

AMMA information system: an efficient cross-disciplinary tool and a legacy for forthcoming projects

Laurence Fleury, Jean-Luc Boichard, Guillaume Brissebrat, Sophie Cloché, Laurence Eymard, Laurence Mastrorillo, Oumarou Moulaye, Karim Ramage, Nicole Asencio, Jean Coppeaux, et al.

▶ **To cite this version:**

Laurence Fleury, Jean-Luc Boichard, Guillaume Brissebrat, Sophie Cloché, Laurence Eymard, et al.. AMMA information system: an efficient cross-disciplinary tool and a legacy for forthcoming projects. Atmospheric Science Letters, 2011, 12 (1), pp.149-154. 10.1002/asl.303 . hal-00612886

HAL Id: hal-00612886
https://hal.science/hal-00612886

Submitted on 2 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ATMOSPHERIC SCIENCE LETTERS
Atmos. Sci. Let. **12**: 149–154 (2011)
Published online 12 November 2010 in Wiley Online Library
(wileyonlinelibrary.com) **DOI:** 10.1002/asl.303

RMetS
Royal Meteorological Society

AMMA information system: an efficient cross-disciplinary tool and a legacy for forthcoming projects

Laurence Fleury,[1,2]* Jean-Luc Boichard,[1] Guillaume Brissebrat,[1] Sophie Cloché,[3] Laurence Eymard,[4] Laurence Mastrorillo,[1] Oumarou Moulaye,[5] Karim Ramage,[3] Nicole Asencio,[2] Jean Coppeaux,[2] Marie-Pierre Devic,[2] Florence Favot,[2] Karine Ginoux,[6] Jean-Philippe Lafore,[2] Jan Polcher,[7] Jean-Luc Redelsperger,[2] Odile Roussot[2] and Michel Tytéca[2]

[1]*OMP Data Centre, CNRS, Toulouse, France*
[2]*GAME, URA CNRS/Météo-France, Toulouse, France*
[3]*IPSL Data Centre, CNRS/IPSL, Palaiseau, France*
[4]*LOCEAN, CNRS/IPSL, Paris, France*
[5]*AGRHYMET Regional Centre, Niamey, Niger*
[6]*CNRS/IPSL, Paris, France*
[7]*LMD, CNRS/IPSL, Paris, France*

*Correspondence to:
Laurence Fleury, Service Données de l'OMP, UMS 831, 14 Avenue Edouard Belin, 31400 Toulouse, France.
E-mail:
Laurence.Fleury@obs-mip.fr

Received: 26 February 2010
Revised: 9 July 2010
Accepted: 7 September 2010

Abstract

In the framework of the African Monsoon Multidisciplinary Analyses (AMMA) programme, several tools have been developed in order to facilitate and speed up data and information exchange between researchers from different disciplines. The AMMA information system includes a multidisciplinary user-friendly distributed data management and distribution system, a reports and quick looks archive associated with a display website and scientific papers exchange systems. All the applications have been developed by several French institutions and fully duplicated in Niamey, Niger. Copyright © 2010 Royal Meteorological Society

Keywords: interdisciplinarity; distributed database; bibliography database; campaign website

1. Introduction

The African Monsoon Multidisciplinary Analyses (AMMA) programme combines one of the largest research field experiments that have been launched in West Africa with the reinforcement of operational observation networks and the setting up of a centralized data management system gathering together all the data produced or used by the programme researchers. The database contains various dataset types (local observations, satellite products, model outputs, human dimension studies data…) either present or historical, either operational or research. The data management and distribution system has been designed to facilitate data exchange between researchers from different disciplines or using different methods. The AMMA database and associated online tools have been developed and are managed by two teams in France (IPSL Data Centre, Palaiseau and OMP Data Centre, Toulouse). Datasets are stored in either one of the Centres depending on their type, but all the datasets can be accessed through a single and user-friendly data request interface. The whole system has been duplicated and is also operational in the AGHRYMET Regional Centre (CRA), Niamey. The AMMA database, as described in Section 2, constitutes a multidisciplinary reference archive about the West African area and will be maintained and enriched by forthcoming projects.

The AMMA information system includes two other components detailed in Section 3. First, a quick looks display website has been developed to meet the 2006 intensive campaign needs and allows users to browse research and operational observations, atmospheric analyses and forecasts, satellite images, reports… The 2006 website is still online and can be used for a first selection of case studies. A derived version dedicated to the needs of the African Centre of Meteorological Application for Development (ACMAD), Niamey, is operational. Second, publication gathering systems have been developed in order to submit and share publications in early stages (draft, submitted…). They constitute a very useful scientific publications archive, particularly for researchers faced with access difficulties to scientific journals. The AMMA publications collaborative website has been duplicated in CRA, Niamey.

2. AMMA data management and distribution system

2.1. Requirements and choices

The AMMA database aims at:

- Storing any data produced or used by the AMMA scientific community, which means a great amount

Copyright © 2010 Royal Meteorological Society

and a wide variety of datasets: *in situ* measurements, satellite products, model outputs, human sciences studies data, value-added products... In order to take advantage of the participating data centres' expertise, the AMMA data are managed either by the IPSL Data Centre or by the OMP Data Centre – previously located at MEDIAS-France, Toulouse – depending on their type. Data type characterizes only the data structure and is not related to the corresponding discipline. Basically, 2D or 3D data are managed at the IPSL. This concerns satellite products, model outputs, maps... 1D data, corresponding to the local observational time series, are managed at the OMP. Data that are qualitative, textual or partly qualitative are managed at OMP. The AMMA data are distributed in two data centres but conform to similar metadata standards and can be accessed through a single web data request user interface. In the Niamey AMMA database, all the datasets are managed by the CRA.

- Rescuing historical data over West Africa. *In situ* data may come from operational networks (radiosondes, meteorological synoptic observations, rain gauges networks...) or previous scientific experiments (COPT81, HAPEX-Sahel, IMPETUS-Benin). Satellite data from past satellite instruments (AVHRR, MODIS, POLDER...) have been collected in the satellite database and outputs coming from widely used reanalyses (ECMWF ERA40, NCEP...) in the model database.
- Enhancing data exchange between researchers from different disciplines and creating collaborations between specialist and non-specialist users. This specific mission had strong implications in the AMMA database choices, in terms of data documentation, data processing and formats. Metadata, i.e. information needed to characterize the data, are compatible with international standards (ISO19115 and DIF, Directory Interchange Format, http://gcmd.nasa.gov/User/difguide/) and systematically documented. Whenever possible, more documentation about data or data processing is made available. Local data are homogenized (parameter names, units...) and provided to the users in standard formats (plain text files using the Comma Separated Value format and NetCDF format files). The output files contain normalized headings with metadata information and present the same columns or blocks order for similar data. Concerning satellite and model data, easy-to-use products have been reprocessed. All the products are reprojected over regular latitude–longitude grids and converted into NetCDF files that contain a metadata header and data arrays.
- Allowing data retrieval following various criteria (location, time period, parameter...). To meet this requirement, several relational databases have been developed. The local observation database is fully relational, while the satellite/model database is a tree-structure archive of files, associated with a metadata relational database. The AMMA data request user interface takes full advantage of the data or metadata base relational structure.
- Providing different data processing levels depending on the user's needs. *In situ* observations are available both in the original format provided by the data producer and in the standard formats once extracted from the relational database management system. It is the data producer's responsibility to define which data levels he/she will provide to the database and which should be distributed to the users, to check the quality of the data and to provide documentation about data processing, if any. All the provided versions of the datasets are stored by the database team (from raw or first provided version to the most controlled) but only the most reliable can be accessed in homogenized formats. The metadata and data bases include flags for data quality and parameter status (e.g. observed *vs* calculated). Satellite geophysical products (level 2 or more), operational or specifically derived for AMMA, are available online. But level 2 data with an original projection not corresponding to a regular grid or level 1 data (radiances, brightness temperatures...) that can be of interest to expert users are available offline. The original satellite or model data centre address is part of the metadata.
- Providing AMMA participants the best and easiest access to data, but taking care of the data producers' intellectual property rights. This is particularly crucial in the framework of a wide multidisciplinary and international programme, where participants do not know each other and may have different data sharing cultures. Moreover, the AMMA programme has been funded by different national and pan-national projects and some data have been reserved for a subset of participants. Database users must register and sign online for the AMMA data and publication policy prior to any data access. The most important rules are the prohibition of any data redistribution and any data use for commercial applications. Database users are expected to contact the data producer in order to propose collaboration. To ensure this happens, any access to a dataset automatically triggers an email to the dataset producer so that he/she is aware of who intends to use the data, and an email to the dataset user reminding him/her to contact the dataset producer. In case of a publication by a data user, co-signature with the data producer, or at least an acknowledgment, is imposed, as well as an acknowledgment of the agencies which have built the AMMA programme. A copy of the publication should be sent to the Project Office. The AMMA database user management complies with the data policies. In particular, the registered users' name, mail, password and access type are stored in a Lightweight Directory Access Protocol (LDAP) directory, that is synchronized between the OMP, IPSL and CRA so that users can access the authorized datasets in any database. Authentication by the

Copyright © 2010 Royal Meteorological Society

LDAP directory is required when a user accesses the data extraction interface or downloads a data request output.

- Being duplicated in the CRA, Niamey. To facilitate this task, as well as for maintenance and future evolutions of the system, the AMMA database and associated online tools are based on robust, standard and free languages and software. The technical choices allow applications handling by any competent Linux system engineer.
- Insuring long-term preservation for the AMMA data. At present, the metadata and data are managed with state-of-the-art technologies and back-up procedure is insured by the existence of three copies (two in France, one in Niger). One of the best ways to preserve data is to keep the data system active and funded, i.e. maintaining a new data flux and serving a user community. In the framework of the AMMA second phase (2010–2020; http://www.amma-international.org/), the AMMA database is associated with many current and forthcoming research projects. In France, its activities will be supported by the scientific agencies until 2020, but funding to sustain the CRA database is still to be secured. Long-term data preservation is every data centre's major concern and needs specific and costly handling. A strategy for the AMMA long-term data preservation has to be decided in the coming years together with the African institutions participating in AMMA.

2.2. Achieved system

The AMMA database is composed of five elements: four different databases and a unique data request user interface. At present, the total volume, including original and processed datasets, is as large as 21 terabytes, one of them being devoted to local observations, ten to satellite products and ten to model outputs. Detailed tables summarizing the AMMA database content are regularly sent by email to all the users and can be downloaded in the user interface documentation corner. The AMMA database has a count of 550 registered users. About 100 data requests are processed every month by the data request user interface.

The local observations database is stored in the PostgreSQL relational database management system. All the 1D geo-localized and dated observations that are numerical values or that can be converted into numerical value with equivalence tables (e.g. codes for crop types), whatever the discipline is (meteorology, hydrology, oceanography, ecology, epidemiology...), are standardized and converted into the PostgreSQL format. At present, the local observations database contains about 200 datasets. They are AMMA Long-term, Enhanced and Special Observing Periods (LOP, EOP and SOP) ground-based, ship and airborne observations, operational data – meteorological data sent to the Global Transmission System (GTS), radiosondes high resolution profiles... – , long-term monitoring data from different networks – AERONET, CATCH, IDAF, PIRATA ... – , historical data... About 90% of the AMMA data have been recovered.

Gridded data are reprojected over regular latitude–longitude grids and converted into the NetCDF format. The products are available in six regions associated with predefined resolutions. Temporal and spatial resolutions for each product are chosen according to the sensor resolution and geophysical parameter variability. The products are sampled or interpolated to fit the predefined resolutions. They are stored in a tree-structure archive of files. Associated metadata are stored in a MySQL relational database. Satellite database distributes 60 operational and research products. They are derived from past and present satellite instruments (AVHRR, Meteosat, MODIS, POLDER, SEVIRI, SSM/I, TMI...) or have been developed within AMMA (Estimation of Precipitation by SATellite Second Generation), EPSAT-SG_AMSR-E surface soil moisture estimation (Kergoat *et al*., 2011). Model database distributes ten sets of outputs from operational meteorological and ocean models, available in different regions associated with predefined time and space resolutions. Distributed outputs include NCEP and ECMWF ERA40 reanalyses, ECMWF, three Météo-France models and MERCATOR ocean analyses during the AMMA programme's period. A special ECMWF AMMA period reanalysis (Bock *et al*., 2011) and a set of the AMMA Land Surface Model Intercomparison Project (ALMIP, Ruti *et al*., 2011) outputs are available too.

Field data that are not numerical values, or that cannot be converted into numerical value using equivalence tables, require the development of specific databases. Up to now, only one case has been achieved. The AMMA database team developed a MySQL relational database and a data capture interface for a socio-economic questionnaire about household adaptation strategies (Mertz *et al*., 2011). Consultation pages and predefined database requests have also been set up. About 1350 questionnaires have been captured and are available from the data request user interface.

A complete metadata base has been designed and developed. It gathers all dataset information, regardless of the data type in the PostgreSQL relational database management system. The catalogue home page displays the data information sorted following different criteria and can be investigated through an online search tool.

A single data request user interface allows users to register, identify, access metadata and request data. Users can either build complex cross-criteria data extraction requests on the data inserted/converted in the different databases or request data in their original format using a files-ordering interface. The administrators' corner allows to sort out registration requests and to enrich the LDAP user directory. More technical information can be found in Fleury *et al*. (2010).

Copyright © 2010 Royal Meteorological Society

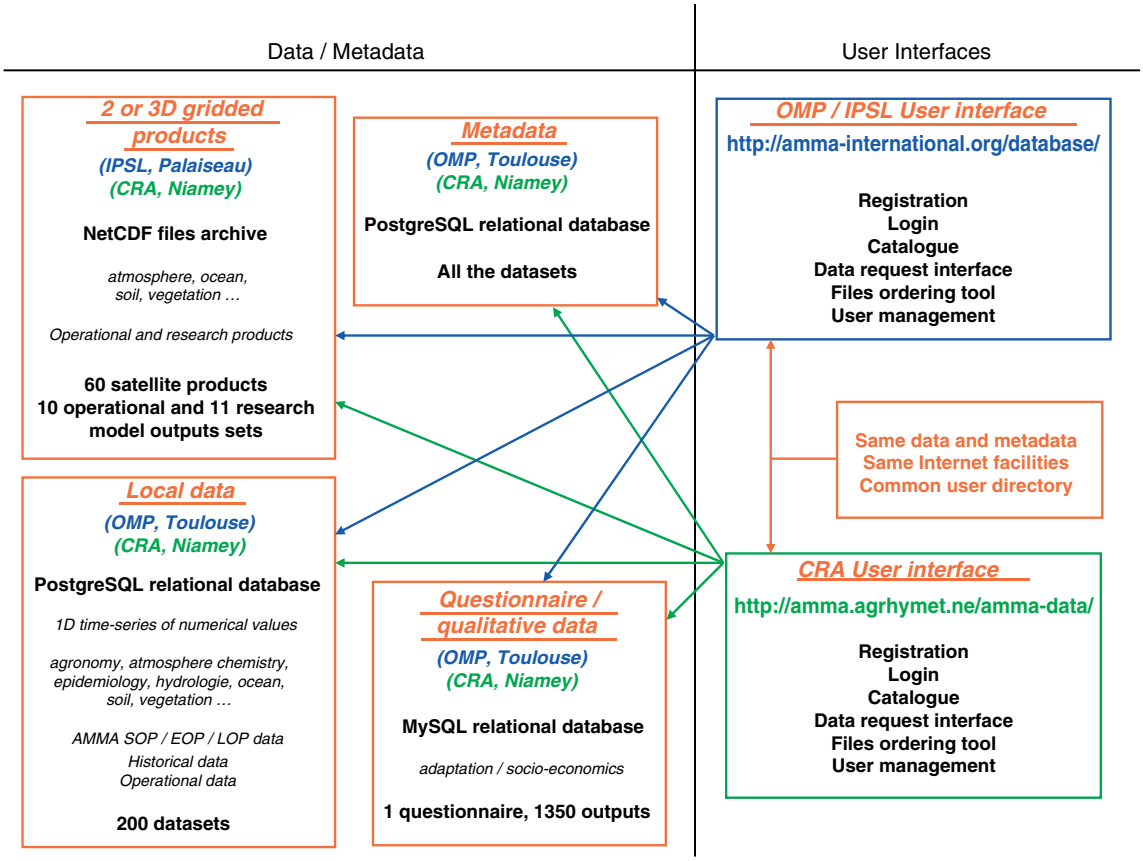


Figure 1. General structure of the twin AMMA databases.

Two copies of the complete system exist. One copy is distributed and managed together by the IPSL and OMP data centres (http://amma-international.org/database); the other is centralized in the CRA (http://amma.agrhymet.ne/amma-data/). The two systems are independent of each other. Data, user list and programs are automatically synchronized by scripts. Up to now, updates are occurring first in the French system and the synchronization scripts are one-way, from the IPSL/OMP to the CRA. When necessary, the reversed synchronization scripts will be set up to duplicate the CRA updates in the IPSL/OMP system. Figure 1 summarizes the twin systems structure.

3. Other components of the AMMA information system

3.1. AMMA operational centre website

The AMMA Operational Forecasting Centre (AOC-F) was set up at the ACMAD in Niamey to meet the operational needs of the airborne and ground-based observational teams during the 2006 SOP, from 1 June to 30 September. AOC-F main responsibility was the provision of weather forecasts in order to plan aircraft operations, and of nowcasts to guide the aircraft missions in real time. The AOC-F provided a framework for sharing knowledge and experiences between West African forecasters, scientists and modellers. Participating West African forecasters are expected to transfer some of the experiences and expertise gained to their national meteorological centres and to constitute a pool of resource persons who can be available for capacity development within the subregion. A Forecasters' Handbook for West Africa is under preparation. It will propose common forecasting techniques and procedures across the subregion, and include SOP real-time case studies.

The AOC-F work was based on the intensive use of the AOC website (http://aoc.amma-international.org) that was developed and tested during the 2006 dry SOP (January–February 2006). The application displays on real-time quick looks of a great amount of products. The quick looks were automatically fetched on various meteorological or data operational centres or created for AOC-F purposes. The displayed products were operational observations (radiosonde profiles, satellite images…), numerical weather products issued from different global or regional forecasts models, and research forecasts or diagnostics. The AMMA observations, either ground-based or airborne, were also displayed together with field or flight reports. AOC-F outputs – briefings reports twice a day, daily bulletins and West African Synthetic Analyses and Forecasts at different times – were also posted on the AMMA website. During the last phase of the field campaign – SOP3 in September 2006 – a set of mesoscale weather forecasts were computed by the AMMA Centre for Operations in DAKar (ACODAK,

Copyright © 2010 Royal Meteorological Society

set up in LPAOSF, Cheikh Anta Diop University) and posted on the AMMA website.

The display website was developed in PhP script. It was installed on servers in Toulouse and Niamey during the whole SOP and in Dakar during SOP3. The servers were synchronized with two-way scripts using compressed rsync protocol. Dedicated equipment and internet subscription were set up in order to ensure a constant data flow rate. New quick looks and reports were posted either in Toulouse, Niamey or Dakar. Website supply was made by automatic scripts, and ftp transfer by the AOC-F team and individual scientists on the field. Products were stored in a structured archive. The file names followed a definite nomenclature and included information (name, type, date, hour...) to be read by the display scripts. The site was operationally maintained and enriched for new display pages by persons in charge in Toulouse and Paris.

The AOC-F website is still online, and constitutes a testimonial view on the AMMA SOP campaign and a preliminary investigation tool for researchers. Browsing through the quick looks may serve to determine interesting situations to be studied. After the SOP, Niamey servers stayed in the ACMAD and website supply has been maintained. A derived version is now operational and is focused on ACMAD needs in the Severe Weather Forecast Demonstration Project (SWFDP, http://aocafrique.amma-international.org/) framework.

3.2. Bibliographic databases

Following the AMMA European programme's consortium agreement, any publication submission requires the consent of the other partners. To meet this rule, a publication management procedure has been set up through the AMMA-EU collaborative portal (https://www.amma-eu.org/). Any publication to be submitted, whatever the type, is uploaded by the author in a directory and an information mail sent to all the programme partners. If no objection is expressed, the publication can be submitted. Then the full text document is moved successively to different directories following its status (submitted, published...). Depending on the publishers' policy, the access to the documents is limited to those with an account on the AMMA-EU website or disseminated publicly. This system allowed sharing of papers at early maturity stages within the other programme's participants and fostering of early collaboration between partners.

In order to extend publication sharing dynamics towards the largest AMMA community and preserve the collection of papers, a collaborative site has been set up online (http://biblio.amma-international.org/ or http://amma.agrhymet.ne/ammaBiblio/). It is based on the WIKINDX tool (http://wikindx.sourceforge.net/) and aims at gathering together published and submitted articles, plus theses, drafts, conference articles, posters and communications concerning the African monsoon. The AMMA publication philosophy is that of open archives. Anybody can access references and their abstracts, usually with a URL link to the publishers. Registered users can access more detailed bibliographic notices and full texts of articles in a version complying with the publisher's policy. Every registered user can easily contribute and enrich or update the bibliographic resources. At present, the AMMA Publication has a count of 950 references corresponding to more than 1400 authors. It is the most exhaustive source of information on the African monsoon available for all and helps enhancing the scientific impact of the AMMA programme.

4. Conclusion

The AMMA programme gathers together hundreds of researchers coming from different disciplines. In order to promote data and information exchange within the community, several cross-disciplinary tools have been developed. They are based on free and robust technologies and software that made their duplication in West Africa possible. The different archives are rich and reveal the tools and user support efficiency, as well as the willingness of programme's participants to share data and publications, even at early research stages.

Thanks to the AMMA programme, the research community for the West African monsoon issues emerged and will now develop its activities in various national and pan-national projects. The different tools detailed here constitute a legacy of the AMMA programme, and will be maintained and improved such as serving the needs of new projects. In France, the system will be maintained at least up to 2020, but long-term funding to sustain the twin databases system is being sought. Hopefully, the AMMA database will become the core of a multidisciplinary reference database for climate and environment of and in West Africa.

Acknowledgements

The support of the AMMA project is gratefully acknowledged (see http://onlinelibrary.wiley.com/doi/10.1002/asl.331/full for full acknowledgement).

The AMMA operational centre display website has been made possible thanks to the active involvement of the CTI team in CNRM, Météo-France and Thomas Fiolleau, LMD, CNRS. The AMMA database team received the efficient help of Véronique Chaffard, LTHE, IRD and Hartmut Höller, DLR in the data recovering task.

References

Bock O, Guichard F, Meynadier R, Gervois S, Agustí-Panareda A, Beljaars A, Boone A, Nuret M, Redelsperger JL, Roucou P. 2011. The large-scale water cycle of the West African monsoon. *Atmospheric Science Letters* **12**: 51–57, DOI: 10.1002/asl.288.

Fleury L, Boichard J-L, Brissebrat G, Cloché S, Eymard L, Mastrorillo L, Oumarou M, Ramage K. 2010. A user friendly distributed database for the African Monsoon Multidisciplinary Analyses programme. In *Proceedings of the Data Management Workshop,*

Copyright © 2010 Royal Meteorological Society

University of Cologne, Germany, 29–30 October 2009, Curdt C, Bareth G (eds). *Kölner Geographische Arbeiten* **90**: 45–51.

Kergoat L, Grippa M, Baille A, Eymard L, Lacaze R, Mougin E, Ottlé C, Pellarin T, Polcher J, de Rosnay P, Roujean J-L, Sandholt I, Taylor CM, Zin I, Zribi M. 2011. Remote sensing of land surface during the African Monsoon Multidisciplinary Analysis (AMMA). *Atmospheric Science Letters* **12**: 129–134, DOI: 10.1002/asl.325.

Mertz O, Mbow C, Reenberg A, Genesio L, Lambin EF, D'haen S, Zorom M, Rasmussen K, Diallo D, Barbier B, Moussa IB, Diouf A, Ø. Nielsen JØ, Sandholt I. 2011. Adaptation strategies and climate vulnerability in the Sudano-Sahelian region of West Africa. *Atmospheric Science Letters* **12**: 104–108, DOI: 10.1002/asl.314.

Ruti PM, Williams JE, Hourdin F, Guichard F, Boone A, Van Velthoven P, Favot F, Musat I, Rummukainen M, Domínguez M, Gaertner MA, Lafore JP, Losada T, Rodriguez de Fonseca MB, Polcher J, Giorgi F, Xue Y, Bouarar I, Law K, Josse B, Barret B, Yang X, Mari C, Traore AK. 2011. The West African climate system: a review of the AMMA model inter-comparison initiatives. *Atmospheric Science Letters* **12**: 116–122, DOI: 10.1002/asl.305.

Copyright © 2010 Royal Meteorological Society