because none of the sophisticated line simplification algorithms allow inclusion of semantic criteria in the simplification system, the line simplification process is conceived here as a selection (or elimination) of points from the original set comprising a line. However, we will not just select the $n^{th}$ point on the line as Lang (1969) did. Through spatial multicriteria analysis we will create an information field of varying probability based on the spatial and semantic constraints applicable to the system; the multicriteria score determines the probability of a point's selection (or elimination).

We begin our multicriteria analysis by defining each point as an alternative that has to be evaluated with respect to constraints and criteria available in the simplification system (figure 2).

The next step is to introduce constraints and criteria of evaluation into the system. We have already provided some constraints and we will use them to select criteria on which to evaluate each of the alternatives (point). Criteria should be well disaggregated, diverse, small in number, amenable to assignment of numerical values (preferably interval or ratio scale), and uncorrelated for maximizing information retrieval (Voogd, 1983). Sinha & Flewelling (2002) for specifically discuss the process of identifying multiple criteria in a multicriteria line simplification context.

In multiattribute analysis, criteria and alternatives make up a 2-dimensional matrix; if alternatives are arranged in rows ($a_{i*}$) and criterion in columns ($c_{*j}$), each cell value ($x_{ij}$) of the matrix indicates the level of contribution each criterion *j* makes toward an alternative *i*. Since such contributions measured for different criteria can be on disparate scales, the raw scores for each criterion are usually standardized to fit a range of 0 - 1, using one of the many techniques available for standardization. Readers should refer to Voogd (1983), Malczweski (2000) for traditional and Wilson & Martinez (1997) for newer methods of standardization. The particular

* 'Generalization' is the term used by geographers for the process of reducing the detail of geographic data through automated heuristics primarily for optimizing visual information content of maps.

standardization method depend on the set of criteria and scales of measurement. each point can be represented as a vector of scores ranging from zero to one (figure 2); each score measures the level of contribution of a criterion to the selection or elimination of that point.

Criteria selection and scoring are the necessary components of the multicriteria line simplification system, but not sufficient in themselves; the other crucial component of the system is the AEM modeler's preference structure, which will be introduced in the system as a one-dimensional weight vector **W**, which has as many elements as there are criteria. These weights represent the (relative) importance of each criterion in the simplification process.

The next step in multicriteria line simplification is definition of an aggregation function that reduces the multicriteria vector of raw scores into a single composite score. Chrisman (1997) provides an excellent discourse on the logic of 'overlaying' and combining criteria spatially into one overall score of desirability. Herwijnen & Rietveld (1999) have formalized the process in the spatial context. We will choose the simplest and most popular method known as Simple Additive Weighing (SAW) (Eastman et al., 1993; Malczweski, 2000). The general idea behind the method is very simple – that of that of weighted averages. Formally SAW can be expressed in matrix form as:

$$\mathbf{X} * \mathbf{W} = \mathbf{A} \tag{1}$$

If $m$ is the number of alternatives and $n$ the number of criteria, **X** is the $m \times n$ matrix of raw criterion scores provided by spatial data analysis, **W** the $n \times 1$ weight matrix provided by the user, and **A** is the $m \times 1$ outcome matrix with final composite scores, for each of the $m$ points. Note that since each weight is expressed as a percentage, and the standardize scores vary from 0 - 1, each composite score will vary from 0 to 100 The composite score allows the ranking of points with respect to their amenability for selection, after simplification. It can also be interpreted as a measure of the probability (likelihood) of selection. Since each point will get a different score normally, this creates an uneven information selection field for a given line. The likelihood of a point being retained in the simplified version of the line varies with the particular constraints imposed, the criterion scores and weights, and the aggregation function.

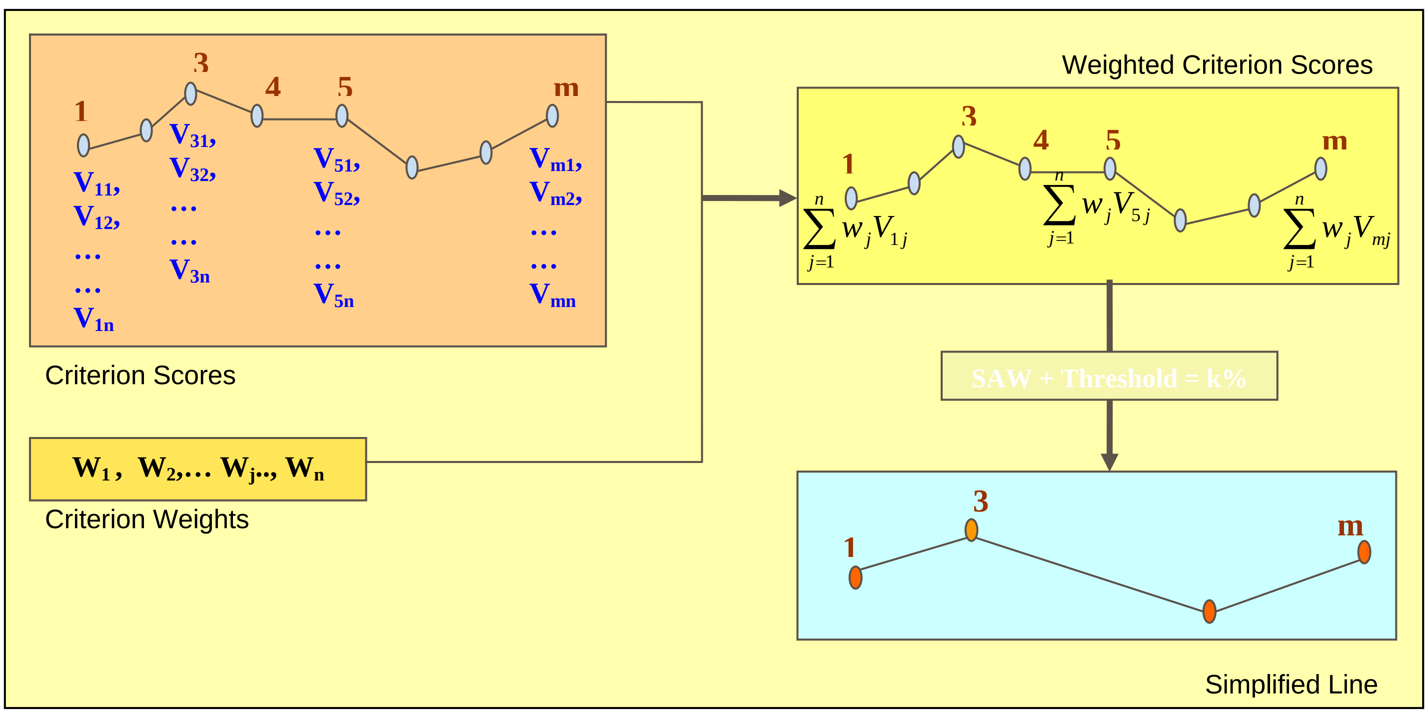



Figure 2. Multicriteria Line Simplification Schema

Following the reduction of multicriteria vectors to a scalar value (composite score) for each alternative, a selection rule is applied to select the best alternative. In our case we are not interested in selecting not only the 'best' point (one which has the highest composite index), but all those points as well that meet the criterion of selection. The selection rule we follow is based on the simple idea of setting a threshold percentage (0 – 100); invoking this threshold will allow one to select only those points whose composite score is greater than the threshold. Setting a low threshold to zero will select all points, and setting it to 100 will eliminate all points. However one has to be careful here because certain constraints need to be considered at this stage of elimination of points; for example, to preempt the elimination of the anchoring points (first and last nodes) for preserving topological continuity between line elements, we enforce the constraint that their composite score is always 100. This can also be done for any other point, whose presence is regarded as mandatory during initial spatial analysis of the groundwater model. The constrained threshold selection rule provides the modeler with another opportunity to enforce his preference after the criteria selection and weight vector specification.

## 7. Calibration of the Multicriteria Line Simplification System

The process described above is a heuristic and there is no theoretical way to prove its optimality. Only empirical results can indicate the optimal configuration of the multicriteria line simplification system. The simplification phase therefore should always be followed by an evaluation phase, during which model performance and results are calibrated. The multicriteria simplification system can be made to produce significantly different results by changing either the system state (scored criteria) or system parameters (criterion weights, aggregation function, thresholds). Generally, if a relatively stable list of criteria can be generated, calibration of the line simplification system will involve only changing the parameters. If the aggregation function is also invariable, then calibration involves repeating simplification of linear analytic elements for different combinations of weights and thresholds.

In an earlier work, McMaster (1986) suggested six different statistical measures to calibrate the performance of a line simplification algorithm. Recently Skopeliti & Tsoulos (2001) have also suggested similarity measures for comparing two lines. However, these are based only on the geometric differences between the original and simplified line. As explained earlier, these measures may not be relevant in our case because our ultimate objective is not simplification itself, but reducing simulation time and also minimizing prediction errors. Geometric measures may classify two simplified versions of the same original line as very similar, but their relative impact on the model may differ significantly depending on the set of points comprising the two simplified versions. Hence we suggest three AEM model dependent measures to calibrate the simplification process.

i) *Simulation time* = $T_i$ $\quad$ i = *1, 2,…,s* (2)

where $T_i$ is the time the model takes to converge to an acceptable solution at a fixed level of precision, and $s$ is the total number of simulations made, each for a different (simplified) model.

ii) *Mean Global RMS Error* = $(\sqrt{\sum_i (y_i - Y_i)^2})/k$ $\quad$ i = *1,2,…,k,* (3)

where $y_i$ is the predicted value for the chosen output parameter from a simplified model for an error node *i*, $Y_i$ is the predicted value of the same parameter as calculated from the original unsimplified model for the same error node *i*, and *k* is the total number of error nodes. An error node is any location in the geographic area where model predictions are valid; typically these are chosen well within the geographic boundary of the groundwater system to avoid spurious errors induced by artificial boundary conditions imposed for modeling purposes. This error measure indicates how well the simplification does over the whole geographic study area.

iii) *Mean Local RMS Error* = $(\sqrt{\sum_j (y_i^j - Y_i^j)^2})/k_j$ $\quad$ i = *1,2,…,*$k_j$ (4)

Local RMS error values are similar to global RMS errors, but are calculated separately for different delineated zones of interest, instead of for the whole study area. The suffix *j* in equation 4 identifies the zones; *y* and *Y* identify the predicted values from the simplified and unsimplified model and $k_i$ refers to the number of error nodes in the $j^{th}$ zone.

While the Mean Global RMS error is an easy way to compare results, it lacks spatial expression. Local RMS errors tend to capture some of that, especially if they are used for spatial statistical analysis later on. This will help identify areas of over or under prediction; additional constraints can then be added or criteria weight vectors modulated to further minimize errors, during the sensitivity analysis phase.

## 8. Case Study

### 8.1 MCLS System Description

In order to verify the above theory, we constructed an AEM groundwater model for the geographic area between Oil and Ischua creek in Cattaraugus County in Western New York, USA. 42 streams and 57 pumping wells were used as inputs to the model. Based on previous calibrations and studies (Hart, 2001) the recharge[†] rate for the area was set to be 0.03 inches/day; the background conductivity was input as 0.3 m/day. No inhomogeneities were input to the system for this case study. SPLIT (Jankovic, 2001) was used as the AEM engine.

Three criteria for multicriteria line simplification were chosen:

i) *Douglas-Peucker Amplitude (DPAmp)*: The amplitude essentially indicates the 'geometric potential' of a point for selection; higher distances mean that it is further away from the local trend of the line and hence its ability to be a 'geometric event' is higher. The raw criterion score is the amplitude (distance) of a point from the baseline in use before the point divides the segment containing it into two segments, during a simplification run of the well known Douglas-Peucker algorithm. The algorithm was first suggested by Douglas-Peucker (1973) and has been the most popular (and often most efficient) algorithm of line simplification since then (White, 1985; McMaster & Shea 1992; Visvalingam & Whyatt, 1991). Cromley (1991) used the same approach to construct a hierarchical tree based on Douglas-Peucker distances; here we use the distances to instead score the geometric criterion representing a geometric event.

ii) *Change in Topographic Slope (Slo)*: The change in topographic slope was calculated along the surface hydrographic features (streams); this criterion is designed to capture semantic events like waterfalls (i.e., sudden change in heads). Higher scores on this criterion for a point indicates that slope changes rapidly near the point. This criterion is designed to help maintain downslope flow of water by scoring points with large curvature higher.

iii) *Log Inverse Distance to Nearest Well (Log1/DW)*: This criterion measures for each point the log of the inverse distance to the nearest well. This was included in the system to model the constraint that stream segments near wells should be preserved to obtain accurate local heads near these anomalous regions. Higher scores indicate that the point is near to a well and hence should be preserved after simplification.

Seven different weighing schemes were designed such that each criterion got 100%, 50% and 25% of the weight. Based on the tuple structure *(DPAmp, Slo, Log1/DW)*, the schemes can be represented as 3 element weight vectors: *<33⅓,33⅓, 33⅓>, <100,0,0>, <0,100>, <0,0,100>,<50,25,25>,<25,50,25> & <50,25,25>*. So for example, the weight vector *(50, 25, 25)* represents a simplification scheme in which DPAmp receives 50% importance and Slo and Log1/DW receive 25% weight each during simplification. For each such scheme, the composite score threshold was varied to choose 25%, 50% and 75% of the total points (sum of all points on all streams) in the system. Thus 21 different simulations were conducted and the simulation time and Global RMS errors calculated for each. Table 1 and 2 present the results for the case study.

[†] Recharge (Infiltration ) is calculated as: Recharge = Precipitation – Surface Runoff. The runoff in this area is typically 70-80% of the precipitation leaving; average precipitation assumed is around 30-40 inches/year.

**8.2 Results**

i) *Simulation Times*: Table 1 below indicates that the simulation time averaged across weight vectors for drops by more than 50% for only a 25% reduction in time and by more than 75% with a 50% decrease in points. Simulation time is not necessarily a function of only the number of points, since average time for all thresholds varies with respect to the weight vectors. The best average time was obtained for *<25,25,50>* but others were almost as time efficient. However, when distance to wells (Log1/DW) (i.e., *<0,0,100>*) was the only criterion used to simplify lines, simulation time was considerably high. Analysis of all the simulation times table indicates that even if the number of line segments (or arity of the set of selected points) is held constant, the simulation time can change depending on which points were selected (as would happen if different schemes were used to select points).

| | | Time(minutes) | | | | |
|---|---|---|---|---|---|---|
| **Schema** | **Weight Vector** | **Unsimplified** | **25% Points** | **50% Points** | **75% Points** | **Average Time** |
| Equal Wts | <33⅓,33⅓, 33⅓> | | 28.350 | 88.733 | 275.850 | *130.978* |
| 100%DPAmp | <100,0,0> | | 26.400 | 140.800 | 226.950 | *131.383* |
| 100%Slope Change | <0,100,0> | | 29.250 | 138.667 | 222.817 | *130.244* |
| 100%Log1/DW | <0,0,100> | | 136.817 | 128.183 | 279.950 | *181.650* |
| Divided Wts | <50,25,25> | | 20.600 | 112.450 | 279.833 | *137.628* |
| Divided Wts | <25,50,25> | | 23.217 | 120.433 | 273.717 | *139.122* |
| Divided Wts | <25,25,50> | *510.683* | 22.917 | 115.933 | 225.767 | ***121.539*** |
| | | **Average Time** | *41.079* | *120.743* | *254.983* | *138.935* |

Table 1: Simulation Times

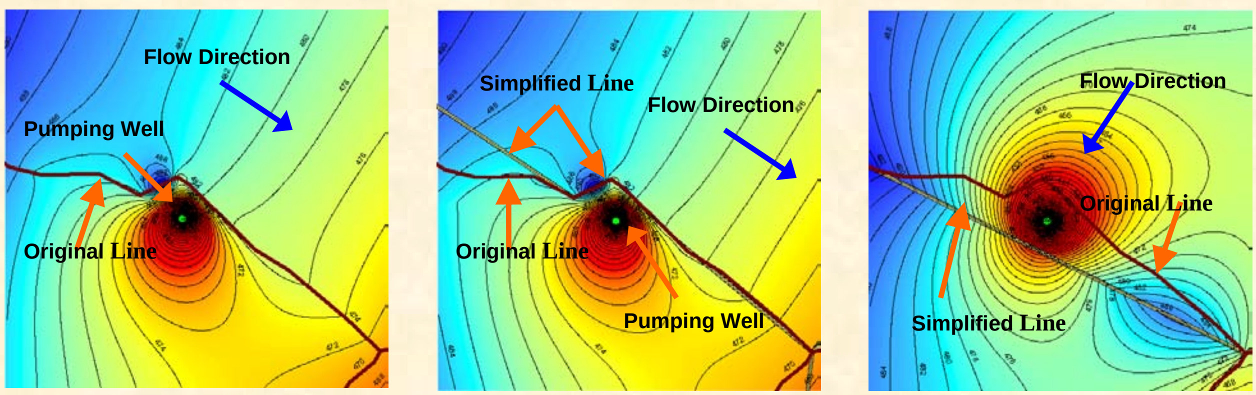
| | | Global RMS Error | | | | |
|---|---|---|---|---|---|---|
| **Schema** | **Weight Vector** | **Unsimplified** | **25% Points** | **50% Points** | **75% Points** | **Average Err** |
| Equal Wts | <33⅓,33⅓, 33⅓> | | *7.460* | *1.153* | *0.060* | *2.891* |
| 100%DPAmp | <100,0,0> | | *0.373* | *0.042* | *0.006* | ***0.140*** |
| 100%Slope Change | <0,100,0> | | *3.267* | *0.485* | *0.219* | *1.324* |
| 100%Log1/DW | <0,0,100> | | *18.983* | *2.989* | *1.171* | *7.714* |
| Divided Wts | <50,25,25> | | *5.574* | *0.120* | *0.219* | *1.971* |
| Divided Wts | <25,50,25> | | *9.713* | *2.664* | *0.063* | *4.147* |
| Divided Wts | <25,25,50> | | *4.945* | *0.279* | *0.095* | *1.773* |
| | | **Average Err** | *7.188* | *1.105* | *0.262* | *2.851* |

Table 2: Global RMS Errors

ii) *Global RMS Errors*: If we look at the Global RMS errors calculated for 450 error nodes (Table 2), groundwater heads are least accurate for all threshold values when Log1/DW is given 100% weight. Hence, it is obvious, while nearness to well is an important criterion for maintaining the position of wells with regard to the stream bank, using it as the only criterion in simplification of polylinear elements tends to be both computationally expensive and highly inaccurate. This is to be intuitively expected since using distance to well as the only criterion imposes no geometric constraints on the movement of the simplified line, i.e., in other words, the actual position of surface features has no contribution to groundwater heads. However, as any groundwater modeler can verify, heads are mostly influenced by surface water flow characteristics, especially in areas of high conductivity. The criterion

'Slo' alone (i.e., <0,100,0>) produces better results because indirectly it does contribute somewhat to the way a stream flows and captures some of the environmental variation generally found along stream reaches.

The most interesting conclusion that can be drawn from the Global RMS errors obtained in this case study is related to the *<100, 0, 0>* (i.e., DPAmp is effectively the only criterion) simplification scheme. Quite contrary to our original hypothesis that cartographic methods of line simplification will not suffice for environmental modeling, the Douglas-Peucker Amplitude, when used to simplify lines, produces the most accurate model results. It appears that on a global scale, geometric simplification, relying only on positional accuracy and vector displacement, has the strongest influence on groundwater heads. Other artificial constraints like well location and abnormal head drops appear to be local events that do not affect estimations of global heads throughout the domain.




3a: Unsimplified Model 3b: <33⅓,33⅓, 33⅓> Simplified 3c: <100,0,0> Simplified

Figure 3: As 3c shows the flow pattern is significantly changed near the well since the <100,0,0> (i.e., pure Douglas-Peucker simplification) shifts the stream segment erroneously south of the well. Flow pattern near the well for <33⅓,33⅓, 33⅓> weight vector, (3b) which gives equal importance to all criteria, remains similar to that observed for the unsimplified (3a) i.e., most accurate model.

| Schema | Weight Vector | Local RMS Err | Time (minutes) |
|---|---|---|---|
| Equal Wts | <33⅓,33⅓, 33⅓> | 0.476 | 36 |
| 100%DPAmp | <100,0,0> | **106.477** | **52** |
| 100%Log1/DW | <0,0,100> | 0.427 | 31 |

Table 3: Local RMS Err and $T_i$ around a well that changes relative position after simplification

iii) *Local RMS Errors*: On a full investigation of the model configuration for all simplification schemes, we realized that due to the scale of digitization of hydrographic data (≈100m), geometry was overwhelming the other criteria. No constraints (e.g. wells on wrong side of stream bank) appeared to have been violated. We therefore modified the model for further academic investigation. A pumping well was artificially introduced such that its position with respect to stream banks will change after simplification with the scheme (*<100,0,0>*). Two new simulations were conducted (unsimplified and simplified (*<100,0,0>*) model).

Figure 3 and Table 3 show Local RMS model errors for the *<100,0,0>* and *<33⅓,33⅓, 33⅓>* weight vectors at 25% threshold, as measured with respect to the new base results (unsimplified model run) around the new well. Figure 3 clearly documents the local change in flow potential around the well for the unsimplified model and the two simplified ones. Flow direction (which is orthogonal to the equipotential contours) changes abnormally for a purely geometric simplification scheme (*<100,0,0>)*. Weighing criteria equally, in this case, prevents the stream segment from crossing over to the other side of the well (figure 3b). Table 3 below indicates that local errors rise phenomenally if incorrect criteria are used to simplify segments but are moderated if appropriate constraints are in place. Simulation time increases too indicating that the model takes relatively longer to converge in case of this new geometric setting.

## 9. Conclusions and Future Work

The stress in this paper has been on the emplacement of semantic constraints while processing raw geospatial data for environmental modeling. The Multicriteria Line Simplification (MCLS) system was introduced to explicitly model relevant constraints on simplification of linear hydrographic features for AEM groundwater modeling. The method is however generic and can be used for any vector based environmental model. It was hypothesized that traditional line simplification heuristics are not efficient since they focus on the geometric caricature of the line and neglect semantic constraints. However, AEM modeling depends heavily on surface feature geometry and hence the Douglas-Peucker Amplitude criterion is a good method of polylinear element simplification when major constraints will not be violated globally. It was also shown that local errors are magnified significantly where constraints having local impacts are violated. In this paper we chose only three criteria and used a global weighing scheme. Polylinear elements in constrained regions should be simplified with different weighing schemes than used globally. This dual approach of weighing should help improve results.

More diverse constraints and spatially varying weight vectors will be needed for more complicated analyses. Other criterion aggregation functions besides Simple Additive Weighing (SAW) should also be considered (see Malczweski, 2000). One should also be careful of the method used for scoring criteria since results are heavily contingent on measurement scales and standardization methods. Future work will consider extending the 0-dimensional point based simplification scheme with l-dimensional line based multiconstraint simplification scheme. Most importantly, since the geometric criterion outperformed others, other geometric criteria besides DPAmp should be explored. If future analyses indicate overwhelming influence of geometric or any other criterion, this simplification scheme will constrain modelers by prescribing minimum relative weights that should be given to such criteria.

## Acknowledgments

This research has been supported by Grant # R82-7961 from the U.S. Environmental Protection Agency's Science to Achieve Results (STAR) program. This paper has not been subjected to any EPA review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred. James Craig was supported by the National Science Foundation Integrated Graduate Education and Research Training (IGERT) program in Geographic Information Science.

## References

Beard M K (1990). Constraints on Rule Formation. In B P Butenfield, R B McMaster (eds.) *Map Generalization: Making Rules for Knowledge Representation*. Longman Group UK Limited 1991, pp 121-135.

Beard M K (1991). Theory of the Cartographic Line Revisited: Implications for Automated Generalization. *Cartographica*, Vol. 28(4), pp. 32-58.

Brassel K E, Weibel R (1988). A Review and Conceptual Framework of Automated Map Generalization. *International Journal of Geographical Information Systems*, Vol.2(3), pp. 229-244.

Buttenfield B P (1985). Treatment of the Cartographic Line. *Cartographica*, Vol. 22(2), pp. 1-26.

Buttenfield B P, McMaster R B (eds.) (1991). *Map Generalization: Making Rules for Knowledge Representation*. Longman Group UK Limited.

Chrisman G M (1997). *Exploring Geographic Information Systems*. John Wiley & Sons.

Cromley R G, Morse S J (1988). Some Comments on R.B. McMaster's 'Geometrical properties of numerical generalization'. *Geographical Analysis*, Vol. 20(4), pp. 336-340.

Cromley R G, Campbell G M (1991). Noninferior Bandwidth Line Simplification: Algorithm and Structural Analysis. *Geographical Analysis*, Vol. 23, pp. 25-38.

Cromley R G (1991). Hierarchical Methods of Line Simplification. Cartography and Geographic Information Systems, Vol. 18(2), pp. 125-131.

Douglas D H, Peucker T K (1973). Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Character. *The Canadian Cartographer*, Vol. 10(2), pp. 112-123.

Eastman J R, Kyem P A K, Toledano J, Jin W (1993). *GIS and Decision Making*. Geneva: The United Nations Institute for Training and Research (UNITAR).

Haitjema H M (1995). *Analytic Element Modeling of Groundwater Flow*. Academic Press, San Diego, California.

Hart E (2001). Characterization of Aquifer Vulnerability Using DRASTIC and GIS for the Olean Creek Watershed, Cattaraugus County, New York. *Master's Thesis*. Department of Geology, University at Buffalo, NY, USA.

Herwijnen M V, Rietveld P (1999). Spatial Dimensions in Multicriteria Analysis. In J C Thill (ed.) *Spatial Multicriteria Decision Making and Analysis: A Geographic Information Sciences Approach*. Ashgate Publishing Company, 11-48.

Jankovic I. (2003) Split. Computer program for analytic-based modeling of single-layer groundwater flow in heterogeneous aquifers with particle tracking, capture-zone delineation and parameter estimation, available from www.groundwater.buffalo.edu

Jankowski P (1995). Integrating Geographical Information Systems and Multiple Criteria Decision-Making Methods. *International Journal of Geographic Information Systems*, Vol. 9 (3), pp. 251-273.

Lang T (1969). Rules for Robot Draughtsmen. *The Geographical Magazine*, Vol. 42(1), pp. 50-51.

Malczweski J (2000). *GIS and Multicriteria Decision Analysis*. John Wiley and Sons, Inc.

McMaster R B (1986). A Statistical Analysis of Mathematical Measures for Linear Simplification. *The American Cartographer*, Vol. 13 (2), pp. 101-121.

McMaster R B (1987). Automated Line Generalization. *Cartographica*, Vol. 24, pp. 74-111.

McMaster R B, Shea K S (1992). *Generalization in Digital Cartography*. Washington DC: Association of America Geographers.

Plazanet C (1995). Measurement, Characterization and Classification for Automated Line Feature Generalization. *Proceedings Auto Carto 12* (*ACSM/ASPRS Annual Convention and Exposition 4*), pp. 59-68.

Skopeliti A, Tsoulos L (1999). On the Parametric Description of the Shape of the Cartographic Line. *Cartographica*, Vol. 36 (3), pp. 57-69.

Skopeliti A, Tsoulos L (2001). A Knowledge Based Approach for the Cartographic Generalization of Linear Features. 20th International Cartographic Conference, August 6-10, 2001, Beijing, China.

Sinha G, Flewelling D M (2002) A Framework for Multicriteria Line Generalization to Support Scientific and Engineering Modeling, *Second International Conference of Geographic Information Science (GIScience 2002)*, Boulder, Colorado.

Strack O D L (1989). *Groundwater Mechanics*. Prentice Hall.

Visvalingam M, Whyatt J D (1990). The Douglas-Peucker Algorithm for Line Simplification: Re-evaluation through Visualization. *Computer Graphics Forum 9*, pp. 213–228.

Voogd H (1983). *Multicriteria Evaluation for Urban and Regional Planning*. London, Pion.

Wang Z, Müller J C (1998). Line Generalization Based on Analysis of Shape Characteristics. *Cartography and Geographic Information Systems*, Vol. 25 (1), pp. 3-15.

Wilson D R, Martinez T R (1997). Improved Heterogeneous Distance Functions. Journal of Artificial Intelligence Research, Vol 6, pp. 1-34.

Weibel R, Dutton G (1998). Constraint-Based Automated Map Generalization. *8th International Symposium on Spatial Data Handling* 1998, pp. 214-224.

White E R (1985). Assessment of Line Generaliation Algorithms using Characteristic Points. The American Cartographer, Vol. 12 (1), pp. 17-28.

Yoon K P, Hwang C L (1995). *Multiattribute Decision Making: An Introduction*. Sage Publications, London.

Zeleny M (1982). *Multiple Criteria Decision Making*. McGraw Hill, New York.