

Schmidt, M., Thiombiano, A., Dressler, S., Hahn-Hadjali, K., Guinko, S. & Zizka, G. (2010). Phytodiversity data — strengths and weaknesses, a comparison of collection and relevé data from Burkina Faso. In: X. van der Burgt, J. van der Maesen & J.-M. Onana (eds), *Systematics and Conservation of African Plants*, pp. 829–837. Royal Botanic Gardens, Kew.

PHYTODIVERSITY DATA — STRENGTHS AND WEAKNESSES, A COMPARISON OF COLLECTION AND RELEVÉ DATA FROM BURKINA FASO

MARCO SCHMIDT¹, ADJIMA THIOMBIANO², STEFAN DRESSLER¹,
KAREN HAHN-HADJALI³, SITA GUINKO² & GEORG ZIZKA¹

¹Forschungsinstitut Senckenberg, Senckenberganlage 25, 60325 Frankfurt, Germany; mschmidt@senckenberg.de

²Université de Ouagadougou, UFR/SVT, B.P. 7021, Ouagadougou 03, Burkina Faso;

³Institut für Ökologie, Evolution und Diversität, J.W.Goethe-Universität, Siesmayerstr. 70, 60323 Frankfurt am Main, Germany

Abstract

In the course of the international BIOTA (**B**iodiversity **M**onitoring **T**ransect **A**nalysis) West Africa project we have built up a large specimen-based database (over 20,000 specimens) as well as one for observation data for Burkina Faso (about 100.000 records from 4,596 relevés). These data are analysed to pinpoint advantages and disadvantages of both sets, comparing representation of rare vs. common species and selected families. Our results confirm biases for both data types and give evidence for using a combination of both for modelling approaches.

Résumé

Données de phytodiversité — ses forces et ses faiblesses, une comparaison des données de collection et de relevé de Burkina Faso. Dans le cadre du projet international BIOTA (**B**iodiversity **M**onitoring **T**ransect **A**nalysis) Afrique de l'Ouest, nous avons mis en place une base de données importante aussi bien à partir des collections d'herbier (plus de 20.000 échantillons) que des données d'observations de terrain pour le Burkina Faso (soit 100,000 entrées correspondant à 4.596 relevés). Ces données sont analysées pour mettre en évidence les avantages et inconvénients de chaque type de données, en comparant la représentation d'espèces rares ou communs et de familles sélectionnées. Nos résultats confirment le caractère biaisé de chacun type de données et la nécessité d'une combinaison des deux types pour des approches de modélisation.

Key words: Biodiversity databases, BIOTA, observation data, relevés, specimens.

1 Introduction

One of the most pressing issues of biological research is to document, explain and project biodiversity changes under increasing human impact and climate change. An essential drawback for this task is the lack of appropriate data, especially from the tropics.

Digitisation of specimen data has become an essential issue for most natural history collections; efforts in this field are increasing and bundled in international initiatives

like GBIF (see <http://www.gbif.org>). Such data can be linked with environmental layers to analyse distribution, macro-ecology, and phytogeography (e.g. Wittig *et al.*, 2004; Thiombiano *et al.*, 2006), model potential distributions (e.g. Chatelain *et al.*, 2001; Wieringa & Poorter, 2004), and use these modelled distributions to calculate maps of phytodiversity (e.g. Schmidt *et al.*, 2005; König *et al.*, 2006). The scale of these studies extends from local to regional, continental and global scale (e.g. Lovett *et al.*, 2000; Küper *et al.*, 2006). The quality of the results of such modelling approaches depends principally on the underlying data.

This leads to the question: which type of data should be used for the modelling approaches? In the course of our longstanding research in West Africa, we accumulated comprehensive sets of observation and specimen data, especially for Burkina Faso. As a model case, we used these data to assess the (dis)advantages of both datasets.

2 Data and Databases

In the course of our long-term studies on phytodiversity and its dynamics in dry Western Africa, two types of comprehensive datasets were accumulated: (a) specimen based data and (b) observation data.

The collections from Burkina Faso in the herbaria of Ouagadougou (OUA) and Frankfurt (FR) have been digitised, using a specimen database based on Microsoft Access and the programs BRAHMS (<http://herbaria.plants.ox.ac.uk/BOL/>) and recently SeSam (<http://sesam.senckenberg.de/>). The latter is used in the Senckenberg Research Institute for different kinds of scientific collections and accessible via the internet. At present, the databases include data from over 20,000 specimens (OUA: 9,273; FR: 11,167). The specimen data from FR are provided online via SeSam which contributes to GBIF. Nevertheless, due to ongoing analyses and research projects, part of the information (collecting locality) in each dataset is not yet accessible.

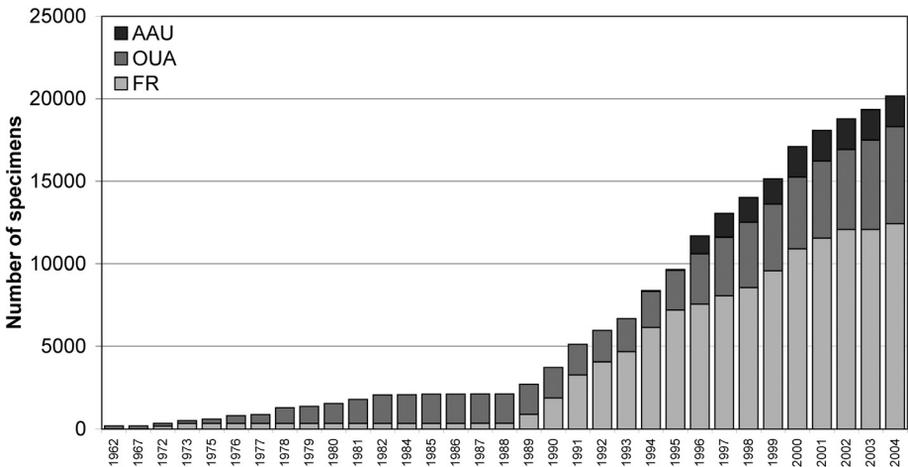


FIG. 1. Collection data for Burkina Faso: Cumulative number of herbarium specimens for the period 1962–2004. The considerable increase of collections from the 1990s on, is principally due to the research projects “SFB 268 — Cultural development and language history in the natural habitat of the West African savanna” and BIOTA.

The second dataset comprises field observation data, usually documented in the form of phytosociological relevés or species lists. Important early sources are the studies of Guinko (1984). From the 1990s on, the interdisciplinary research projects “Cultural development and language history in the natural habitat of the West African savanna” (SFB 268) and BIOTA (**B**iodiversity **M**onitoring **T**ransect **A**nalysis; started 2000) produced a wealth of observation data from Burkina Faso and also an increased number of herbarium specimens (Fig. 1). The observation dataset includes relevés from the BIOTA and SFB 268 projects (Hahn, 1996; Kéré, 1996; Küppers, 1996; Böhm, 1998; Denschlag, 1998; Ataholo, 2001; Sieglstetter, 2002; Müller, 2003; Krohmer, 2004; Schmidt, 2006). The data were pooled in a specially designed database (“VegDa”). About 100,000 records (from 4,596 relevés) are by now included in VegDa (see below).

For our diversity analyses (e.g. Schmidt, 2006: 41–49; Schmidt *et al.*, this volume), specimen and observation data were combined and specimen data added from the Herbarium Jutlandicum (AAU), published on its website (http://herb42.bio.au.dk/aau_herb/default.php). These extensive data for Burkina Faso allow also a comparative analysis of the two sets. To detect biases, we tested them for documentation of rare and common species, selected families, and spatial pattern of records.

2.1 VegDa — a database for digitising observation data

To store the observation data from our research projects, the first author developed the database VegDa, based on Microsoft Access. In the beginning used only for digitising our own observation data, its main use has now become the compilation of existing, hitherto scattered vegetation data from theses and reports. It is possible to enter different types of observations, ranging from simple inventories to relevés including abundance or frequency of species, such as according to Braun-Blanquet (1964) and BIOTA standard (Schmiedel & Jürgens, 2005) relevés.

VegDa is a relational database with several tables, the most important ones being (a) a species list including information on life form, distribution, photosynthetic type, (b) a series of tables for the taxonomic hierarchy, (c) a synecological table linking species to relevés, including stratum and cover/frequency and finally (d) a relevé table with position and surface area of the plots, vegetation structure, and information on land use, soil type, etc.

VegDa offers some analytic functions including the calculation of the biodiversity indices Shannon-Weaver, Evenness and Simpson. Export functions are also available, such as a routine producing a csv file which has no restrictions in the number of rows and columns and can be used in programs like CAP (Pisces, 2002) or SPSS (<http://www.spss.com>).

VegDa is continuously being upgraded; currently we are aiming at inclusion of dendrometrical data, an import/export function and an export of data to a Google Earth kml-file (<http://code.google.com/apis/kml/documentation/index.html>) for easy geographical visualisation. A desideratum in the long run is an online version. The program is distributed free of charge by the first author.

2.2 Comparison of collection and observation data

Some differences between specimen and observation data are obvious and account for their preferential uses: specimens are stored in collections and are available for revision. Sometimes they are determined by more than one scientist and thus accumulate taxonomic knowledge. Specimen data are regarded to be determined more reliably (at least they are available for verification/inspection) and are preferentially

used in analyses of distribution and in modelling approaches. However, the broader the data basis used for analyses, the less obvious are these advantages. With the increase of digitised collections, large data sets are available and used for analyses (e.g. all specimens from an area or country). It has to be assumed that in these cases, a considerable number of misidentifications is included. Such “errors” can vary considerably between collections, depending on the intensity of scientific curation and taxonomic expertise for the relevant systematic groups or geographic area. We have no study at hand that provides figures for the average amount of misidentifications in herbarium collections.

Such studies have been performed for observation data by Scott & Hallam (2003), Archaux *et al.* (2006), and Archaux *et al.* (2007). The former report a misidentification rate in vegetation monitoring of 4.3%. Archaux *et al.* (2006) also showed that sampling time and species richness of the plots had an influence on the amount of identification errors.

Because we have been collecting and analysing specimen as well as observation data, we tested the assumed bias of both data types:

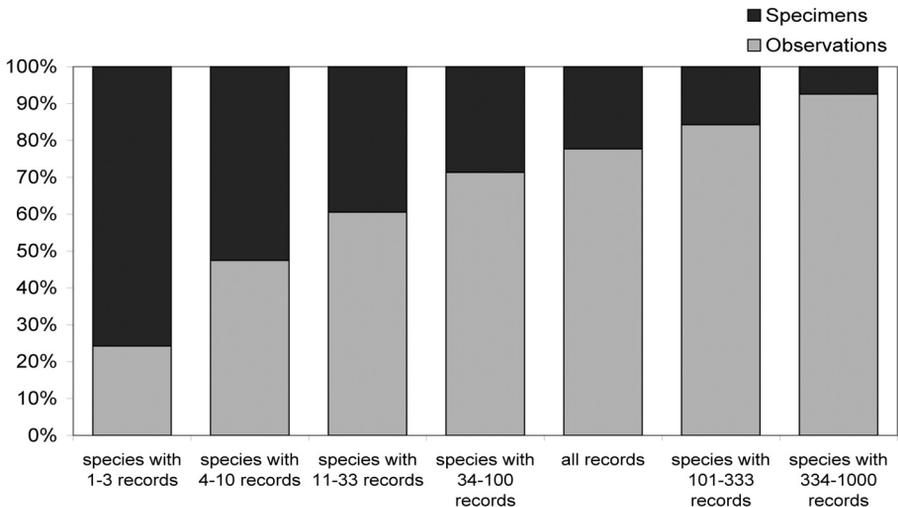


FIG. 2. Representation of rare and frequent species in collection and observation data for Burkina Faso, expressed as % of all records within a frequency class (defined by number of records per species).

When collecting in the field, common plants are rarely considered because time, manpower and place in the press are limited. The same holds true for groups that are difficult to collect, e.g. palms. We therefore expect such groups to be under-represented in herbaria and also in “Florals”, which principally rely on herbarium collections. Lebrun *et al.* (1991: 241) note in their plant catalogue for Burkina Faso for the common palm *Borassus aethiopicum*: “Le plus grand des palmiers africains (rônier) existe au Burkina Faso mais nous n’en connaissons aucun échantillon d’herbier.” And indeed, recently a second *Borassus* species from Burkina Faso was described (Bayton *et al.*, 2006) based on only few herbarium collections and mainly detected by careful observation of living specimens. On the contrary, observation data can be assumed to represent primarily common species and not to be correlated with the technical problems of

collecting. Fig. 2 confirms these assumptions, showing that rare species are very well documented in collections but heavily under-represented in observations. The opposite holds true for widespread species. Looking at the records all together, the observations account for nearly 80% of all records.

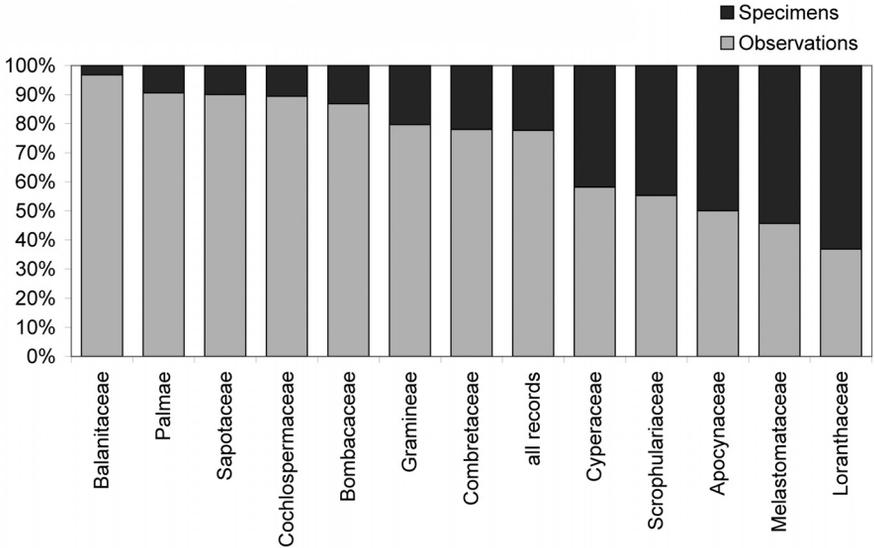


FIG. 3. Proportion of collection and observation data for selected families (expressed as percentage of total number of records), compared to the proportion among all records.

Looking at the representation of selected families in collections and observations, those comprising widespread, almost exclusively woody plants are under-represented in the collections (e.g. Palmae, Balanitaceae, Sapotaceae, Cochlospermaceae, Bombacaceae, Fig. 3).

Families over-represented among the collections are predominantly herbaceous and thus more easily herborised (e.g. Cyperaceae) or are characterised by a high percentage of relatively rare species (e.g. Apocynaceae, Melastomataceae). The Scrophulariaceae s.l. of the study area are predominantly small and inconspicuous plants and possibly are easily overlooked.

Gramineae and Combretaceae appear to be represented among collections and observations more or less on average, in spite of their different life form and habit. The Gramineae are dominant, the Combretaceae common elements of the West African savanna vegetation, and representatives of both can be easily herborised. Moreover, Combretaceae have been extensively collected over the years in the course of a taxonomic and ecological study (Thiombiano *et al.*, 2006).

An analysis of the geographical origin of the data revealed also differences between the two data sets. We chose 56 grid cells of $1^\circ \times 1^\circ$ to cover a rectangle ($8^\circ\text{--}16^\circ\text{N}$, $7^\circ\text{W--}3^\circ\text{E}$) around the territory of Burkina Faso and calculated the records for each of the cells as a proportion of the total records. In this case, the specimen based records are compared not to observation data for a single species but to relevés.

Fig. 4 illustrates different patterns for our two datasets. The observation data are obviously more clustered than the specimen data with c. 44% of the records from one grid cell! Actually, such data are recorded for only for 15 grid cells.

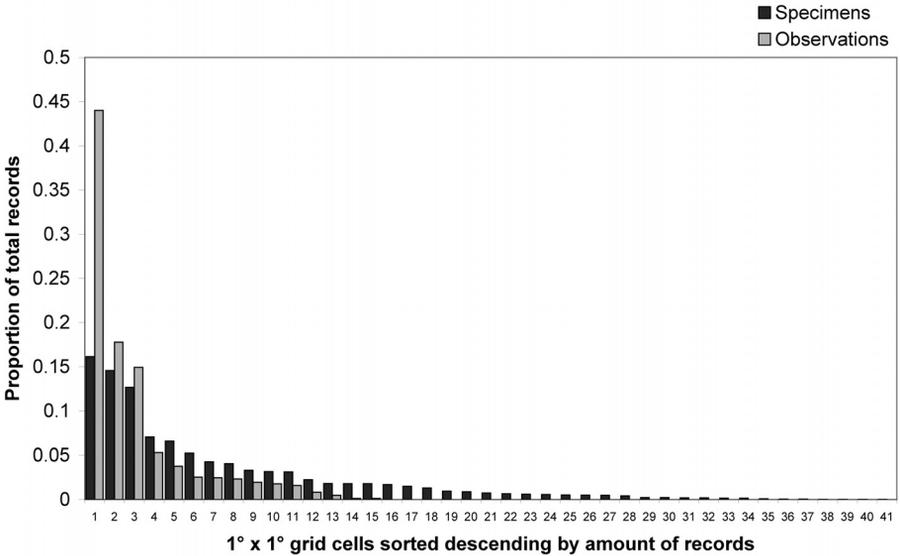


FIG. 4. Geographic distribution of records, expressed as the number of records per $1^\circ \times 1^\circ$ grid cell. Specimen based records are compared to relevés.

The specimen data display a more even distribution. For 41 grid cells such data are recorded, the best documented cell contributing only c. 16% to the data.

These differences mirror different approaches and design of research projects and related field work. Most of the studies on the vegetation have been rather local and produced a considerable number of relevés in only few grid cells (e.g. Böhm, 1998; Müller, 2003). For the Burkina Faso data it is obvious that the vegetation studies providing the observation data have been performed within a limited geographical range, covering only 27% of the grid cells. Collecting activities as the base of specimen data have been more widespread and included 73% of the grid imposed on the country's area.

3 Conclusions and perspectives

The investigated datasets for Burkina Faso display biases of specimen and observation data.

Specimen data have rare species well documented but frequent species are under-represented. Also insufficiently represented are groups like palms which are cumbersome to collect. Observation data document frequent species well but have rare and inconspicuous species under-represented. At least for our datasets from Burkina Faso, the observation data are geographically more clustered, while specimen data are more evenly distributed and cover a considerably higher part of the area. As far as the weaknesses in the documentation of rare and frequent species are concerned, the two data sets are complementary. Our results are in favour of a combination of both data types to minimise biases, in spite of the fact that misidentifications in observation data cannot be revised. Misidentifications occur to a certain extent in observation data, but also in specimen data. With the rapid increase of data available for distribution and

diversity analysis via the internet and their use for local, regional and continental analyses, revising the determination of specimens — though possible — is no longer realistic and feasible. A certain amount of misidentifications will therefore have to be accepted in analyses based on large datasets.

Although observation and collection data are complementary in some respect, the wealth of observation data must not be neglected. We believe that observation data have been scientifically under-exploited until now but will gain more importance in the future with the increasing pressure to address the issues of climate change and its consequences for biodiversity.

A flaw hampering the use of observation data is the often rapidly decreasing availability. Most of these data are collected in the course of scientific theses or reports. Only part of the data — if any — become published in international journals, while the original datasets (typically tables of relevés) are prone to be stored in archives or libraries and become difficult to trace and access.

The effective use of observation data requires adequate software tools which easily allow combination with specimen data. The database VegDa proved to be adequate for this purpose. Still the problem remains, how digitising and easy-to-access storage of observation data can be achieved on a broader scale. A now renowned example from Germany is VegetWeb (<http://www.floraweb.de/vegetweb>), a similar database from the United States of America is VegBank (<http://www.vegbank.org>).

We recommend that the digitisation of observation data collected in the course of scientific studies is made obligatory. “Data nodes” should be set up, where these data can be stored permanently. For Burkina Faso, such a “data node” has been identified at the Herbarium of the University of Ouagadougou (OUA).

Acknowledgements

We thank the German Ministry for Education and Research (BMBF) for funding the BIOTA project, and the contributors to our databases.

References

- Archaux, F., Gosselin, F., Berges, L. & Chevalier, R. (2006). Effects of sampling time, species richness and observer on the exhaustiveness of plant censuses. *J. Veg. Sci.* 17 (3): 299–306.
- Archaux, F., Berges, L. & Chevalier, R. (2007). Are plant censuses carried out on small quadrats more reliable than on larger ones? *Pl. Ecol.* 188 (2): 179–190.
- Ataholo, M. (2001). *Pflanzensoziologische Untersuchungen der Segetalarten in der Sudanzone Westafrikas*. Dissertation, J. W. Goethe-Universität, Frankfurt am Main.
- Bayton, R. P., Ouédraogo, A. & Guinko, S. (2006). The genus *Borassus* (Arecaceae) in West Africa, with a description of a new species from Burkina Faso. *Bot. J. Linn. Soc.* 150 (4): 419–428.
- Böhm, M. (1998). *Dorfvegetation in Burkina Faso*. Diploma thesis, J. W. Goethe-Universität, Frankfurt am Main.
- Braun-Blanquet, J. (1964). *Pflanzensoziologie – Grundzüge der Vegetationskunde*, 3. Aufl. Springer, Wien.
- Chatelain, C., Gautier, L. & Spichiger, R. (2001). Application du SIG IVOIRE à la distribution potentielle des espèces en fonction des facteurs écologiques. *Syst. Geogr. Pl.* 71: 313–326.

- Denschlag, J. (1998). *Ethnobotanische und pflanzensoziologische Untersuchungen der Gehölzvegetation bei den FulBe im Südosten von Burkina Faso (Westafrika)*. Diploma thesis J. W. Goethe-Universität, Frankfurt am Main.
- Guinko, S. (1984). *Végétation de la Haute Volta*. Dissertation, Université de Bordeaux III.
- Hahn, K. (1996). *Die Pflanzengesellschaften der Savannen im Südosten Burkina Fasos (Westafrika). Ihre Beeinflussung durch den Menschen und die naturräumlichen Gegebenheiten*. Dissertation, J. W. Goethe-Universität, Frankfurt am Main.
- Kéré, U. (1996). *Die Dorf- und Savannenvegetation in der Region Tenkodogo (Burkina Faso)*. Dissertation, J. W. Goethe-Universität, Frankfurt am Main.
- König, K., Schmidt, M. & Müller, J. (2006). Delineating patterns of plant diversity in the Sahel zone of Burkina Faso: Modelling of environmental envelopes with high resolution remote sensing data. In: A. Röder & J. Hill (eds), *Proceedings of the first International Conference on Remote Sensing and Geoinformation Processing in the Assessment and Monitoring of Land Degradation and Desertification (RGLDD-05)* Trier, Germany, September 7–9 2005: pp. 317–323. Trier.
- Krohmer, J. (2004). *Umweltwahrnehmung und -klassifikation bei Fulbegruppen in verschiedenen Naturräumen Burkina Fasos und Benins*. Dissertation, J. W. Goethe-Universität, Frankfurt am Main.
- Küper, W., Sommer, J. H., Lovett, J. C. & Barthlott, W. (2006). Deficiency in African plant distribution data — missing pieces of the puzzle. *Bot. J. Linn. Soc.* 150 (3): 355–368.
- Küppers, K. (1996). *Die Vegetation der Chaine de Gobnangou*. Dissertation, J. W. Goethe-Universität, Frankfurt am Main.
- Lebrun, J. P., Toutain, B., Gaston, A. & Boudet, G. (1991). *Catalogue des plantes vasculaires du Burkina Faso*. éd. I.E.M.V.T., Maisons-Alfort.
- Lovett, J. C., Rudd, S., Taplin, J. & Frimodt-Moller, C. (2000). Patterns of plant diversity in Africa south of the Sahara and their implications for conservation management. *Biodivers. Conserv.* 9 (1): 37–46.
- Müller, J. V. (2003). *Zur Vegetationsökologie der Savannenlandschaften im Sahel Burkina Fasos*. Dissertation, J. W. Goethe-Universität, Frankfurt am Main.
- Pisces Conservation (2002). *Community Analysis Package 2.0*. Pisces Conservation Ltd, Lymington.
- Schmidt, M. (2006). *Pflanzenvielfalt in Burkina Faso — Analyse, Modellierung und Dokumentation*. Dissertation, J. W. Goethe-Universität, Frankfurt am Main.
- Schmidt, M., Kreft, H., Thiombiano, A. & Zizka, G. (2005). Herbarium collections and field data-based plant diversity maps for Burkina Faso. *Diversity & Distrib.* 11 (6): 509–516.
- Schmidt, M., Thiombiano, A., Ouédraogo, A., Hahn-Hadjali, K., Dressler, S. & Zizka, G. (2010). Assessment of the flora of Burkina Faso. In: X. van der Burgt, J. van der Maesen & J.-M. Onana (eds), *Systematics and Conservation of African Plants* (this volume). Royal Botanic Gardens, Kew.
- Schmiedel, U. & Jürgens, N. (2005). Biodiversity Observatories. A new standardised monitoring tool for biodiversity studies. In: M. Veste & C. Wissel (eds), *Beiträge zur Vegetationsökologie der Trockengebiete und Desertifikation*. UFZ-Bericht 01/2005: pp. 87–91.
- Scott, W. A. & Hallam, C. J. (2003). Assessing species misidentification rates through quality assurance of vegetation monitoring. *Pl. Ecol.* 165 (1): 101–115.
- Siegstetter, R. (2002). *Wie die Haare der Erde — Vegetationsökologische und soziokulturelle Untersuchungen zur Savannenvegetation der Südsudanzone Westafrikas und ihrer Nutzung und Wahrnehmung durch die ländliche Bevölkerung am Beispiel der Region Atakora im Nordwesten Benins*. Dissertation, J. W. Goethe-Universität, Frankfurt am Main.

- Thiombiano, A., Schmidt, M., Kreft, H. & Guinko, S. (2006). Influence du gradient climatique sur la distribution des espèces de Combretaceae au Burkina Faso (Afrique de l'Ouest). *Candollea* 61 (1): 189–213.
- Wieringa, J. J. & Poorter, L. (2004). Biodiversity hotspots in West Africa; patterns and causes. In: L. Poorter, F. Bongers, F. Kouame & W. D. Hawthorne (eds), *Biodiversity of West African forests*. An ecological atlas of woody plant species: pp. 61–72. CABI Publishing, Oxford.
- Wittig, R., Schmidt, M. & Thiombiano, A. (2004). Distribution maps of the species of the genus *Acacia* L. in Burkina Faso. *Etudes Fl. Vég. Burkina Faso* 8: 19–26.