IMPLEMENTATION OF A COMPREHENSIVE AND EFFECTIVE GEOPROCESSING WORKFLOW ENVIRONMENT

Rob Lemmens[1*], Bert Toxopeus[2], Luc Boerboom[3], Martin Schouwenburg[4], Bas Retsios[5]
Willem Nieuwenhuis[6], Chris Mannaerts[7]

University of Twente, Faculty of Geo-Information Science and Earth Observation, The Netherlands
[(1)] r.l.g.lemmens@utwente.nl [(2)] a.g.toxopeus@utwente.nl [(3)] l.g.j.boerboom@utwente.nl
[(4)] m.l.schouwenburg@utwente.nl [(5)] v.retsios@utwente.nl
[(6)] w.nieuwenhuis@utwente.nl, [(7)] c.m.m.mannaerts@utwente.nl

Commission IV, WG IV/4

Keywords: Workflow, Geoprocessing, Remote Sensing, Earth Observation

ABSTRACT:

Many projects and research efforts implement geo-information (GI) workflows, ranging from very basic ones to complicated software processing chains. The creation of these workflows normally needs considerable expertise and sharing them is often hampered by undocumented and non-interoperable geoprocessing implementations. We believe that the visual representation of workflows can help in the creation, sharing and understanding of software processing of geodata. In our efforts we aim at bridging abstract and concrete workflow representations for the sake of easing the creation and sharing of simple geoprocessing logic within and across projects.

We have implemented a first version of our workflow approach in one of our current projects. MARIS, the Mara Rangeland Information System, is being developed in the Mau Mara Serengeti Sustainable Water Initiative (MaMaSe). It is a web client that uses the Integrated Land and Water Information System (ILWIS), our open source Remote Sensing and GIS software. It aims to integrate historic, near real time and near future forecast of rainfall, biomass, carrying capacity and livestock market information for the sustainable management of rangelands by conservancies in the Maasai Mara in Kenya. More importantly it aims to show results of a carrying capacity model implemented in a comprehensive geoprocessing workflow.

In this paper we briefly describe our software and show the workflow implementation strategy and discuss the innovative aspects of our approach as well as our project evaluation and the opportunities for further grounding of our software development.

1. INTRODUCTION

The University of Twente develops the Integrated Land and Water Information System (ILWIS, http://52north.org/communities/ilwis/). ILWIS is an open source, C++ based, Earth Observation and GIS software. ILWIS delivers a wide range of features including import/export, digitizing, editing, analysis and visualization of geodata (ILWIS, 2016). ILWIS is currently being renewed and transformed into a more modular platform called ILWIS-Objects along with a redesigned plug-in platform and APIs. Researchers, trainers and students can now easily implement, store and share their methods via software, in addition to their written reports. Part of this effort is the development of a plug-in for QGIS. After a prototype for QGIS 2, we are now embarking on an integration in QGIS 3.

One of the design principles for ILWIS is to support education in Remote Sensing and GIS operations which drives us to embark upon a highly interactive drag-and-drop interface (see Figure 1), the use of map thumbnails and linked views: changes in the data properties window are immediately visible in the map window. The underlying framework, ILWIS-Objects, supports the extension of functionality through Python scripts and the creation of new connectors to data formats and other software libraries. This creates a low threshold for researchers, students and teachers to make tailored applications and demonstrators.

ILWIS-Objects is based on the QT framework (C++) together with the GEOS framework(https://trac.osgeo.org/geos/), the boost libraries (https://www.boost.org/) and PROJ.4 (https://proj4.org/).
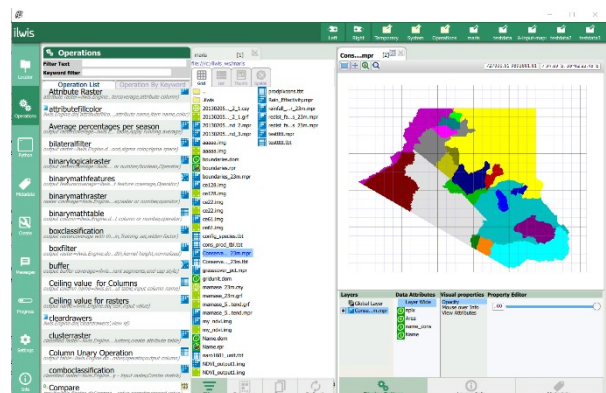

Figure 1. User Interface of the ILWIS desktop, showing its operation and data pane and the map with conservancies in the Maasai Mara in Kenya, the use case focus of this paper (see Section 3.1).

The SQLite (https://www.sqlite.org/) database is used as internal database. The backend is an extensible set of plugins for processing and accessing data. At the moment there are plugins for GDAL(2+), PostGIS, various spread sheet formats, OpenCV, the GNU Scientific library and a set of in-house developed algorithms for raster and vector processing.

The desktop client (ILWIS 4) is based on the QML framework (https://doc.qt.io/qt-5.10/qmlapplications.html, part of QT) and uses three.js as main engine for rendering spatial data.

2. WORKFLOW MODELING

We created a workflow builder that facilitates a drag-and-drop canvas with which the user can visually create geoprocessing workflows and perform a debugging-style visual validation of each part. Its metadata annotations clarify the semantics of the elements. The graphical user interface is shown in Figure 2. The ILWIS workflow builder stands out in supporting the handling of aggregated workflows, loop constructs and conditional statements in order to support the larger GI workflows and make them insightful for people who want to reuse/reproduce them. This helps users with different skill levels to create and communicate geoprocessing workflows. These workflows can be shared through a JSON file with other applications. The JSON file gives a tree-structured representation of workflow inputs and outputs for the Web. Work is underway to smoothen the conversion of JSON workflows to Python code and vice-versa.
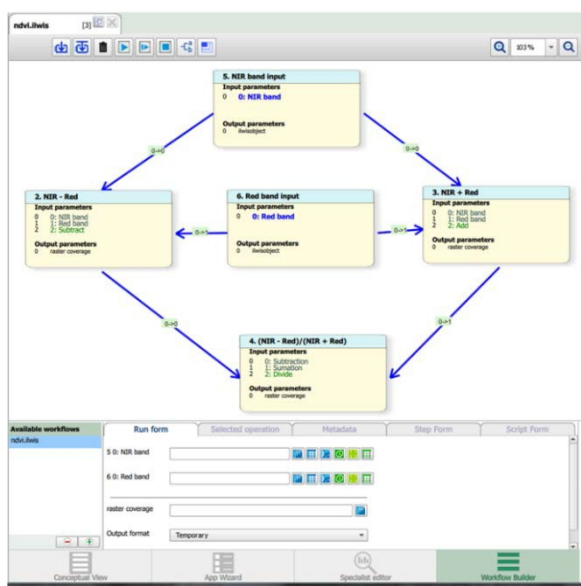

Figure 2. Workflow builder in ILWIS. Depicted is the workflow for the calculation of a Normalized Difference Vegetation Index (NDVI).

To enable a step-wise workflow construction process we distinguish between 1) abstract workflows, in which humans express their process logic, but not necessarily know or want to express the actual data and software operations and 2) concrete workflows, which contain process steps executable by a specific software (Lemmens et al., 2016). By bridging abstract and concrete workflow representations, users can better understand the geoprocessing logic and share workflows within and across projects. In a follow-up development we are connecting to geodata and geoprocessing ontologies (De Carvalho Diniz, F. (2016) in order to be able to translate concrete workflows into abstract workflows and – in the other direction- to support users to create workflows from abstract to concrete ones.

Workflows have been developed to standardize business processes. An example is the ArcGIS Workflow Manager, a proprietary product of ESRI Inc. They have also been developed to control complex scientific analyses. Examples are Dinamica-Ego, a freeware, but not open source, product of the Centro de Sensoriamento Remoto. The former is more used in vector data environments, whereas the latter is exclusively used in raster data

environments. The ILWIS workflow manager is similar to the latter example aiming to control complex raster data analysis with the goal to perform system analytic workflow analysis as can be done in system dynamics software packages such as Stella or Vensim. However, it is also similar to the former, in that it aims for workflows to be used in regularly recurring management practices.

3. WORKFLOW USAGE

We will now describe how our current workflow environment is put to practice. The software development was partly project-driven, giving us the opportunity to test it in a real-life environment from the start.

3.1 Use case application of ILWIS 4 workflow builder

We have implemented our current workflow approach in one of our projects. MARIS, The Mara Rangeland Information System, has been developed in the Mau Mara Serengeti Sustainable Water Initiative (MaMaSe), funded by the Netherlands embassy in Kenya. It is a web client that uses ILWIS. It aims to integrate historic, near real time and near future forecasts of rainfall, biomass, and carrying capacity as well as livestock market information for the sustainable management of rangelands by conservancies in the Maasai Mara in Kenya. More importantly, it aims to show results of a carrying capacity model implemented in a comprehensive geoprocessing workflow. And these analyses are to be used regularly by conservancies to establish and control their grazing management plans.

MARIS (including interactive modelling) is to be used by agencies such as wildlife conservancies, banks and insurance companies to predict the results of certain management actions. These results, i.e., spatial and temporal changes, of carrying capacity of the rangelands for wildlife and livestock are presented in the form of maps and tables, in the MARIS web client, all produced by ILWIS workflows.

3.2 Mara middle basin ecosystem

The Mara Middle Basin is a typical example of a semi-arid savannah ecosystem e.g., zones V (Bekure et al., 1991, Braun and Mungai, 1981, Touber, 1983) in eastern Africa. The ecosystem is characterised by low and seasonal bimodal rainfall. The long rainy season is (in general) from March until May and the plains having the higher rainfalls and lower in the hilly areas, while the short rainy season is from October till the end of December Rainfall in the intermediate period are again highest in all the hilly areas and lower in the plains.
In rangelands areas such as Narok District, the close relationship between rainfall and primary production makes rainfall the most important climatic variable. The seasonal distribution of rainfall governs the fundamental patterns of range utilisation by the nomadic pastoralists and migratory wildlife, while longer term fluctuations can affect their overall abundance and the health, economic welfare and political stability of the people dependent upon them (Norton-Griffiths, 1977).

3.3 Problem analysis and problem structuring

Forage shortage or overgrazing, as a result of both overstocking of livestock and a surplus of wildlife is considered as one of the most severe problems concerning the conservation of the ecosystem. The presence of livestock in the Maasai Mara reserve, especially during the dry season, when the availability of forage and

water is often problematic, was mainly due to the poor distribution of water resources (drinking water) in the ecosystem, leading to localised overstocking of livestock and finally trespassing into the core zone.

On the other hand, high human population growth has led to changes in land use, like increased cultivation of (irrigated) crops. This has led to fencing of plots (to protect the agricultural field against damages, caused by wildlife) and harassment of wildlife, which in turn might result in the loss of the traditional migratory routes of wildlife during the wet season. Wildlife tends to stay longer or permanent in certain areas, leading to overstocking (overgrazing).

For grazing management, we developed a spatio-temporal model of forage (grass) biomass production, another model of forage consumption by livestock and wildlife and a third model that subtracts grassland consumption from grassland production to establish rangeland carrying capacity. These three models were implemented in three different ILWIS workflows.

3.4 Forage productivity

The productivity of rangelands can be described as a function of rainfall, soil condition, topography, and structural vegetation cover, at 23-meter resolution. Particularly grass cover was of interest to assess carrying capacity for livestock and wildlife (see Figure 3). The structural vegetation cover map was compiled through interpretation of satellite images (TM, NDVI_MODIS-terra), google images, data collected during field surveys, of which each mapping unit represents the percentage cover of the different vegetation layers (e.g., perennial grasses).
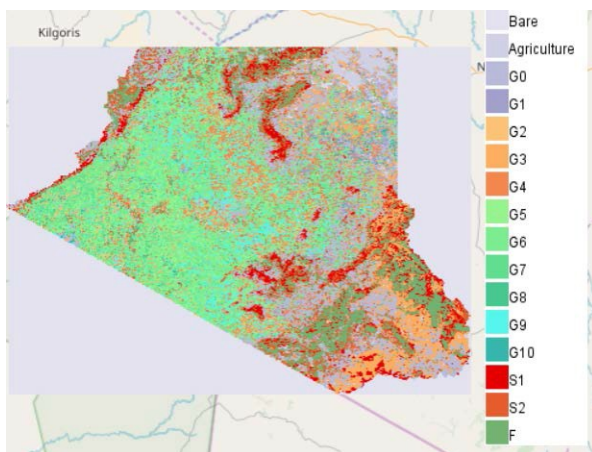

Figure 3. The structural vegetation cover map with the different floristic grass compositions per cover type.

The rainfall map is obtained from different satellite products. Together these products make a nearly continuous sequence of historic rainfall data and near future rainfall forecast. The satellite data used are: Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS), Tropical Applications of Meteorology using SATellite (TAMSAT), Meteosat Second Generation (MSG/MPE) and The Global Forecast System (GFS). This resulted in a graph indicating the decadal rainfall, both from the past as well as 10 days predictions (see Figure 4).
The dry matter (DM) production map is calculated by multiplying the seasonal rainfall map with the grass cover- and an empirical rainfall effectivity map (Van Wijngaarden, 1985):
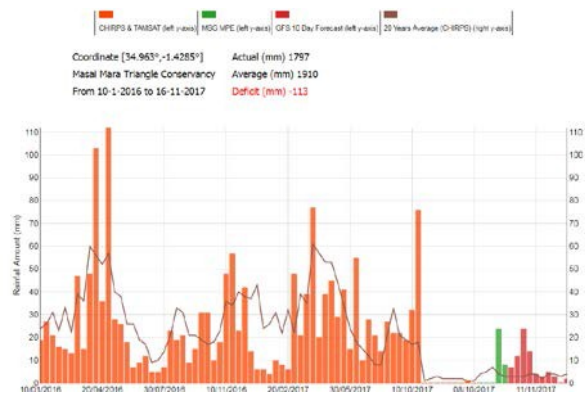

Figure 4. The perennial grass cover map

1DM production (gr/m2) = Rainfall (mm) * Grass cover (%) * Rainfall effectivity factor. The workflow calculating the forage productivity, compiled in ILWIS 4, is presented in Figure 5.
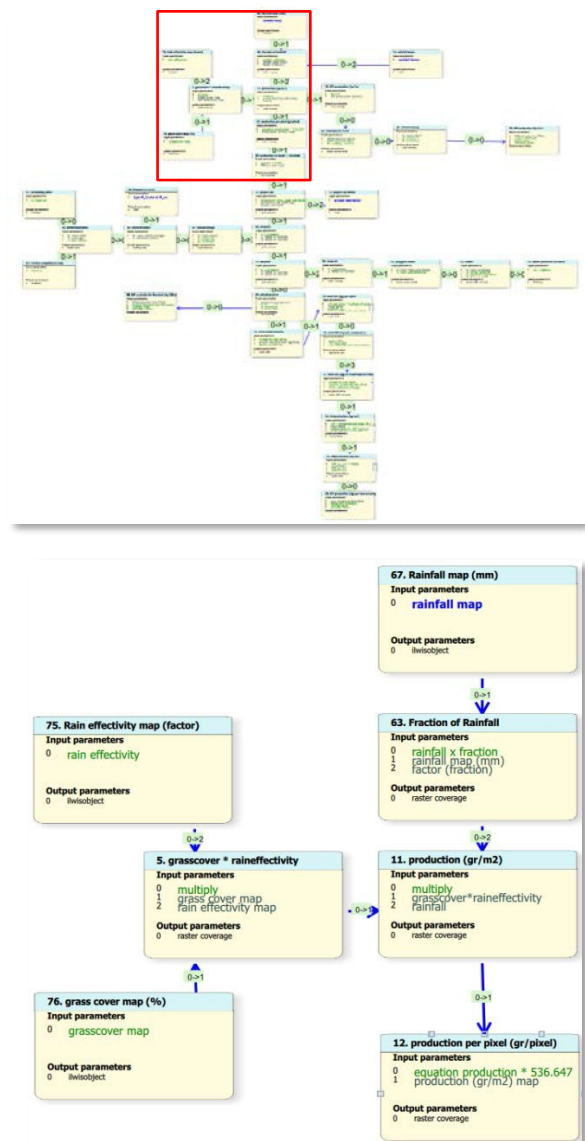

Figure 5. The ILWIS 4 production workflow (overall and an enlarged part)

3.5 Forage demand

For wildlife and livestock, the forage demand of each species is summed to obtain a map with the location and extent of the total wildlife and livestock forage demand, also at 23-meter resolution. For the model the herbivores monitored by DRSRS (Peden, 1984) are used to estimate the density and distribution of the herbivores. DRSRS carried out seasonal/yearly aerial census both on livestock and on wildlife to estimate the distribution and the number of large herbivores in the Maasai Mara ecosystem. The model is calculating the forage needs for grazers and mixed feeders only, leaving out the browsers. In order to determine the location and extent of the forage demand by livestock for the long dry season, the length of the long dry season is a multiplication of the average number of days per months (30.4), the dry matter intake value of a species and the distribution and density of a species. Figure 6 shows the full algorithm looping (inside the box) through the 19 herbivore species.
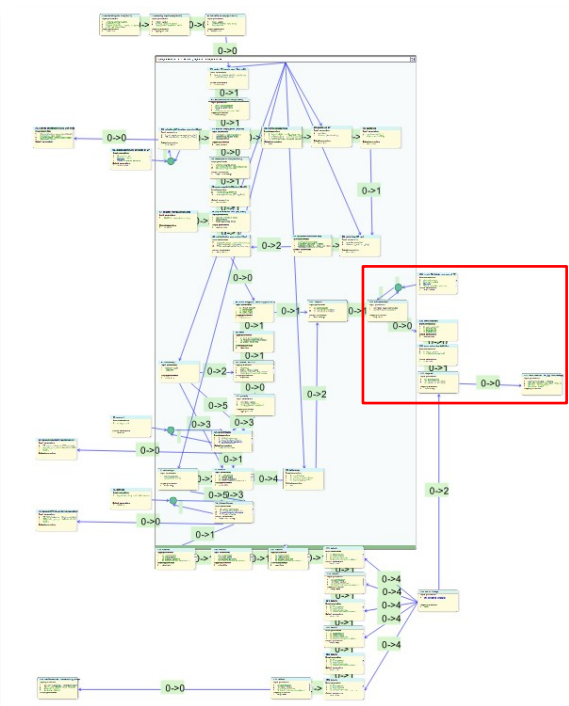
3.6 Carrying Capacity: assessment of the status of the rangelands

Although all kinds of methods and formulas to estimate carrying capacity for a certain rangeland have been used (Andere, 1981, Mwichabe, 1988, Pratt and Gwynne, 1977, Western and Praet, 1973), the estimation of the status of rangelands (e.g., over- or under grazing) in the model will be based on the "final" forage availability in relation to the number of livestock (demand for forage) in that area.

The actual forage availability map for livestock has been combined with the conservancies map, to calculate the redistribution of the livestock forage demand per conservancy over the actual forage availability within the conservancies, assuming that the status of the rangelands is an indicating for the distribution rate of the livestock forage demand within the conservancy.

The result is the carrying capacity map or the so called "conservancies management map", which is a map, with forage values ranging from - kg/ha to + kg/ha, indicated per 23m pixel, and the real pixel values in kg, indicating the amount of shortage or surplus of forage, according to the available forage (on a sustainable base) minus the demand for forage during the year (season) of calculation. For management purposes, the tabular (statistical) information of actual available forage, forage demand and the status of the rangelands (over- or understocked) during that period is calculated per conservancy.

3.7 Scenario running

In the web client users can register for a conservancy. After registering the option 'scenario running' becomes available for the user. With this option, the user is able to change stocking densities, select different historic or near future rainfall regimes (more dry or wet, longer or shorter) and other input variables, to analyse the effect of these different management options given given exogenous conditions (Figure 7). Each change will automatically be incorporated in the ILWIS 4 workflow and executed accordingly.

Figure 7. Example of one of the scenario options

3.8 Discussion

The implementation in ILWIS has a number of special and unique features. First it is implemented in the ILWIS client-server architecture that leads to conceptualization of conversion of workflows to a JSON format and tree representation of inputs and outputs as well as their metadata, for instance explain how outputs were generated from corresponding inputs. Consequently, workflows can be communicated to this (or other) web client(s) end users. But more importantly, web client users can change inputs and re-run models. Moreover, on the back-end, improvements can be made to models by modellers without the need to involve web developers. Combined with the simplicity of establishing workflows, this opens up possibilities for researchers with little programming experience to capture end-user behaviour as they use their workflow applications.

Figure 6. The ILWIS 4 demand workflow (overall and an enlarged part)

Second, the ILWIS software is freely available open source software that has a language interface with python, and is to be integrated with the latest version of QGIS through a plug in. This will make the features of workflow modelling available to a wide range of users.

The carrying capacity assessment workflows as well as the web client were tested in three workshops with managers of ten conservancies in the Maasai Mara. Although the analyses were quite complex and educational background of workshop participants varied from high school to MSc degree, all participants could clearly follow the logic of the algorithms by means of the graphical workflow representation, giving credibility to the analysis and leading to interest in implementation and use.

4. CONCLUSIONS AND OUTLOOK

At this stage the MARIS model is operational and can be accessed via: http://mara.rangelands.info.ke. (the local server in Kenya). After registration the user can download all data and also run his own rangelands management scenario's. On YouTube a video (https://www.youtube.com/watch?v=iGc22qVMrGg) is also available to demonstrate the functionality of the programme. Implementation of the information system (and the interactive modelling system) through the local counterpart organisations is the next activity. The use of all the (updatable) variables like maps, tables and factors for the ecosystem are described and tried out by the implementing agencies to check the efficiency, user-friendliness and reliability of the model in question.

The ILWIS workflow builder makes complex analyses communicable and its web clients create the possibility for online communication and inclusion in recurrent management processes. Other application areas could for instance be automated drought assessment and insurance payment schemes, or more generally, spatial risk assessment.

We believe that visual workflows form a good basis for the understanding and exchange of knowledge between humans when it comes to data processing and data handling procedures. We therefor plan to make visual workflows part of our next project, AfriAlliance, as part of its geo-information infrastructure. The AfriAlliance project (https://afrialliance.org/), aims to prepare Africa for future climate change challenges by creating the opportunity for African and European businesses, governments and individuals to work together in the areas of water innovation, research, policy, and capacity development. One of the innovative aspects of AfriAlliance is the multiple information source gathering approach. The inventory is based on a triple-sensor approach, looking into the capacity and efforts on human sensors (citizens, geo-crowdsourcing and communications), physical in-situ sensors (like weather station-based meteorological and ground-based water resource and hydrological networks such as stream gauging, groundwater, etc.) and space-based satellite sensors. Visual workflows will be disseminated to practitioners and researchers to enable them to understand and replicate the geoprocessing workflows deployed by others.

5. REFERENCES

Andere, D.K., 1981. Wildebeest Connochaetes taurinus (Burchell) and its food supply in Amboseli Basin. African Journal of Ecology, Vol: 19, pp: 239-250.

Bekure S., P.N. de Leeuw, B.E. Grandin and P.J.H. Neate, 1991. An analysis of the livestock production system of Maasai pastoralists in eastern Kajiado District, Kenya. ILCA Systems Study 4, Addis Ababa, 154 pp.

Braun, H.M.H. and D.N. Mungai, 1981. The Agroclimatic zones of Kenya's Arid and Semi-Arid Lands. Kenya Soil Survey, Report S676/DNM/JAO, Nairobi, 11 pp.

De Carvalho Diniz, F. (2016) Composition of semantically enabled geospatial web services. Enschede, University of Twente Faculty of Geo-Information and Earth Observation (ITC), 2016.

ILWIS Open Source GIS/RS software. Video introduction (2016) https://vimeo.com/user29453510/review/153355429/1c1a97df8 4

Lemmens, R., Farifteh, J., Piccinini, C., Retsios, B., Schouwenburg, M., Bonina, J. (2016). Visual and coded geoprocessing workflows based on ILWIS and Python. Journal of GeoPython, 1(1), 21–24. Retrieved from http://www.geopython.net/pub/Journal_of_GeoPython_1_2016. pdf

Mwichabe, E.S., 1988. The grasslands of the Loita-Aitong-Mara plains (Narok District, Kenya). Their composition, distribution, ecology and productivity. MSc Thesis ITC, Enschede, 103 pp.

Norton-Griffiths, M., 1977. Aspects of climate of Kajiado District. Wildlife Conservation and Management Department, Ministry of Tourism and Wildlife and UNDP/FAO Wildlife Management Project, Document no. 13, KEN/71/526, 62 pp.

Peden, D.G., 1984. Livestock and Wildlife Population Inventories by District in Kenya 1977 - 1983. Technical Report No. 102, KREMU, Ministry of Planning and National Development, Kenya, 46 pp.

Pratt, D.J. ad M.D. Gwynne, 1977. Rangeland management and ecology in East Africa. Hodder and Stoughton, London.

Touber, L., 1983. Soils and vegetation of the Amboseli-Kibwezi area. Kenya Soil Survey, Reconnaissance Soil Survey Report no. R6, Nairobi, 270 pp.

Western, D., 1982. Amboseli National park: Enlisting landowners to conserve migratory wildlife. Ambio, vol. XI, no.5, pp: 302-308.

Western, D. and V. Praet, 1973. Cyclic changes in the habitat and climate of an East African ecosystem. Nature London, pp: 104-106.

Wijngaarden, W. van, 1985. Elephants, trees, grass, grazers; relationship between climate, soil, vegetation and large herbivores in a semi-arid savanna ecosystem (Tsavo, Kenya). PhD thesis, Wageningen, 159 pp.