Evaluating Hydrology Preservation of Simplified Terrain Representations

Ph. D. Student: Christopher Stuetzle (CS Dept., RPI)
Ph. D. Advisors: W. Randolph Franklin (ECSE Dept., RPI)
and Barbara Cutler (CS Dept., RPI)

Collaborators: Jonathan Muckell (ECSE Dept., RPI), Marcus Andrade (DPI - Univ. Fed. Vicosa - Brazil), Jared Stookey (ECSE Dept., RPI), Metin Inanc (CS Dept., RPI), Zhongyi Xie (CS Dept., RPI)

ABSTRACT

We present an error metric based on the potential energy of water flow to evaluate the quality of lossy terrain simplification algorithms. Typically, terrain compression algorithms seek to minimize RMS (root mean square) and maximum error. These metrics fail to capture whether a reconstructed terrain preserves the drainage network. A quantitative measurement of how accurately a drainage network captures the hydrology is important for determining the effectiveness of a terrain simplification technique. Having a measurement for testing and comparing different models has the potential to be widely used in numerous applications (flood prevention, erosion measurement, pollutant propagation, etc). In this paper, we transfer the drainage network computed on reconstructed geometry onto the original uncompressed terrain and use our error metric to measure the level of error created by the simplification. We also present a novel terrain simplification algorithm based on the compression of hydrology features. This method and other terrain compression schemes are then compared using our new metric.

Categories and Subject Descriptors

Computing Methodologies [**Computer Graphics**]: Computational Geometry and Object Modeling

1. INTRODUCTION

Terrain data is being sampled at ever increasing resolutions over larger geographic areas requiring special compression techniques to manipulate the data. Typically the effectiveness of a terrain compression technique is how well it minimizes the root mean square or the maximum error between the original terrain and the reconstructed geometry [7]. This metric is not always the best choice for preserving hydrological information, since channels and ridges, essential for the calculation of drainage networks [13], might be lost. For example, a scheme which naively interpolates the terrain between two points on opposite banks of a river can flatten the terrain and block flow.

Direct ground truth measurements can be used to determine the amount of water and various hydrology statistics. This can be expensive, time consuming, and require accessing remote locations. Rapid technological advances are making it possible to acquire accurate, high-resolution elevation data, allowing more accurate computer simulation of hydrology. It is essential that the scientific community have the tools available that can efficiently store and manipulate large terrain datasets [1]. Accurate hydrological simulations allow better understanding of regions at greatest risk of flooding, preparation for the threat of natural disasters, and tracking and predicting the flow of pollutants. This work could also be applied to segmentation of terrain based on watersheds or other flow based models, such as volcanic flow.

ACM GIS '08 November 5-7, 2008. Irvine, CA, USA
©2008 ACM ISBN 978-1-60558-323-5/08/11 ...$5.00.

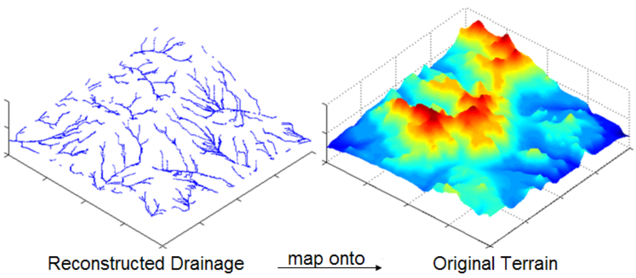


Figure 1: To compute the potential energy error, the drainage is computed on the reconstructed terrain and then mapped onto the original terrain. The amount of water flowing uphill and downhill influences the metric. The highest elevations are visualized in dark red and the lowest elevations are dark blue.

2. PRIOR ART

2.1 Digital Hydrology Methods

Various methods and metrics have been defined for computing and comparing digital drainage networks to ground truth, real world drainage [14].

One such method, the D8 model, assigns flow in one of the eight possible directions. In the SFD (single flow direction) version of the D8 model the entire amount of flow from each cell is distributed to the lowest adjacent neighbor. This is not the case in the MFD (multi-flow direction) version in which the flow is fractionally distributed to all the lower adjacent neighbors. A slightly more sophisticated MFD model is the D∞ model. As the name indicates, flow can travel in an infinite number of directions and is not limited to eight cardinal and diagonal directions. The amount of water leaving each cell is distributed to one or more adjacent cells based on the steepest downward gradient [11].

Another method for hydrology calculation is the digital elevation model network or DEMON model [3]. Rather than modeling flow as a point source that flows to an adjacent neighbor, DEMON captures the flow by contributing and dispersal areas. The motivation for using a method such as DEMON is that the representation allows for flow width to vary over non-planar topography. However, this can introduce loops and inconsistencies in the hydrology.

Elevation data is only an approximation for the actual terrain and is prone to collection and sampling errors that cause unrealistic depressions. To counter this, some methods have been extended to allow water to flow uphill out of local minima (basins) until spilling over an edge. The flow network thus runs uphill in situations when there is not an adjacent lower elevation. These methods expand the drainage networks until they flow off the edge of the terrain. In Terraflow [4, 12], the path of least energy is used to flow uphill until reaching the spill point. The main benefits of Terraflow are the ability to avoid dataset issues, construction of long continuous river flow, and scalability on massive datasets. The disadvantages are that this approach may miss realistic drainage basins and have poorer performance on non-massive datasets than simpler methods.

For the methods listed above, the inputs are a DEM (Digital Elevation Model) and a flow accumulation threshold. The outputs are a flow direction grid and a flow accumulation grid. The flow direction grid specifies the direction of flow and the flow accumulation grid records the amount of flow. A cell is considered part of the drainage network if its flow accumulation is larger then the specified threshold.

2.2 Approximating Terrain using Over-determined Laplacian PDEs

To reconstruct a dense terrain matrix from a subset of the original elevation data, we use the Over-determined Laplacian Differential Equations (ODETLAP) method [7]. ODETLAP can process not only continuous contour lines but isolated, irregularly-spaced points as well. The surface produced tends to be smooth while preserving high accuracy to the known points. Local maxima are also well preserved. Alternate methods generally sub-sample contours due to limited processing capacity, or ignore isolated points.

Starting with the Laplacian for every non-border point:

$$4z_{ij} = z_{i-1,j} + z_{i+1,j} + z_{i,j-1} + z_{i,j+1} \qquad (1)$$

we add a second equation for each known point:

$$z_{ij} = h_{ij} \qquad (2)$$

where h_{ij} stands for the specified elevation and z_{ij} is the computed elevation for the point. Thus, the system of linear equations is over-determined, i.e., the number of equations exceeds the number of unknown variables, so instead of solving it for an exact solution, an approximated solution is obtained. The user defines a parameter R that determines the relative importance of accuracy versus smoothness.

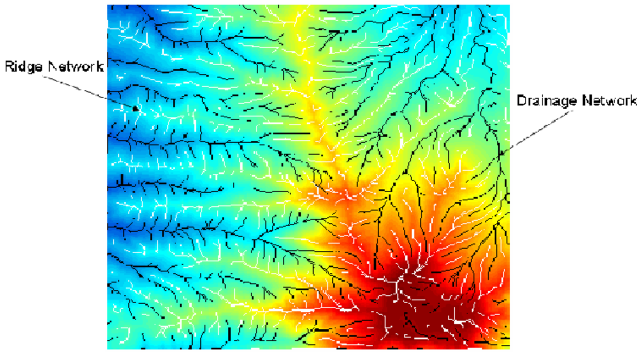


Figure 2: The ridge-river network, with rivers in black and ridges in white.

We have explored the use of Triangulated Irregular Networks, Visibility, and Level Set Components to discover important points that reflect the terrain structure for use in ODETLAP [15].

3. OVERVIEW

Our goal is to preserve not only the overall terrain structure, but also important hydrology features. Our research contributions include:

1. A new metric for measuring the amount of hydrology error introduced by a terrain simplification algorithm. The metric is based on the amount of water that (incorrectly) flows uphill.

2. Introduction of a new geometry terrain feature we call the *ridge network*. This network is used in our hydrology compression scheme and also has applications in observer siting and path planning.

3. Efficient computation of both the drainage network and ridge network using a system of linear equations.

4. Introduction of a new compression method that is *hydrology-aware*. By specifically targeting the compression we can minimize the amount of drainage network error on the reconstructed terrain.

Our experiments have shown that points on the ridge network and drainage network are effective in capturing the hydrology. The ridge-river technique computes both the rivers and ridges, and simplifies the line network to capture the most significant points.

4. OUR CONTRIBUTIONS

4.1 Ridge-River Network Calculation

We compute the drainage network using a standard D8 model [11] based on steepest descent flow. Each cell flows to the lowest adjacent neighbor and flow is forbidden from traveling uphill. We also introduce and compute the *ridge network* in a similar fashion from the inverted terrain, I_e, which is quite simply computed from the original elevation matrix, E, by negating all elevations.

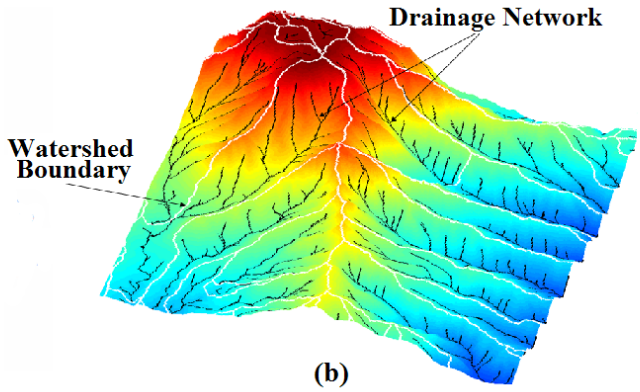


Figure 3: An example of our computed drainage network and the corresponding watersheds.

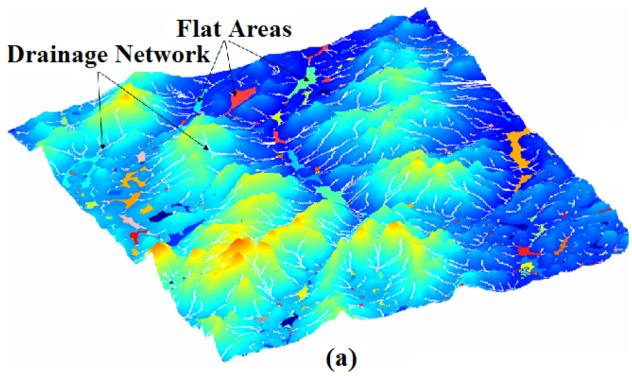


Figure 4: Visualization of flat plateau regions.

E is used to compute the drainage network and I_e to compute the ridge network, which can be done in parallel. We refer to the combination of networks as the *ridge-river network* [10], as seen in Figure 2. To the best of our knowledge, no previous work has performed drainage network computation on the inverted terrain. While the resulting ridge network has no direct physical interpretation, it does share features with the hydrology watersheds (Figure 3). We have found the ridge network to be useful for terrain compression, observer siting, and path planning.

Unlike other methods that use flooding [1], our method computes flow using a system of linear equations $Ax = b$ where x is an unknown N^2 length vector equal to the amount of water accumulation at each cell and b is the initial flow or "rain" at each cell, often with all entries equal to 1. Matrix A is a $N^2 \times N^2$ sparse matrix: the identity matrix with additional non-zero entries to represent flow between neighboring cells. For instance, if cell X_1 receives flow from cell X_2 and X_5, row 1 in matrix A will contain non-zero elements in columns 1, 2, and 5. Therefore the number of non-zero entries in matrix A is bounded by $2N^2$, where N is the size of the $N \times N$ DEM. The upper bound of $2N^2$ is determined since there will be N^2 non-zero entries to load the identity matrix. All other non-zero entries represent flow from one cell to one other cell. There can be at most N^2 additional non-zero elements, since each cell can flow in only one direction. Taking advantage of the sparse nature of matrix A, the linear system can be solved efficiently.

An important problem that needs to be addressed is the occurrence of plateaus, which are regions where the flow direction can not be trivially determined based on steepest descent (Figure 4). To deal with these cases, the plateaus are first identified using a variant of the fast Union-Find algorithm developed by Franklin and Landis [8]. The input is a $3N-2$ by $3N-2$ binary matrix and the output contains a list of components, with each component representing one plateau. Once identified, the flow directions for flat areas are set using a similar strategy to Terraflow [12]. A breadth-first search assigns directions towards the root or spill point. Spill points are identified as cells in a flat component that contain a nonzero direction. In other words, a cell in the component that has an adjacent cell with a smaller elevation. Flat areas that have no spill points are determined to be sinks. The directions of every cell in a sink are assigned to flow to this point.

After assigning directions to every plateau and sink, the final flow network can be computed. The linear system of equations is modified to include the directions assigned to the plateaus and sinks. The flow is recomputed and the final flow accumulation grid and flow direction matrix is determined. Figure 3 and 4 show examples of the drainage network, ridge network, and watershed boundaries.

The benefits of our flow calculation method include simplicity, scalability, and consistency (there is never a flow loop). However, like other digital hydrology simulation methods, we cannot guarantee robust construction of the actual hydrology network due to sampling and dataset inaccuracies that often unrealistically block flow.

4.2 Drainage Network Error Metric

Standard metrics for evaluating the effectiveness of terrain simplification algorithms use root mean squared (RMS) and maximum error. These measurements are ineffective for evaluating the loss of drainage network structure. Therefore, one of the main purposes of our research is to introduce a metric geared towards measuring this error.

It is important to note that the goal of our hydrology metric is not to compare the reconstructed hydrology against an absolute truth. As mentioned above, hydrology computed on a digital representation may have significant errors due to sampling and data collection inaccuracies. Therefore, our hydrology metric does not compare the reconstructed drainage network to the true drainage network. Rather, our metric takes the flow direction grid and the flow accumulation grid computed on the reconstructed terrain and maps it onto the original, uncompressed DEM (Figure 1).

To compute the accuracy of the reconstructed drainage network, the gradient, the amount of flow contributing to each cell, and whether the flow travels uphill or downhill on the original data are taken into account. The total downhill and uphill energies are computed as a summation of the gradient, $|E_i - E_{r(i)}|$, where E is the elevation matrix and E_i is the elevation of the i^{th} cell. $r(x)$ specifies the receiving cell for the flow out of cell i on the reconstructed terrain. Thus, $E_{r(i)}$ is the elevation of the cell that is coupled with cell i through flow. The gradient is weighted by the amount of flow, variable W_i, through the cell. Variable $EnergyDown$ is the sum of cells in the matrix where the reconstructed flow network (correctly) travels downhill on the original terrain. Conversely, $EnergyUp$ is the summation of cells where the

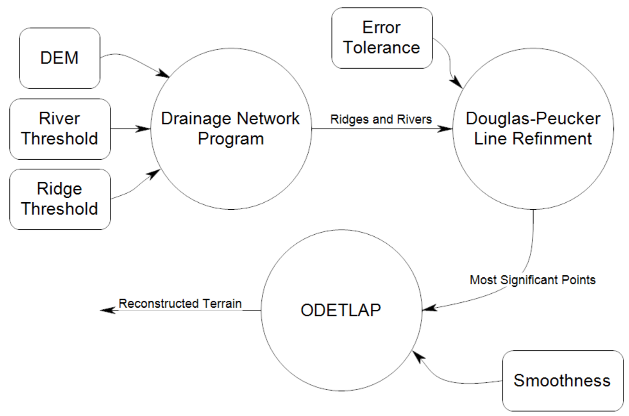


Figure 5: Flow chart of the ridge-river compression method. Inputs are in boxes and programs in circles.

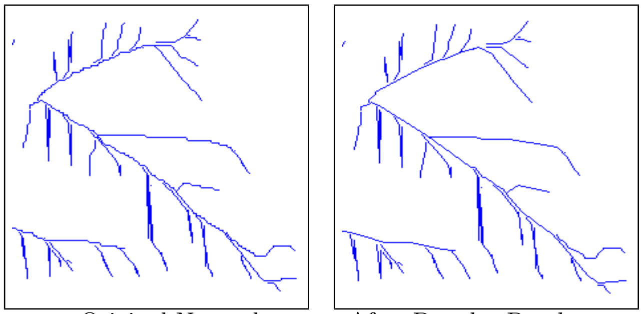


Figure 6: Simplifying the original drainage network using Douglas-Peucker. The refined line network is reduced by a factor of 3 with little visible difference.

flow travels (incorrectly) uphill. The final *Error* is determined as the ratio of the total upward energy divided by the total downward energy.

$$EnergyDown = \sum_i max(0, E_i - E_{r(i)}) * W_i$$

$$EnergyUp = \sum_i max(0, E_{r(i)} - E_i) * W_i$$

$$Error = \frac{EnergyUp}{EnergyDown}$$

To compute the energy error metric, the flow is computed on the reconstructed DEM (described in §4.1). The error is determined by comparing the flow direction matrix computed on the reconstructed geometry with the elevation matrix from the original DEM. A perfect reconstruction has zero uphill flow and a metric value equal to zero. Therefore, the closer the metric is to zero, the more accurate the reconstructed drainage network.

4.3 Network Simplification for Hydrology-Aware Compression

The output of the drainage computation is a flow accumulation grid, where each cell contains an integer corresponding to how many other cells contribute flow to that point. Cells above a predefined threshold are considered significant and are added to the river (or ridge) network. We note that this initial representation (a dense set of cells) is somewhat redundant and can be simplified before storage in our novel compressed format.

The drainage and ridge networks are simplified using the Douglas-Peucker[5] line refinement algorithm. This algorithm selects the most significant points needed to reconstruct a line within a given error tolerance. This tolerance specifies the maximum distance the line can deviate from the original. The higher the tolerance, the fewer points required and the greater the difference between the original network and the reconstructed network. The output from the Douglas-Peucker algorithm is an ordered list of the most significant points needed to reconstruct the line. These points form the basis of our compressed terrain representation. As Figure 6 illustrates, when the tolerance is set appropriately there is a significant reduction in number of control points with negligible visual difference in the river network.

The simplified network segments are then efficiently written to a file using delta encoding to achieve the compressed format. Figure 5 presents a flow chart describing the ridge-river terrain simplification technique for compressing and reconstructing the significant hydrology structure of a terrain.

4.4 Hydro-ODETLAP for Terrain Reconstruction

To reconstruct the terrain from the sparse set of points on the ridge and river networks, we use ODETLAP (§2.2). To more accurately capture the structure of the hydrology, the ODETLAP equations are modified for points selected on the ridge-river network. Because river points are known to be relatively lower that their neighbors we modify the Laplacian equation (Eqn. 1) for these points as follows:

$$4z_{ij} = z_{i-1,j} + z_{i+1,j} + z_{i,j-1} + z_{i,j+1} - D_R \qquad (3)$$

where D_r stands for decrement for the rivers. This variable is an integer corresponding the number of meters the rivers lie below the average of the 4 neighbors. Similarly, ridge network points are higher then the average of their four neighbors, thus for ridge network points, the equation becomes:

$$4z_{ij} = z_{i-1,j} + z_{i+1,j} + z_{i,j-1} + z_{i,j+1} + I_R \qquad (4)$$

where I_R is an integer corresponding to the increment for the ridges. We found that setting $D_R = I_R = 2$ has been effective. In future work we plan to study how varying this parameter affects the results and investigate ways to automatically select an optimal value and/or vary this value as appropriate throughout a terrain. This modification to the original ODETLAP equations yields an impressive reduction in the error, as shown in Figure 7.

5. RESULTS

Our primary focus has been to develop and present a metric that accurately captures the amount of error introduced into a reconstructed drainage network. Guided by this metric, we created an algorithm for achieving high compression ratios without significantly altering the hydrology. Results are shown in Table 1.

We compare our new compression technique to a Triangulated Irregular Network (TIN) [6] and JPEG2000 [9] image compression on a sample of six datasets we have standardized for our testing. JPEG2000 obtains a low percentage of

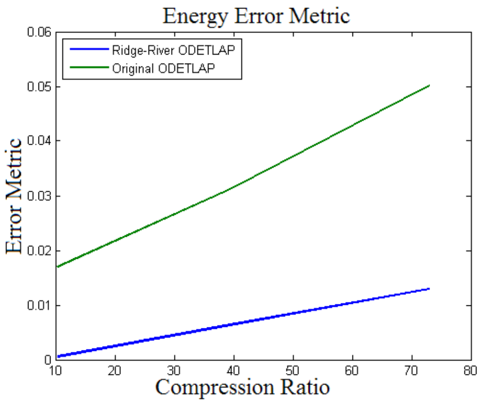


Figure 7: Modifying the ODETLAP equations to better represent ridges and rivers has a drastic decrease in the amount of hydrology error. Both plotted lines above use the same set of points.

cells that flow uphill, which correlates to a fairly low hydrology error. The ridge-river technique is effective in achieving high compression ratios with a fairly low error, however, it currently does not consistently beat JPEG2000. We are confident that small modifications to the current ridge-river method will allow us to achieve a significantly better hydrology error.

We are investigating further modifications to the ODETLAP equations, and to automatically select optimal parameters. For example, we will fill in the river network between the simplified river and ridge points using the Bresenham line rasterization algorithm [2].

Visual inspection of the reconstructed drainage networks correspond to the measurement errors determined by our new metric (Figure 8).

6. DISCUSSION

The modular design of our terrain simplification approach facilitates substitution of different algorithms in place of the ones focused on in this paper. For instance, Terraflow or ArcGIS could be used to compute the ridge-river network. Also, a different line simplification technique could be used instead of Douglas-Peucker. This allows modification to fit the specific objectives of the user and application.

Points on the ridges and rivers of the terrain are important for preserving the hydrology. Rather than use an existing algorithm we discovered that inverting the terrain and running the drainage network provides a quick, effective method for approximating the ridge network. This approach can be done with any drainage network program. The ridges are important in terrain compression for extracting and exploiting terrain structure, but also have other GIS applications such as visibility, siting, hydrology, and edge detection.

7. CONCLUSION AND FUTURE WORK

The potential energy metric introduced in this paper provides a quantitative measurement of the amount of error introduced into hydrology by a terrain compression technique. This value corresponds to a visual examination of the drainage networks, with higher error corresponding to fragmented and unrealistic flow directions (flow traveling uphill).

	Compr.	Ridge-River		JPEG2000		TIN	
	Ratio	%up	error	%up	error	%up	error
hill1	13	2.05	0.0023	0.12	0.0020	0.79	0.0432
	32	3.16	0.1149	0.18	0.0030	1.11	0.0502
	54	2.46	0.2316	0.24	0.0082	1.33	0.0600
hill2	14	0.85	0.0005	0.21	0.0010	1.25	0.0333
	37	1.21	0.0063	0.31	0.0017	1.80	0.0304
	60	1.39	0.0129	0.46	0.0047	2.43	0.0421
hill3	11	2.65	0.0026	0.10	0.0059	0.76	0.0311
	27	4.33	0.0075	0.11	0.0051	0.77	0.0434
	47	2.70	0.0100	0.13	0.0161	0.85	0.0405
mtn1	16	3.75	0.0267	0.41	0.0026	3.96	0.0563
	39	4.96	0.0530	0.80	0.0036	5.11	0.0583
	60	5.91	0.0611	1.33	0.0067	6.28	0.0667
mtn2	16	3.93	0.0769	0.40	0.0033	4.42	0.0748
	38	5.15	0.1169	0.75	0.0033	5.72	0.0874
	59	6.21	0.1377	1.32	0.0067	7.09	0.0904
mtn3	15	3.10	0.0254	0.40	0.0015	4.16	0.0592
	39	4.33	0.0493	0.78	0.0027	5.63	0.0624
	61	5.13	0.0639	1.40	0.0050	6.63	0.0650

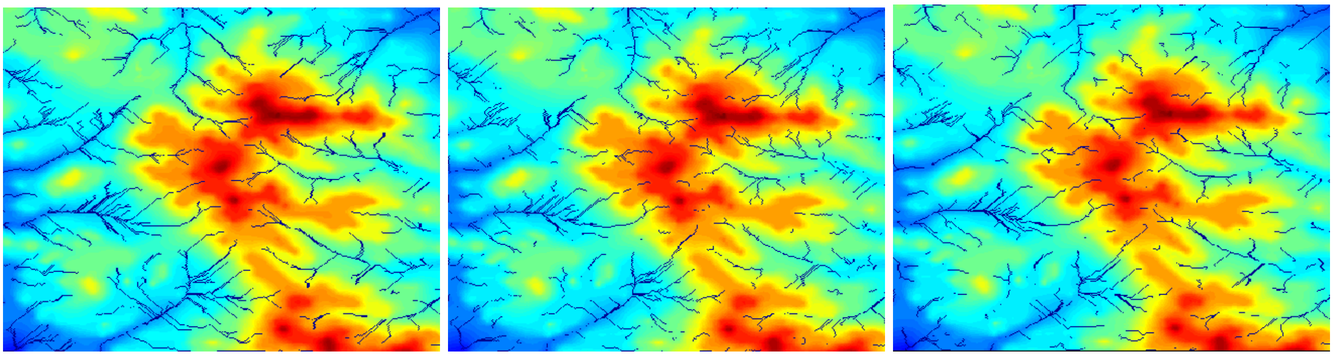
Table 1: In addition to the Oahu dataset, we use three hilly and three mountainous 400 by 400 datasets sampled at 30m resolution. Each dataset is compressed by 3 different lossy compression schemes at 3 different levels. For each, the percent of flow uphill and the energy error metric is presented.

The original DEM is an approximation of the real world terrain surface and not necessarily hydrologically-accurate, due to dataset and sampling errors. Flow can travel in different directions than the original drainage network, yet contain low error if the flow directions are reasonable. Standard terrain compression evaluation metrics such as root mean squared error and maximum error are ineffective in evaluating the amount of error introduced during lossy compression, as they do not take into account important hydrology features.

With terrain being sampled at ever increasing resolutions, it becomes more important to store and manipulate large elevation datasets efficiently, and evaluate the error introduced by lossy compression. Current techniques for compressing these datasets may lose important information, essential for applications such as hydrology. Understanding how compression affects important domain-specific 3D terrain structures will allow the GIS community to effectively evaluate the accuracy of different compression strategies.

There are several possible extensions for this work, including generalizing the metric to include the speed the water travels at and the area over which the water is spread. Deciding which metric is more useful could be left to the application, and the terrain would be compressed according to the chosen metric. Then, the metrics could be compared in a variety of situations to determine which applications each is best suited for.

Additionally, there are many extensions for the hydrology compression technique. We are currently investigating more modifications to the ODETLAP equations to further take hydology error into account. Additionally, in its current state the compression only saves the beginning and ending point of each river. For more accuracy, the reconstruction could interpolate the points along the river as known points



Original, Error=0.0 | Comp, Ratio=14, Error=0.0005 | Comp. Ratio=60, Error=0.0129

Figure 8: The images show the a 400×400 hill2 dataset sampled at 30m resolution and compressed using the ridge-river technique. The color regions represent the elevations with blue being low and red corresponding to high elevation. The black regions shows the significant drainage network above the threshold of 100. The higher potential energy error metric correlates with a visible difference in the drainage network. Notice how the high error corresponds to short fragmented drainage networks.

for ODETLAP during reconstruction.

8. ACKNOWLEDGEMENTS

This research was supported by NSF grants CCR-0306502 and DMS-0327634,a by DARPA/DSO/GeoStar, and by CNPq - the Brazilian Council of Technological and Scientific Development.

9. REFERENCES

[1] L. Arge, J. S. Chase, P. N. Halpin, L. Toma, J. S. Vitter, D. Urban, and R. Wickremesinghe. Efficient flow computation on massive grid terrain datasets. *GeoInform.*, 7(4):283–313, 2003.

[2] J. Bresenham. A linear algorithm for incremental digital display of circular arcs. *Commun. ACM*, 20(2):100–106, 1977.

[3] M. C. Costa-Cabral and S. J. Burges. Digital elevation model networks (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas. *Water Resources Research*, 30:1681–1692, 1994.

[4] A. Danner, T. Molhave, K. Yi, P. K. Agarwal, L. Arge, and H. Mitasova. Terrastream: from elevation data to watershed hierarchies. In *GIS '07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, pages 1–8, New York, NY, USA, 2007. ACM.

[5] D. Douglas and T. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *The Canadian Cartographer.*, 2(10):112–122, 1973.

[6] W. R. Franklin. Triangulated irregular network program, 1973. `ftp://ftp.cs.rpi.edu/pub/franklin/tin73.tar.gz` (accessed 23 May 2006).

[7] W. R. Franklin, M. Inanc, Z. Xie, D. M. Tracy, B. Cutler, M. A. Andrade, and F. Luk. Smugglers and border guards - the GeoStar project at RPI. In *15th ACM International Symposium on Advances in Geographic Information Systems ACM GIS 2007*, Seattle, WA, USA, Nov. 2007.

[8] W. R. Franklin and E. Landis. Connected components on 1000x1000x1000 datasets. In *16th Fall Workshop in Computational Geometry*, Smith College, Northampton, MA, 10-11 Nov 2006. (abstracts only).

[9] JPEG Committee. JPEG 2000. `http://www.jpeg.org` (accessed 14 March 2008).

[10] J. Muckell, M. Andrade, W. R. Franklin, B. Cutler, M. Inanc, Z. Xie, and D. M. Tracy. Drainage network and watershed reconstruction on simplified terrain. In *17th Fall Workshop in Computational Geometry*, IBM T.J. Watson Research Center, Hawthorne, New York, 9-10 Nov 2007. (abstracts only).

[11] D. G. Tarboton. A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research*, 2(33):309–320, 1997.

[12] L. Toma, R. Wickremesinghe, L. Arge, J. S. Chase, J. S. Vitter, P. N. Halpin, and D. Urban. Flow computation on massive grids. In *ACM-GIS*, pages 82–87, 2001.

[13] J. V. Vogt, R. Colombo, and F. Bertolo. Deriving drainage networks and catchment boundaries: a new methodology combining digital elevation data and environmental characteristics. *Geomorph.*, 53(3-4):281–298, 2003.

[14] J. P. Walker and G. R. Willgoose. On the effect of digital elevation model accuracy on hydrology and geomorphology. *Water Resources Research*, 35:2259–2268, 1999.

[15] Z. Xie, W. R. Franklin, B. Cutler, M. A. Andrade, M. Inanc, and D. M. Tracy. Surface compression using over-determined Laplacian approximation. In *Proceedings of SPIE Vol. 6697 Advanced Signal Processing Algorithms, Architectures, and Implementations XVII*, San Diego CA, 27 August 2007. International Society for Optical Engineering. paper 6697-15.