IES
INSTITUTE OF
ENVIRONMENTAL
STUDIES
University of Zimbabwe

ON THE DESIGN OF AN UNRESTRICTED DATA UNIT (UDU) CONCEPTUAL MODEL FOR GEO-SPATIAL DATA

D. Semwayo and Matambanadzo

ON THE DESIGN OF AN UNRESTRICTED DATA UNIT (UDU) CONCEPTUAL MODEL FOR GEO-SPATIAL DATA

D. Semwayo and P. Matambanadzo

2000

ES Working Paper 14
Published by the Institute of Environmental Studies,
University of Zimbabwe, Harare, Zimbabwe

Institute of Environmental Studies
University of Zimbabwe
P O Box MP167
Mount Pleasant
HARARE

Telephone 263-4-302603/332039 (direct) or 303211 ext. 1937
Fax 263-4-332853/333407
E-mail: campbell@africaonline.co.zw

All Working Papers of IES are peer reviewed For reviewing documents in 1999 we thank: Simon Choge, Delalie Dovie, Guilford Hapanyengwi, Grant Hauer, Taelo Letsela, Marty Luckert, Elias Madzudzo, David Maingi, Patrick Mamimine, E. Manzungu, Sibongile Moyo, Nontokozo Nemarundwe, Sheona Shackleton and Terry Veeman

ON THE DESIGN OF AN UNRESTRICTED DATA UNIT (UDU) CONCEPTUAL MODEL FOR GEO-SPATIAL DATA

D. Semwayo and P. Matambanadzo

Institute of Environmental Studies, University of Zimbabwe

Abstract

The maturing spatial information sciences have led to greater free flow of spatial information. More than ever before, scientists within and between scientific disciplines appreciate the need to exchange environmental information to avert irreversible environmental disasters, hence the development of environmental information systems.

Existing ecological classification schema are seen as an impediment to environmental data exchange between scientific disciplines. However, this paper will show that the perceived different ecological classification systems, though different, are not as incompatible as they might appear to be. It will be shown that the perceived problem of classification schema incompatibility is one of object definition and data structuring and lack of adequately structured meta-data.

The advent of the Internet and it's associated technologies has led to immense possibilities for data exchange. Coupled with those opportunities are perhaps equally frightening possibilities of the use of data of undefined quality obtained from remote databases lacking adequate documentation on the data sets.

The development of an elegant data model based on the concept of object hierarchies and their associated behavioural attributes enables the capture, storage, and retrieval of data objects in a way that enables the aggregation of the objects into several ecological classification schema. Such a framework would facilitate the exchange of data between scientists and nations with seemingly different ecological classification systems. By carefully capturing meta-data incorporating it, and propagating it through the different hierarchical schema via the development of supporting logical model constructs, it is hoped that the data model will promote the informed multiple use of data from differently focused ecological classification and aggregation schemes from distributed sources.

Introduction

When scientists delineate **ecological "units/ systems**[1]**"** they are in fact mapping a **portion** of ecological space, (also refered to as a landscape), depending on the mapping objective. Zonneveld, 1995, describes a landscape as a complex of relationship systems, together forming (also by virtue of it's physiognomy) a recognisable part of the earth's surface, and is formed and maintained by mutual action of abiotic and biotic forces as well as human actions. Ecological systems are by nature inter-related, and in mapping activities we are applying a **reductionist** method of abstraction from the general to the specific in order to zero in on a particular theme of interest (Zonneveld, 1995). This is all done, ostensibly, to enhance our understanding of the underlying and **inter-related processes** in a chosen portion of ecological space.

[1] An ecosystem is comprehensive whole "holon", a correlative complex including physiognomy of vegetation, relief, soil conditions, and land use (man's culture- activities), the visual and otherwise known aspects as well (as yet) unknown ones, which may be discovered after more intensive study.
(Smuts, 1932)

Ecological modeling

Proponents of modelling methods that use the systems[2] approach argue that the ecosystems can only be modelled as a whole and not as segments as is the case in scientific disciplinary approaches. ***Von Bertalanffy*** (1973) argues that those problems with multivariable interactions and hierarchic orders cannot (even in Physics) be solved by the Cartesian way of looking from the small to the large system. He goes on further to say that it is sufficient to state that a systems approach, strongly supported from elsewhere in science, is the most appropriate to tackle such a complicated subject as landscape with it's strong hierarchical structure.

Molenaar, 1997, brings in an interesting dimension to the debate. He contends that processes on the earth's surface can only be monitored and managed if they are understood in their geographical context, partly defined by the scale range at which these processes work . This view is shared by other researchers (van Gils *et al.*, 1991) Perdizao and Annina 1997)

[2] A system can be thought of as being a set of relationships

Figure 1 above illustrates the problems inherent in the partitioning of ecological space with the internal processes and interaction between it's members. The overlaying of thematic layers in GIS is in fact an attempt at unifying the partitioned ecological space to rediscover the linkages that occur naturally in reality, (see fig 2 below).

Spatial patterns may be discerned from an overlay procedure. However this method of data integration falls short in defining the natural complex inter-relationships between the various "themes" or features (represented by "thematic layers") simply because a cartesian approach is used; whereas in an ecological system, *the sum of the parts is not equivalent to the whole.*

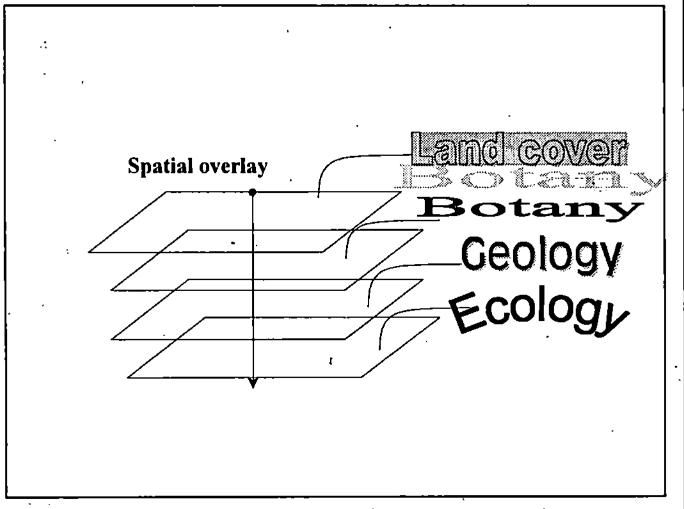

Figure 1: Thematic integration using the overlay procedure

Conceptual framework

This paper is concerned with the design of a **conceptual framework** that supports different **ecological thematic mappings** using the same or similar objects represented as **data units** in a hierachical structure and stored in a geographical information system[3]. In other words we are concerned with the design of a **data model** with **structures** that support the multiple use of ecological data for different scientific applications (mappings) at different scales of observation. The argument here is that we are capturing data from a given ecological space (geographical space) as observable features / objects with specific attributes and **inter-relationships** between themselves. Thus, given *apriori* knowledge of the required information products from a geographical information system, other users can use the same objects to produce different (thematic) information products with little or no additions and subtractions of attributes by primarily incorporating **metadata**[4] about the terrain objects.

Metadata

Metadata in this context is data about geospatial data. Metadata in a GIS environment is used to;

a) indicate availability of data needed to determine the sets of data that exist for a geographic location.

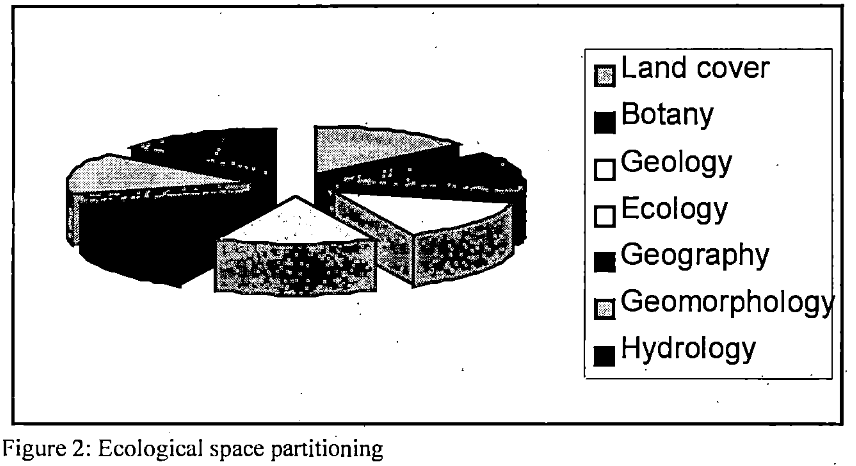

Figure 2: Ecological space partitioning

[3] A geographical information system that enables, capture, modeling, manipulation, retrieval, analysis and presentation of geographically referenced data (Worboys 1996)

[4] Metadata is data or information about data

b) provide information on the accessibility restrictions, if any, placed on a data set.
c) determine fitness for use, that is, data needed to determine if a set of data meets a specific need
d) determine data needed to process and use a set of data.

Metadata conveys the information on a data set; identification, data quality, lineage, spatial data representation data, amongst others. This paper is concerned with the issues relating to c) and d) above. It is envisaged that, in a feature or object based GIS data modelling environment, metadata can be incorporated in the object schema as an attribute of the object to facilitate the informed use of stored data by multiple users.

Kufoniyi, 1997, describes the major components of terrain objects. The terrain object has geometric elements (geometric data) and thematic elements (attribute data), see fig.3 below. He further describes a structure of spatial relations broken down to; topologic relations, metric attributes and a spatial order.

The proposal here is to extend this concept by incorporating meta-data, figure 2. The meta-data will play a part in the description of thematic relationships between terrain objects. The proposed model will look at **thematic order** and how it can be influenced by using **meta-data** to evaluate the use or non-use of data objects at different classification and aggregation levels.

Multiple use of objects

The **attributes** of the objects to be captured will depend on the area of thematic investigation. For instance in the case, the **object** tree, a botanist may define attributes like age, height, leaf shape, canopy cover and so forth. A herbalist may define attributes like age, leaf size, leaf medicinal properties, bark medicinal properties and so forth. From the foregoing it can be seen that both scientists are interested in the same object, tree. They each extract attributes of interest. The object tree does not change it's inherent form as a result of the "artificial" partitionings. It's **geographical location** is stable. The biotic and abiotic influences on the tree remain the same whether we call the tree an Acacia tree or medicine tree. Thus, it makes sense to store the object tree as a **data unit** in a database as it's location, nature and form are stable. We can then simply add other attributes of interest and relevance to the definition of the object tree. Different biotic and abiotic influences (processes) may cause the tree to have variations to it's form (defined by the status of the attributes), but the identification of the tree as defined by it's geographical location remains stable, at least until it dies.

As may be appreciated, the different scientists have different methods of data collection and data definition. This causes problems in **data exchange** and **re-usability**. In addition, their requirements in terms of thematic and locational data may differ. Issues of whether a tree's age is accurately assessed may be trivial to the botanist, but of paramount importance to the herbalist whose precise knowledge of the tree's age may be critical to the usefulness of the medicinal properties of the leaf. This is where **meta-data**[5] becomes useful. We shall argue that by developing a hierarchical structured data model closely related to the naturally existing ecological systems' objects (as perceived by man) and (**encapsulating**) incorporating meta-data on; *purpose of data collection*, *methods of survey*, *locational accuracy*, *error propagation* etc, the next system user can make informed decisions on the suitability of the use of the data residing in a GIS database for his thematic investigations. It is hoped that this will in essence enhance the intelligent use of data.

Aggregation hierarchies and metadata: The Unlimited Data Unit (UDU)

Brode & Ridjanovic (1984), Nyerges (1980), Nyerges (1991a), Molenaar (1996), Huizing (1993) describe a feature based data model of a feature based landscape which embraces the hierarchical ordering of phenomena[6]. At different levels of

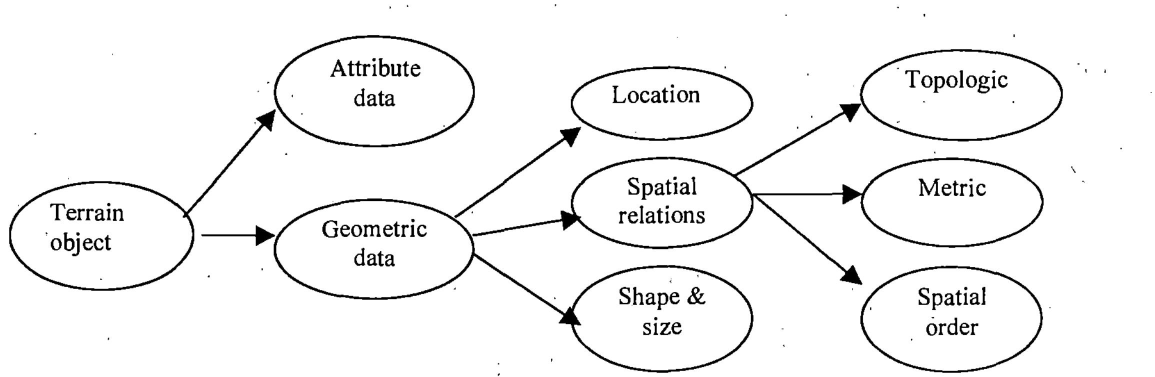

Figure 3: Components of a terrain objects Kufoniyi, 1997

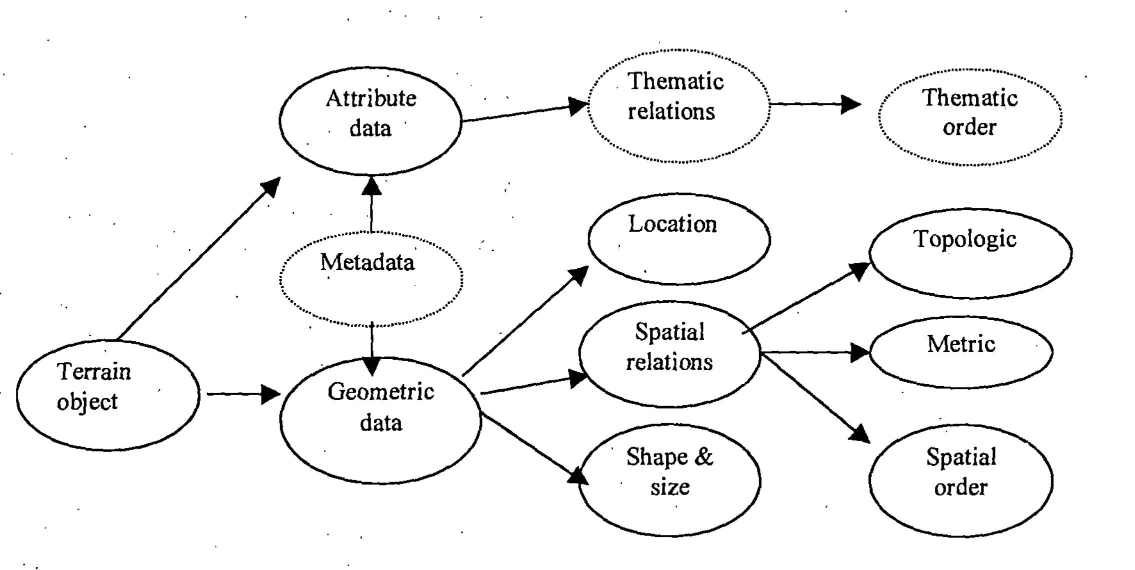

Figure 4: Components of a terrain object with metadata (modified after Kufoniyi, 1997)

observation. The data model defines the following basic components: **classification hierarchies**[7], **aggregation hierarchies**[8] and **associations.** This paper will use this general structure to attempt a data definition for land cover classification and aggregation schema and extend the concept by incorporating meta data (on the nature of object abstraction and classification or aggregation processes) as part of the attribute of terrain object at different hierarchical levels. See figure 4 above.

A land cover classification system is developed to serve the individual objectives of, the ecologist, the botanist, the vegetation scientist and the geographer. These scientists deal with classification systems (1) and (2) (figure 4 above). One can intuitively imagine that a scientist mapping vegetation cover, classification system (1) above, might start at the field level, identifying individual trees, *Brachestigia, Julbernardia*, and *Cliffortia*. At a higher level in a classification hierarchy (object class), one might define two class types called (*or with an instance of*) sub-family *Caesalpinioideae*, and sub-family *Compositae*. The class members for each class being sub-family *Caesalpinioideae*,(*Brachestigia, Julbernardia*) and sub-family *Compositae*, (*Cliffortia*). A further *superclass*, family Fabaceae would have as it's members (*Caesalpinioideae*), other qualifying sub-family types.

A vegetation scientist interested in mapping vegetation structure on the other hand would use a top down approach which begins by mapping the global patterns in land cover and, depending on the detail required, define lower sub-strata of canopies (classification system (2) above. Such a classification method would result in the definition of the mapping context for each hierarchical level of classification. This could approximate to levels I and II of the Anderson levels of vegetation classification (Skidmore *et al*, 1997), for example **woodland** at level I and open woodland (say less than 40% woody cover) at level II. Such classification would most likely use satellite data.

It will be realised that beyond the mapping scientists, other users rely on the land cover classification schemes developed. These include land use planners, resource managers, wild life managers, watershed managers, and land developers. Most of these ancillary users usually have to use land cover maps from different sources. For example, an agricultural land use planner may need to work with a combination of structural and floristic vegetation cover maps to make decisions on range management planning. The planner deals with associations of landscape features and thus uses data units from classification systems (1) and (2) (figure 4 above), to come up with a resource cover map (classification system 3). Assuming data is stored in a GIS database, then to facilitate the intelligent use of such data the planner would require information on the data, i.e. **metadata**. The following information would be critical:

- Source scale and resolution

[7] Classes are collections of objects with the same attribute structure. Class hierarchies at several generalisation levels can be defined with their attributes and their intentions.

[8] An aggregation hierarchy expresses the relationship between a specific aggregated object and it's constituent parts at different levels.

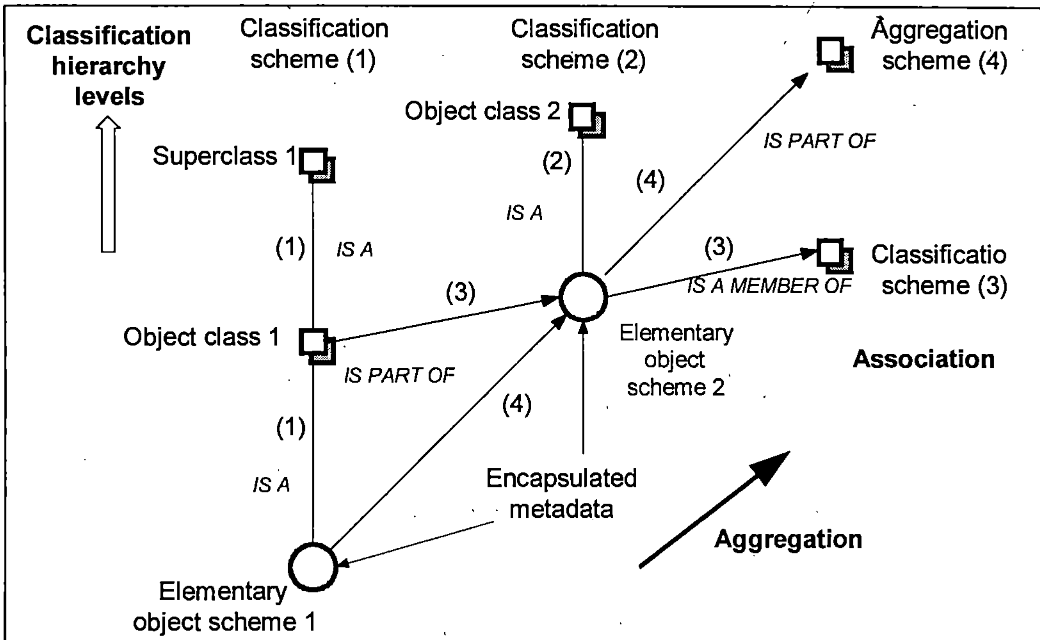

Figure 5: Classification and Aggregation hierarchies incorporating meta-data

- Original purpose or context of data collection
- Co-ordinate projection system (geographic data) used
- Quality information on data source, and method of data collection
- Methods of data processing

Potential or actual incorporated thematic and spatial errors (fuzziness)

The proposed data model describes an architecture not limited by unknown information on the data. The new model will have data "units" that incorporate meta-data which will enable the processing and use of data from different sources, hence the term **"Unlimited Data Unit" (UDU)** model .

5. *Future work*

This paper has touched on the issues of geo-spatial data modelling in a GIS environment. It has been shown that ecological systems occupy geographic space and have to be modelled using a systems approach taking into account their spatial nature.

An object oriented method has been used to develop the UDU model. This presents a hierachical framework that incorporates metadata to enhance the informed multi-use of objects.

By carefully capturing meta-data, incorporating it and **propagating** it through the different hierarchical schema via the development of supporting **logical model constructs**, it is hoped that the data model will promote the **informed** multiple use of data from differently focused ecological classification and aggregation schemes. Future work will involve the use of object oriented constructs to develop a logical data model to develop structures to support the incorporation of metadata within objects representing world features.

References

Benyon, D., (1997) Information and Data modelling, McGraw-Hill

Cheng T., Molenaar M., (1997) A process-oriented spatio-temporal data model to support physical environmental modeling: *Proceedings of the 8th International Symposium on Spatial Data Handling, pp.418-430*

Cheng T., Molenaar, M.,Dynamics of fuzzy objects: Proceedings of the International Workshop on Dynamic and Multi-dimensional GIS, Hong Kong, pp49-63

Kufoniyi, O., (1997) Spatial coincidence modelling, automated database updating and data consistency in vector GIS

Lambin, E.F.(1997) Modelling and monitoring land cover change processes in tropical regions: Progress in Physical Geography 21(3) pp375-393

Molenaar, M. Fritsch, D. (1991) Combined data structures for vector and raster presentations in Geographic Information Systems.

Perdigao.,V , Annoni, A.(1997) Technical and methodological Guide for Updating Corine land Cover Data Base.

van Gils,H., Huizing, H., Kannegieter, A., van der Zee, D. The Evolution of the ITC system of rural land use and land cover classification, *ITC Journal* 1991-3

Von Bertalanffy, L. (1973) General Systems theory, Foundations, Development, Applications.

IES Advisory Board

IES Working Paper Series

1. Frost, P.G.H., Moyo, S., Katerere, Y. and Rukuni, M. 1994. Institute of Environmental Studies planning study. 85 pp.
2. Matose, F., Mudhara, M and Mushove P. 1997. The woodcraft industry of the Bulawayo-Victoria Falls road. 35 pp.
3. Vermeulen, S. 1997. Sharing a state-owned resource with local residents: the case of wood in Mafungabusi State Forest and Gokwe communal area 11 pp.
4. Mandondo, A. and Jackson, J. 1996. Management and performance of *Eucalyptus camaldulensis* in Murewa and Mutoko districts of Zimbabwe. 9 pp.
5. Mandondo, A. 1996. Applied interdisciplinary research and training on natural resource management: a situation report for southern Africa. 15 pp.
6. Manjengwa, J.M. 1997. A preliminary survey of environmental awareness in some secondary school pupils in Zimbabwe. 17 pp.
7. Gwaai Working Group, 1997. Local-level valuation of village woodlands and state forest: Cases from Matebeleland South in Zimbabwe. 29 pp.
8. Nhira, C. 1998. Land use planning and woodland management. A case study of local control and regulatory capacity on household and communal woodland resources in Zimbabwe. 22pp.
9. Mhlanga, L. and Mapaure, I. 1999. Vegetation studies of selected islands and adjacent mainland on Lake Kariba, Zimbabwe. 20 pp.
10. Mapaure, I. 1998. An annotated bibliography of research undertaken in Sengwa Wildlife Research Area (SWRA) and its immediate surrounds. 28 pp.
11. Matambanadzo, P., Semwayo, D. and Hapanyengwi, G. 1999. Towards designing an integrated framework for geo-information based on the persistent paradigm. 10 pp.
12. Frost, P. and Mandondo, A. Improving rural livelihoods in semi-arid regions through management of micro-catchments. 19 pp.
13. Campbell, B., de Jong, W., Luckert, M., Mandondo, A., Matose, F., Nemarundwe, N. and Sithole, B. (in press) Challenges to proponents of CPR systems - despairing voices from the social forests of Zimbabwe.
14. Semwayo, D. and Matambanadzo, P. (in press) On the design of an unrestricted data unit (UDU) conceptual model for geo-spatial data.
15. Mabalauta Working Group (in press). The ecology, control and economics of the Ilala palm in Sengwe communal area, Zimbabwe.
16. Nemarundwe, N. (in press). Institutional collaboration and social learning for forest management in Chivi District, Zimbabwe.
17. Mukamuri, B. and Kozanayi, W. (in press). Institutions surrounding use of marketed bark products: The case of *Berchemia discolor*, *Warburgia salutaris* and *Adansonia digitata*.
18. Hot Springs Working Group (in press).Household livelihoods marketing and resource impacts: A case study of bark products in Eastern Zimbabwe.

Yuan, M., (1996) Temporal GIS and spatio-temporal modeling, Proceedings of the Third International Conference/Workshop on Integrating Geographical Information Systems and Environmental Modeling

Other Publications

1. Institute of Environmental Studies Strategic Plan. 1995
2. Campbell, B. and Kamukondiwa, W. 1995. Zimbabwe Soil Biology and Fertility Project.
3. Campbell, B. 1996 (ed) The miombo in transition: Woodlands and welfare in Africa. CIFOR, Bogor, Indonesia.
4. Environmental Research Co-ordinating Committee. 1998. Strategic Directions for Environmental Research in Zimbabwe. 12 pp.