Function modeling: improved raster analysis through delayed reading and function raster datasets

Hogland, John S.[1], Anderson, Nathaniel M.[1], Jones, J. Greg[1]

[1] US Forest Service, Rocky Mountain Research Station, Missoula, MT

Abstract

Raster modeling is an integral component of spatial analysis. However, conventional raster modeling techniques can require a substantial amount of processing time and storage space, often limiting the types of analyses that can be performed. To address this issue, we have developed Function Modeling.

Function Modeling is a new modeling framework that streamlines the raster modeling process by utilizing lazy reading methodologies. To assess the efficacy of Function Modeling, we compared the processing time and storage space required to execute six simulation using our newly developed methodology and conventional modeling techniques. Our findings indicate that Function Modeling substantially reduces both processing time and storage space when compared to conventional modeling. Outside of simulations, we have used Function Modeling to characterize the impacts of fuel treatments on soil erosion given fire disturbance, estimate basal area, trees, and tons of above ground biomass per acre, identify locations in need of forest management, calculate forest residuals given multiple management prescriptions, and integrate spatially explicit delivery cost models with forest residual estimates in a fraction of the time and storage space it would take to perform similar analysis using conventional modeling methodologies.

Overall, Function Modeling significantly improves how raster models are processed. To facilitate the use of Function Modeling, we built an object oriented .NET library called RMRS Raster Utility. RMRS Raster Utility is free, readily available, has an intuitive user interface, and directly plugs into Environmental Science Research Institute (ESRI)'s software as an ESRI add-in.

Introduction

Raster modeling is an integral component of spatial analysis and remote sensing. It has been used to address a broad array of questions in forest resources management, ranging from estimating forest carbon (Patenaude et al. 2005) to soft mass availability (Reynolds-Hogland et al. 2006). However, current model processing can require a substantial amount of time and storage space, often limiting the types of analyses that can be performed. As a compromise, some analysts re-sample data to perform analyses, which sacrifices data fidelity. While this can make the data more manageable, it can also blur the relationships between dependent and independent variables. In a recent project we were faced with this dilemma, forcing us to take a closer look at raster modeling and develop novel ways to address typical constraints.

After reviewing the conventional raster modeling techniques, it is clear that the way raster models process data can be improved. Specifically, raster models are composed of multiple spatial operations. Each operation reads data from a given raster dataset, transforms the data, and then creates a new raster dataset (Figure 1). While this process is intuitive, creating and reading new raster datasets at each step of a model comes at a high processing and storage price, prompting two questions: 1) Do we need to create intermediate outputs to get a final model output? 2) If we remove intermediate outputs from the modeling process, what improvements in processing and storage can we expect?

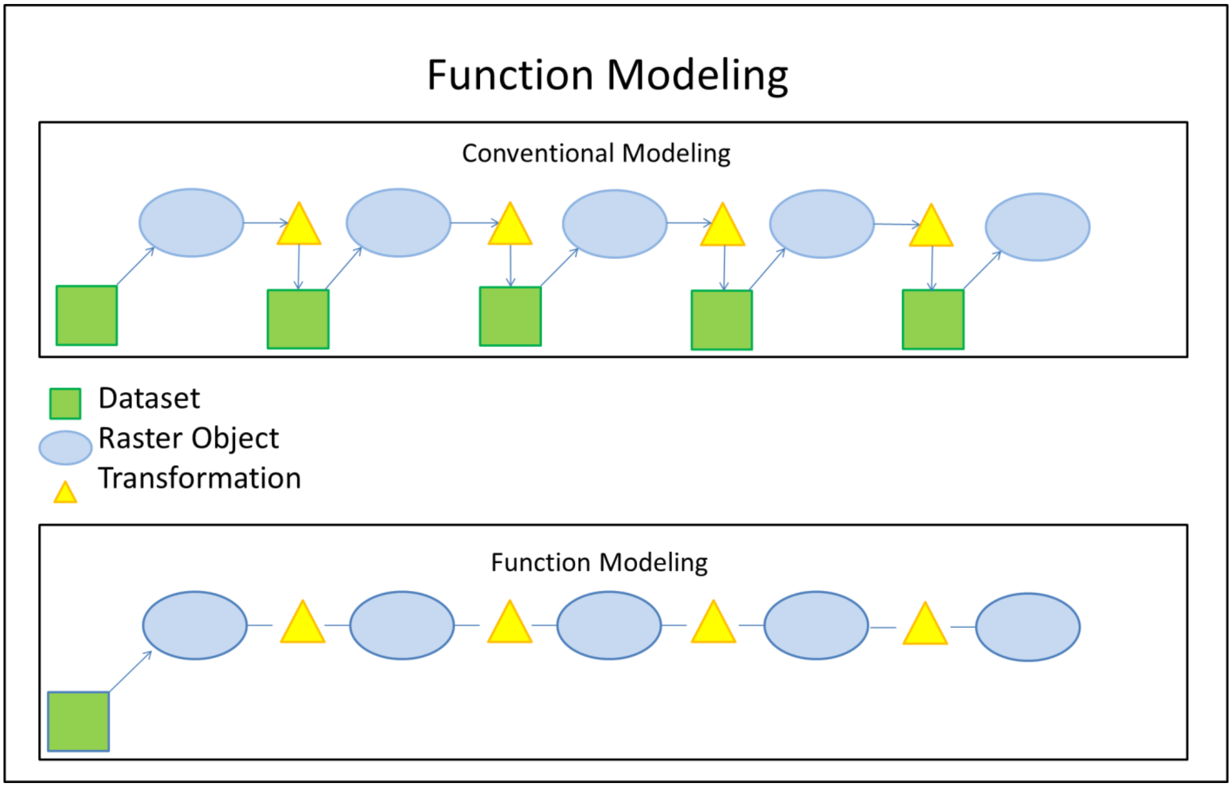

Figure 1. Conceptual idea behind Function Modeling. For each raster transformation (yellow triangles) within a conventional model, an intermediate raster dataset is created (green block). Intermediate raster datasets are then read into a raster object (blue ovals) and transformed to produce another intermediate raster dataset. Function Modeling does not create intermediate raster datasets and only stores the type of transformation occurring to source raster datasets. When raster data are read, this technique performs the transformations to the source datasets without creating intermediate raster datasets.

Methods

To address these questions we developed a modeling framework, called Function Modeling, that allows users to model functions as opposed to raster datasets using an object-oriented design, .NET framework, Environmental Systems Resource Institute (ESRI)'s ArcObjects, and function raster datasets (ESRI 2010). Next, we designed, ran, and recorded processing time and storage space associated with six simulations that varied the size of the raster datasets. Finally, to compare and contrast Function Modeling (FM) with conventional modeling (CM) techniques, we used linear regression. All conventional modeling techniques were directly coded against ESRI's ArcObject Spatial Analyst classes to minimize conventional modeling processing time.

Spatial modeling scenarios within each simulation ranged from one arithmetic operation to twelve operations that included arithmetic, logical, conditional, focal, and summary type analyses (Table 1). Each modeling scenario created a final raster output and was run against six raster datasets ranging in size from 1,000,000 to 121,000,000 total cells, incrementally increasing in size by 2000 columns and rows at each step. Cell numeric precision remained constant as floating type numbers across all scenarios and simulations.

Table 1. Spatial Operations used to compare and contrast Function Modeling and conventional raster modeling. Superscript values indicate the number of times an operation was used within a given model. Model number also indicates the total number of processes performed for a given model.

Model	Spatial Operation Types
1	Arithmetic (+)
2	Arithmetic (+) & Arithmetic (∗)
3	Arithmetic (+), Arithmetic (∗) & Logical (>=)
4	Arithmetic $(+)^2$, Arithmetic (∗) & Logical (>=)
5	Arithmetic $(+)^2$, Arithmetic (∗), Logical (>=) & Focal (Mean,7,7)
6	Arithmetic $(+)^2$, Arithmetic $(*)^2$, Logical (>=) & Focal (Mean,7,7)
7	Arithmetic $(+)^2$, Arithmetic $(*)^2$, Logical (>=), Focal (Mean,7,7), & Conditional
8	Arithmetic $(+)^3$, Arithmetic $(*)^2$, Logical (>=), Focal (Mean,7,7), & Conditional
9	Arithmetic $(+)^3$, Arithmetic $(*)^2$, Logical (>=), Focal (Mean,7,7), Conditional, & Convolution (Sum,5,5)
10	Arithmetic $(+)^3$, Arithmetic $(*)^3$, Logical (>=), Focal (Mean,7,7), Conditional, & Convolution(Sum,5,5)
11	Arithmetic $(+)^3$, Arithmetic $(*)^3$, Logical (>=), Focal (Mean,7,7), Conditional, Convolution(Sum,5,5), & Summary(∑)
12	Arithmetic $(+)^4$, Arithmetic $(*)^3$, Logical (>=), Focal (Mean,7,7), Conditional, Convolution(Sum,5,5), & Summary(∑)

Results

FM significantly reduced both processing time and storage space when compared to CM (Figures 2 and 3). To perform all six simulations FM took 26.13 minutes and a total of 13,299.09 megabytes while CM took 73.15 minutes and 85,098.27 megabytes. Theoretically, processing time and storage space associated with creating raster data for a given model, computer configuration, set of operations, and data type should be a linear function of the total number of cells within a rater dataset and the total number of operations within the model:

$$\text{Time}_i \text{ (seconds)} = \beta_i(\text{cells x operations})_i$$

$$\text{Space}_i \text{ (megabytes)} = \beta_i(\text{cells})_i$$

where i denotes the technique, either FM or CM in this case. Our empirically derived FM Time and Space models nicely fit the linear paradigm (R^2=0.96 and 0.99, respectively). However, for the CM techniques the model for Time (R^2=0.91) deviated slightly and the model for Space (R^2=0.10) deviated significantly from theoretical distributions, indicating that in addition to the modeling operations, extra operations occurred to facilitate raster processing and that some of those operations produced intermediate raster datasets.

FM and CM time equation coefficients indicate that, on average, FM reduced processing time by approximately 286%. Model fit for FM and CM space equations varied greatly between techniques and could not reliably be compared in the same manner. For the CM methodology, storage space appeared to vary depending on the types of operations within the model and often decreased well after the creation of the final dataset because of delayed removal of intermediate datasets. Overall though, FM reduced total storage space by 640% when compared to the CM.

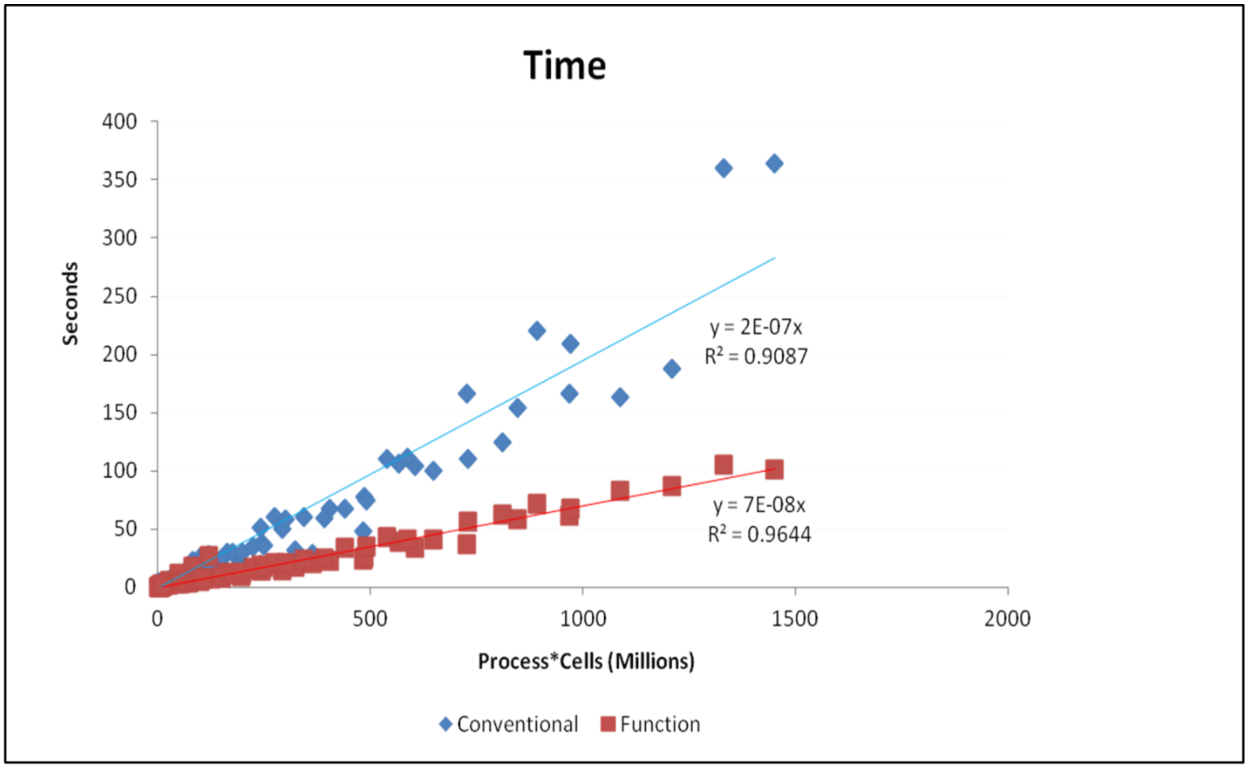

Figure 2. Function Modeling and conventional modeling regression results for processing time.

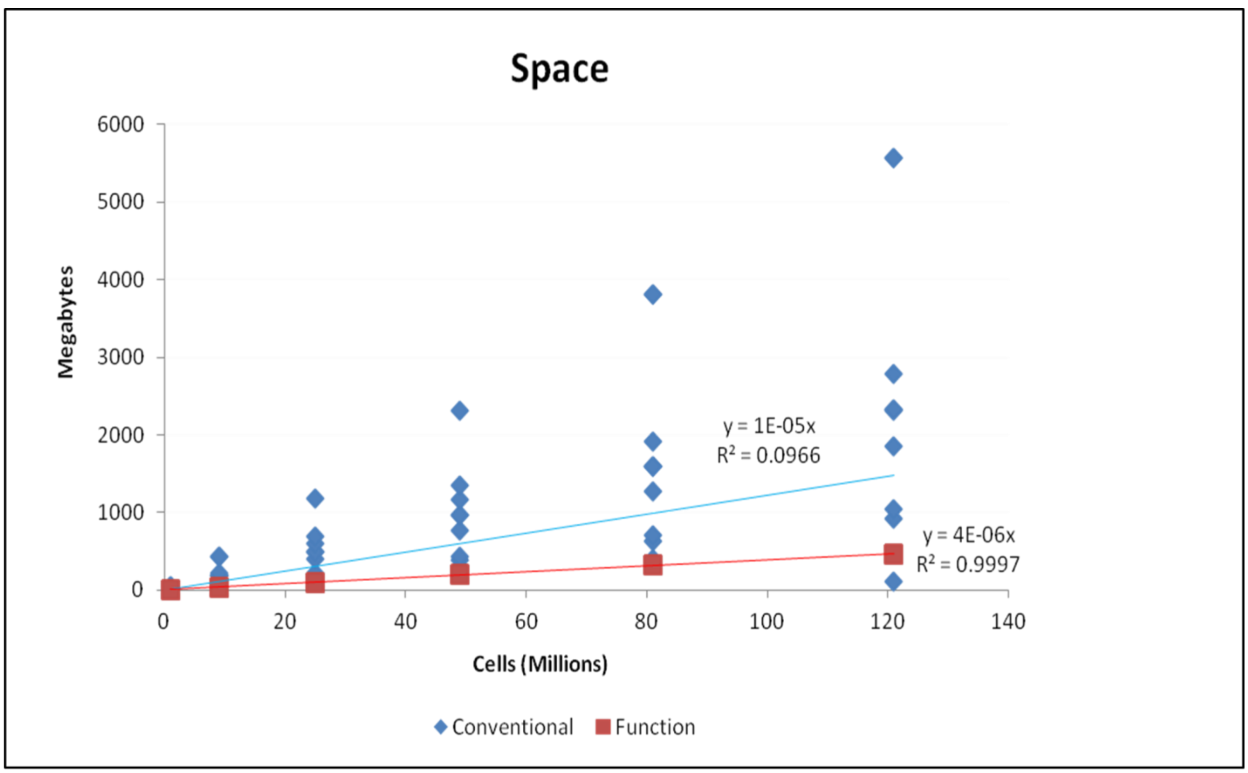

Figure 3. Function Modeling and conventional modeling regression results for storage space

Discussion

CM typically creates intermediate raster datasets. As a model becomes more complex, the number of intermediate datasets increases, which significantly increases the amount of processing time and storage space required to run a model. FM does not create intermediate datasets and significantly decreases model processing time and storage space. In this study we compared FM to CM methodologies by creating final raster outputs from multiple, incrementally complex models. However, FM does not need to create permanent raster datasets. Instead, all functions occurring to source raster datasets can be stored and used in a dynamic fashion to display, visualize, symbolize, retrieve, and manipulate function datasets at a small fraction of the time and space it takes to create a final raster output. For example, if we were to perform the same six simulations presented here but replace the requirement of creating a final output raster dataset with creating a function modeling dataset, FM would take less than 1 second and would require less than 50 kilobytes of storage to finish the simulations.

To facilitate the use of FM we created a user-friendly ESRI toolbar (Figure 4) that provides quick access to a wide array of spatial analyses, while allowing users to easily store and retrieve FM models (RMRS Raster Utility 2012). Moreover, FM models can be loaded into ArcMap as a function raster dataset that can be used interchangeably with all ESRI raster spatial operations. While function modeling is extremely efficient and uses almost no storage space, there may be circumstances when one might want to combine the idea of FM with creating intermediate raster datasets. For example, if the same FM model is used multiple times to create a new raster dataset and the associated time it takes to calculate each operation within that original FM model is greater than the combined read and write time of creating an intermediate raster datasets, then it would be advantageous to create an intermediate dataset. Nonetheless, results from this study demonstrate that raster modeling does not require the creation of intermediate datasets and FM substantially reduces processing time and digital storage. Though straightforward in concept, our work represents the first time the two methodologies have been empirically compared and our publicly available toolbar effectively operationalizes the concepts of FM.

Figure 4. RMRS Raster Utility toolbar is a free, public source, object oriented library packaged as an ESRI add-in that simplifies data acquisition, raster sampling, and statistical and spatial modeling while reducing the processing time and storage space associated with raster analysis. To find out more about the project please visit our website (http://www.fs.fed.us/rm/raster-utility/).

References

ESRI. 2010. ArcGIS Desktop Help 10.0 – Rasters with functions. Available online at http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Rasters_with_functions/009t0000000m000000; last accessed Sept. 4, 2012.

Patenaude, G., R. Milne, and T.P. Dawson. 2005. Synthesis of remote sensing approaches for forest carbon estimation Reporting to the Kyoto Protocol. *Environmental Science and Policy* 8(2): 161–178.

Reynolds-Hogland, M., M. Mitchell, R. Powell. 2006. Spatio-temporal availability of soft mast in clearcuts in the Southern Appalachians. *Forest Ecology and Management* 237(1–3): 103-114.

USFS. 2012. RMRS Raster Utility. Available online at http://www.fs.fed.us/rm/raster-utility; last accessed Sept. 4, 2012.