A Generic Vegetation Growth Sub-model in a Large Human/Environment Interaction Model of East Africa

Timothy Gulden[1], Chenna Reddy Cotla[1], Atesmachew Hailegiorgis[1] and Andrew T. Crooks[1],

[1] Center for Social Complexity, George Mason University, Fairfax, VA 22030
{tgulden, ccotla, ahailegi, acrooks2}@gmu.edu

Abstract. This paper describes a preliminary, lightweight and robust method for simulating vegetation growth that has basic validity in the face of remotely sensed vegetation data while being simple enough to retain conceptual and computational tractability when it is incorporated into a large agent-based model of human subsistence, conflict, and displacement in East Africa. The sub-model predicts daily vegetation values for 2.5 million 1km2 land grid cells using remotely sensed monthly rainfall data. It has been informally validated against remotely sensed, bi-monthly normalized difference vegetation index (NDVI) data. We believe that the approach presented in this sub-model is uniquely well suited to representing dryland vegetation dynamics within a context of a large, human-environment interaction model such as the RiftLand model of which it is part.

Keywords: Human-Environment Interaction, Computational Ecology, Environmental Simulation, Remote Sensing.

1 Introduction

This paper describes a lightweight and robust method for simulating vegetation growth that has basic validity in the face of remotely sensed vegetation data while being simple enough to retain conceptual and computational tractability when it is incorporated into a large agent-based model of human subsistence, conflict, and displacement in East Africa. The sub-model must be simple for two reasons: first, it must have a small enough memory footprint and fast enough execution to avoid slowing down the rest of the model; second, it must be easy enough to understand so that it can be presented as part of the larger modeling effort without requiring too much faith on the part of a critical reader. We seek to achieve these goals by basing the sub-model in fundamental ecological theory while calibrating and verifying the model against extensive remotely sensed data. The result is not a precise fit between model output and data, but rather a general agreement that is robust to various ecosystem types (of which there are many) and an array of weather conditions. While such a model would not serve if vegetation growth itself were the object of study, we believe it has real advantages for a model of the this sort where there are so many potential moving parts that simplifying abstractions must be made at all levels. Furthermore, the presence of extensive data against which to validate the sub-model allows us to develop an objective understanding of how well the sub-model is performing.

This vegetation sub-model is a fundamental part of the RiftLand agent-based model that seeks to model human subsistence, conflict and displacement in East Africa at multiple scales (Cioffi-Revilla, 2011). The RiftLand model covers a 2.5 million square kilometer area that includes all or parts of Kenya, Ethiopia, Sudan, Uganda, Democratic Republic of Congo, Rwanda, Burundi and Tanzania. Vegetation growth and human subsistence are modeled within this area at a spatial resolution of 30 arc seconds (approximately $1km^2$) and a temporal resolution of 1 day. People are modeled at the household level. Ethnic and national identities are based on anthropological literature. Rural households divide their time between herding and farming as dictated by environmental conditions and, in both modes of subsistence, are heavily dependent on rainfall for survival.

When drought comes to an area, people there are stressed and may come into conflict with one another and/or their subsistence activities may fail (see Barnett and Adger, 2007; Reuveny, 2007). In this case, they may be forced out of their accustomed lifestyle, becoming internally displaced people (IDPs). Displaced households may make further moves to cities and camps that have a limited capacity in terms of coping with increases in populations and often as a result, become overwhelmed. Eventually, this migration can place pressure on international borders creating refugee flows between states and international tensions.

The RiftLand model thus involves small-scale human/environment interactions and aggregates these interactions up to larger scales. These larger scales include tension between different livelihood types (herding and farming), neighboring ethnicities, urban and rural populations, interacting cities, various interest groups within a polity, and neighboring states. This makes for a very complex modeling environment where keeping each part as simple as possible (but no simpler!) is essential. Too complex and the model is slow and hard to understand - too simple and the basic dynamics of population stress and movement are unrecognizable. The dynamics of land cover, specifically vegetation, underlie much of the activity within the model.

Our focus within this paper is the sub-model representing vegetation growth that is both spatially and temporally explicit. This vegetation growth model is crucial for the agent-based model that operates on top of such land cover. However, as noted above, we want the vegetation model to be simple (easily explained) and efficient (in terms of computer resources used). The standard measure of vegetation in a regional context is that of Normalized Difference of Vegetation Index (NDVI). NDVI measures the differential reflection of green vegetation in the visible and infrared portion of the spectrum. This can be expressed as:

$$\text{NDVI} = (\text{IR} - \text{R})/(\text{IR} + \text{R}) \qquad \textbf{(1)}$$

Where IR = Infrared reflectance and R = Red reflectance.

NDVI is highly correlated to the fraction of absorbed photosynthetically active radiation and directly indicates the photosynthetic capacity of the land cover. It also correlates with many vegetation indices such as green leaf biomass, leaf area index, and annual net primary productivity. Many potential applications of the data have been explored. For instance, these data may be used as input for modeling global biogeochemical and hydrological processes and global and regional climate, for characterizing land surface biophysical properties and processes, including primary production and land cover conversion.

In the African context, NDVI has been used to assess the green vegetation cover in various environments at a range of spatial and temporal scales. To mention few, NDVI has been used in classifying continental and regional land covers (Tuker *et al.,* 1985a), monitoring land use change (Lambin and Ehrlich, 1997), monitoring vegetation dynamics (Anyamba and Tucker, 2005), assessing herbaceous biomass and dry matter accumulation (Tucker, *et al,* 1985b), monitoring drought condition (Hutchinson, 1991). Importantly for our purposes here, NDVI has been shown to scale approximately linearly with net production of vegetation (Prince, 1991).

Nicholson *et al*. (1990) argued that exploring the relationship between NDVI and soil moisture could provide greater understanding of the environmental constraints on vegetation growth. However, acquiring soil moisture data on a large scale is costly and in most developing countries, in many cases, it is still near to impossible. The most feasible way could be to use rainfall as a proxy. Hence many studies have been conducted to determine the relationship between NDVI and rainfall. These studies tried to measure the cross-correlation between time series of NDVI and rainfall. Although the studies differ in their details, the outcome from most of the studies indicated that there is a strong positive relationship between rainfall and NDVI.

The current state of the art in estimating NDVI given time-series rainfall data is using the geographically weighted regression (GWR) method (see Fotheringham *et al.,* 2002). This method has been applied with considerable success to African drylands by Gaughan *et al*. (2010). In the current context, however, this approach has two problems: First, it is computationally somewhat complex and would be difficult to implement efficiently within the model. Second, it is not fundamentally a dynamic technique. In this application, we are interested not only in having accurate estimates of vegetation given a rainfall time series, but also in how this growth pattern reacts to other shocks that are endogenous to the model - chief among them being grazing by pastoral herds. The method presented below seeks to retain some of the accuracy of the GWR method, while maintaining the fast processing capabilities and dynamic properties that are essential to this modeling application.

2 Methodology

We make use of two remotely-sensed datasets in this simple vegetation model, one for rainfall and one for vegetation. For rainfall, we use Tropical Rainfall Measuring Mission (TRMM[1]) 3B43. The TRMM 3B43 data are provided with a temporal resolution of one month and a spatial resolution of 0.25 by 0.25 degree (approximately 30km by 30km). The data are provided both at global and continental scale.

[1]The data used in this effort were acquired as part of the activities of National Aeronautics and Space Administration’s (NASA's) Science Mission Directorate, and are archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC).

Following the assumption of linear relationship between NDVI and vegetation, we estimate vegetation using the MODIS/TERRA NDVI data set, specifically MOD13A2[2]. The MOD13A2 data are provided every 16 days at a spatial resolution of 1km by 1km as tiles of approximately 1200km by 1200km. We recombine and clip these to the study area using ArcGIS. NDVI values are converted to kilograms of dry matter based on assessments of good pasture in Northern Kenya as estimated by de Leeuw and Tothill (1993). Both datasets have a temporal span from 2001 to 2008.

We use ArcGIS to reformat these data before bringing them into the model. Since both rainfall and NDVI data are provided in hierarchical data format (HDF) format, which is not directly importable by ArcGIS, we first converted the HDF format into a Grid using ArcGIS, particularly using "Make NetCDF Raster Layer" tools from the Multidimension Tools toolbox of ArcGIS to create the raster layers. We use the mosaic to raster function in ArcGIS to aggregate all tiles of the NDVI data that are located within the study area. We then clip the mosaic data by masking the RiftLand boundary. Since rainfall data are provided at global scale, we extract the area of interest by masking the RiftLand boundary. We then project both the NDVI and rainfall data to the coordinate system. Both the NDVI and Rainfall data are then converted to ASCII files and used in the model.

We base our vegetation sub-model on a logistic growth curve along the lines of one described by France and Thornley (1984).

$$V_{t+1} = V_t + GV_t(1 - V/V_{max}) \qquad (\mathbf{2})$$

Were V is the mass of the vegetation on a parcel, G is a growth rate, and V_{max} is the maximum possible vegetation on a single pixel. While this curve has a solid basis in biological theory, we find that we need to modify it slightly by adding a floor and a ceiling to the vegetation amounts. We enforce:

$$0.1 < V < 0.9 \qquad (\mathbf{3})$$

This prevents the model from becoming trapped at zero or V_{max}. Our justification for imposing these constraints have both theoretical and empirical basis. In theoretical terms, the floor value can be thought of as representing roots, seeds, nubs, etc., - inedible and hard-to-kill parts of vegetation that allow it to recover when conditions permit. If an area is truly devoid of plant matter, plants will not grow there no matter how much it might rain, thus any area that recovers after rains must have some amount of plant matter even in the harshest of times. Similarly, the upper value can be thought of as reflecting the work of natural pests and grazers - even under the best conditions, a parcel never produces more than 90% of its potential. Empirically, we find that NDVI values for vegetated parcels in the region generally vary between 0.1 and 0.9. Thus the required model dynamics, theoretical intuition, and observation all point to the imposition of floor and ceiling constraints with values similar to these.

The real action in the model comes from the growth rate (*G*). We model *G* as a function of rainfall, existing vegetation, and land quality.

---

[2]These data are distributed by the Land Processes Distributed Active Archive Center (LP DAAC), located at the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center.

$$G = G_{base} * (aR_t - bV_t + cQ) \quad (\mathbf{4})$$

G_{base} is the base growth rate. At present, G_{base} is uniform for all parcels in the model, though further investigation may reveal that different base growth rates are appropriate for different ecosystem types.

R_t is the rain in the current time step. As mentioned above, our data are calendar monthly totals and we assume consistent daily rainfall for each day within that month. This is, of course, wildly unrealistic. We justify this assumption by thinking of this more as soil moisture, which is a better predictor of plant growth (see Nicholson *et al.,* 1990). Soil moisture will be highly correlated with rainfall, but much less volatile. The parameter 'a' controls how strongly rainfall alters the growth rate.

V_t is the vegetation on the parcel at the present time. The idea here is that more vegetation requires more moisture to maintain a positive growth rate. Thus, if the vegetation requires more rain than is available ($bV_t > aR_t$), the growth rate goes negative and the amount of vegetation available will shrink. As with 'a', the parameter 'b' controls the water demands of the vegetation and the importance of this term.

Finally, Q is the quality of the land. This is an empirically derived, catch all measure that accounts for such things as elevation, ecosystem type, soil type, subsurface hydrology and anything else that alter the way that vegetation relates to rainfall. We derive Q for each parcel by performing a simple uni-variate regression for each parcel using average rainfall as the independent variable and average NDVI as the dependent variable and then using this regression to predict average NDVI from average rainfall. Subtracting predicted average NDVI from observed average NDVI (i.e. the residual) gives a measure of land quality. Where more vegetation grows than one would expect from the given rainfall, the land can be said to be good. Where less grows, the land is less good. While there are many ways that this measure might be incorporated into the model, we have simply used it to further modify the level of rain required to maintain a positive growth rate. The parameter 'c' controls how strongly land quality impacts vegetation growth.

3 Results and Analysis

The vegetation growth sub-model produces general agreement with observed NDVI in a wide variety of situations. While formal validation has yet to be done (and is beyond the scope of this paper) we present graphs of several representative individual parcels comparing predicted to observed NDVI along with rainfall in Figure 1.

To examine the performance of the sub-model in different climatic conditions, we present three graphs comparing model output to NDVI data (see Figure 1). Graph 'a' shows a very dry area near Lake Turkana in Northern Kenya. Graph 'b' shows a moderate area near the coast in Somalia. Graph 'c' shows a farmed area in the lush hills of Burundi. It should be noted that the NDVI data is somewhat noisy and contains occasional zero values which appear as single-period downward spikes.

These are almost certainly artifacts of data collection or data processing and are not a target of the modeling effort.

We find that the growth model, as presented, produces reasonable fits in most ecosystem types. The size of the spikes and dips in vegetation levels associated with wet and dry periods generally correspond to those observed. Also, the timing of these spike and dips is approximately accurate in most cases. While visual inspection of a few pixels does not constitute formal validation, it provides us with some confidence that our functional form is adequate and that further investment in more formal calibration and validation is warranted.

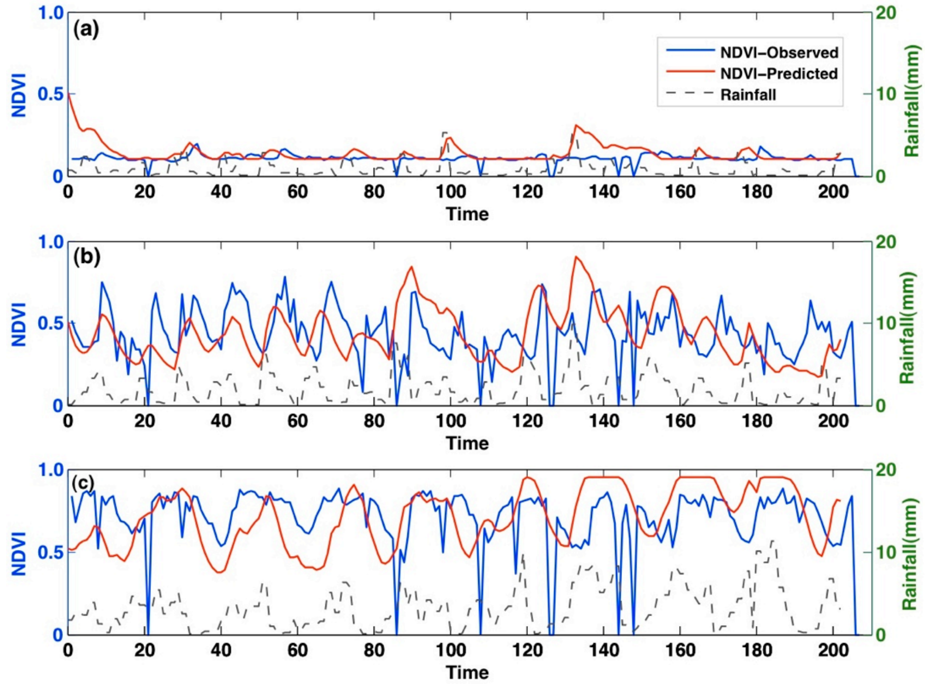

Fig. 1. Observed and simulated vegetation with rainfall for three individual 1km^2 parcels in different climatic regions of the study area. Time intervals are at 16 day periods, the frequency with which NDVI data is collected and spans the years 2001 to 2008. (**a**) a dry area near Lake Turkana in Northern Kenya; (**b**) a moderate area near the coast in Somalia; (**c**) a farmed area in the lush hills of Burundi.

4 Discussion

While the sub-model presented here has only loose basis in biological and ecological first principles, it is quite simple and produces results that appear (at this point in our analysis) to be valid for the application for which the sub-model is intended. We are seeking to model environmental stresses and how they reverberate through the social system in a large part of Africa, resulting in displacement of people, conflict between these people, pressure on international borders, changes in the legitimacy of governments, etc. In order to achieve this goal, we need a vegetation growth sub-model that captures the general dynamics of plant growth in Eastern Africa while abstracting from the details of particular ecosystems and running quickly enough to enable the rest of the model sub-components to be added while still running on the available hardware and software platforms.

It is worth noting that the functional form of the sub-model evolved as a result of an iterative verification and validation process. Our initial prototype growth sub-model, was taken from our earlier HerderLand model (Kennedy *et al.,* 2010), which did not perform well in the context of a much broader set of ecological conditions and real rainfall data used within RiftLand. This failure, however, was not apparent to us until we started critically examining its performance on individual land parcels. Once the problem was identified, we rebuilt the sub-model in MatLab to facilitate rapid adjustment and testing. By comparing the vegetation output of representative parcels with NDVI data, we were able to quickly form and test new hypotheses about the relationship between rainfall and vegetation growth and to quickly focus in on the most significant issues in the model.

It is important to point out that this is a fundamentally empirical model - it is descriptive rather than process oriented. If our goal were to understand vegetation growth, then this model would not serve us well. However, our goal is not to understand vegetation growth, but rather to understand environmental stress, displacement, and conflict. The model needs to serve only as a description of the behavior of the vegetation - not as an explanation for this behavior. Descriptive models of this sort can sometimes outperform process-oriented models, particularly when the first principles of the system in question are not fully captured (Haefner, 2005).

The next steps for this model include more formal calibration and rigorous quantitative validation (in the vain of authors such as Pontius *et al.,* 2008; Visser and Nijs, 2006). Finally, we will merge this sub-model with the full RiftLand model including agents with pastoral and farming behavior. This will require recalibration of the sub-model as we endogenize human impact on vegetation. If our models of human/environment interaction are performing well, it should be possible to achieve better fits with human activity represented than without. The ability of the human/environment interaction sub-models (herding and farming) to improve the performance of the vegetation sub-model will serve as a check on the validity of all three sub-models and on their interactions.

While this sub-model is not, in itself, an example of computational social science, the environmental dynamics that it produces are essential to the success of a large human-environment interaction model such as RiftLand. A massive social science model of this makes unique demands of environmental sub-models, requiring them to be relatively robust, yet simple and lightweight. For this reason, we believe the model presented here to be uniquely suited to social science modeling. A model in computational ecology would likely be much more detailed, but also more complex, more computationally intensive, and possibly less robust across different ecosystem types. The contribution here, therefore, is an approach to modeling ecology that is lightweight and robust enough to produce a meaningful environment over which the social science components of the RiftLand model can operate.

5 Acknowledgements

This work was supported by the Center for Social Complexity at George Mason University and by the Office of Naval Research (ONR) under a Multi- disciplinary University Research Initiative (MURI) grant no. N00014-08-1-0921. The authors would like to acknowledge input from the Mason-HRAF Joint Project on Eastern Africa (MURI Team). The opinions, findings, and conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of the sponsors.

References

1. Anyamba, A., and Tucker, C. (2005), Analysis of Sahelian Vegetation Dynamics Using NOAA-AVHRR NDVI Data from 1981–2003, *Journal of Arid Environments*, 63(3): 596 - 614.
2. Barnett, J., and Adger, W. N. (2007), Climate Change, Human Security and Violent Conflict, *Political Geography*, 26(6): 639-655.
3. Cioffi-Revilla, C., Gulden, T., Kennedy, W., Coletti, M. (2011), MASON RiftLand: An Agent Based Model for Analyzing Conflict, Disasters, and Humanitarian Crises in East Africa. Working Paper, Mason-Yale Joint Project on Eastern Africa, Center for Social Complexity, Krasnow Institute for Advanced Study, George Mason University, Fairfax, Virginia 22030 USA. Available from the first author.
4. de Leeuw, P.N., and Tothill, J. C. (1993), The concept of Rangeland Carrying Capacity in Sub-Saharan Africa — Myth or Reality. In: Behnke Jr., R.H., Scoones, I. and Kerven, C. (Eds.), *Range Ecology at Disequilibrium. New Models of Natural Variability and Pastoral Adaptation in African Savannas*, Overseas Development Institute, London, pp. 77–88.
5. Lambin, E. F., and Ehrlich, D. (1997), Land-cover Changes in Sub-Saharan Africa (1982–1991): Application of a Change Index Based on Remotely Sensed Surface Temperature and Vegetation Indices at a Continental Scale, *Remote Sensing of Environment*, 61(2): 181–200.
6. France, J., and Thornley, J. H. M. (1984), *Mathematical Models in Agriculture*. London: Butterworths.
7. Fotheringham, A.S., Brunsdon, C., and Charlton, M.E. (2002), *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, Wiley, Chichester.
8. Gaughan, A.E., Stevens, F.R., and Gibbes, C. (2010), Linking Climate to Landscape: Investigating the Response of Vegetation to Precipitation Variability in the Okavngo Kwando Zambezi Catchment of Southern Africa. Presented at the *Association of American Geographers*. April 22, 2010. Washington, D.C.
9. Haefner, J. (2005), *Modeling Biological Systems Principles and Applications*. Springer, New York.
10. Hutchinson, C.F. (1991), Uses of Satellite Data for Famine Early Warning in Sub-Saharan Africa, *International Journal of Remote Sensing* 12(6): 1405 - 1421.
11. Kempes CP, West GB, Crowell K, and Girvan M. (2011), *Predicting Maximum Tree Heights and Other Traits from Allometric Scaling and Resource Limitations*. PLoS ONE 6(6): e20551. doi:10.1371/journal.pone.0020551
12. Kennedy, W.G., Hailegiorgis, A.B., Rouleau, M., Bassett, J.K., Coletti, M., Balan, G.C., Gulden, T., and Cioffi-Revilla, C. (2010), *MASON HerderLand: Origins of Conflict in East Africa*. Presented at the Annual Meeting of the Computational Social Science Society, 5-6 Nov 2010, Tempe, AZ.
13. Nicholson, S.E., Davenport, M.L. and Malo, A.R. (1990), A Comparison of the Vegetation Response to Rainfall in the Sahel and East Africa, Using Normalized Difference Vegetation Index from NOAA AVHRR, *Climatic Change,* 17(2-3): 209–241.
14. Prince, S.D. (1991), A Model of Regional Primary Production for Use with Coarse Resolution Satellite Data, *International Journal of Remote Sensing,* 12(6): 1313-1330.
15. Pontius, R.G., Boersma, W., Castella, J.C., Clarke, K.C., Nijs, T., Dietzel, C.K., Duan, Z., Fotsing, E., Goldstein, N.C., Kok, K., Koomen, E., Lippitt, C.D., McConnell, W., Sood, A.M., Pijanowski, B., Pithadia, S., Sweeney, S., Trung, T.N., Veldkamp, A. and Verburg, P.H. (2008), 'Comparing The Input, Output, and Validation Maps for Several Models of Land Change', *The Annals of Regional Science,* 42(1): 11-37.
16. Reuveny, R. (2007), Climate Change-induced Migration and Violent Conflict. *Political Geography*, 26(6): 656-673.

17. Tucker, C. J., Townshend, J. R. G., and T. E. Goff, (1985a), African Land-Cover Classification Using Satellite Data, *Science*, 227(4685): 369-375.
18. Tucker, C.J., Vanpraet, C.L., Sharman, M.J., and G. Van Ittersum, (1985b), Satellite Remote Sensing of Total Herbaceous Biomass Production in the Senegalese Sahel: 1980-1984, *Remote Sensing of Environment*, 17(3): 233-249.
19. Visser, H. and de Nijs, T. (2006), The Map Comparison Kit, *Environmental Modelling & Software*, 21(3): 346-358.