DATA NOTE

A spatiotemporal atlas of hydropower in Africa for energy modelling purposes [version 1; peer review: 1 approved, 1 approved with reservations]

Sebastian Sterl, Albertine Devillers, Celray James Chawanda, Ann van Griensven, Wim Thiery, Daniel Russo

1 Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Brussels, 1050, Belgium
2 Department of Earth and Environmental Sciences, KU Leuven, Leuven, 3001, Belgium
3 Center for Development Research (ZEF), University of Bonn, Bonn, 53113, Germany
4 International Renewable Energy Agency (IRENA), Bonn, 53113, Germany
5 Mines ParisTech, Paris, 75272, France
6 IHE-Delft Institute for Water Education, Westvest 7, Delft, 2611AX, The Netherlands

Abstract
The modelling of electricity systems with substantial shares of renewable resources, such as solar power, wind power and hydropower, requires datasets on renewable resource profiles with high spatiotemporal resolution to be made available to the energy modelling community. Whereas such resources exist for solar power and wind power profiles on diurnal and seasonal scales across all continents, this is not yet the case for hydropower. Here, we present a newly developed open-access African hydropower atlas, containing seasonal hydropower generation profiles for nearly all existing and several hundred future hydropower plants on the African continent. The atlas builds on continental-scale hydrological modelling in combination with detailed technical databases of hydropower plant characteristics and can facilitate modelling of power systems across Africa.

Keywords
Hydropower, energy modelling, Africa, resource profiles, renewables, decarbonization

This article is included in the Societal Challenges gateway.
Corresponding author: Sebastian Sterl (sebastian.sterl@vub.be)

Author roles: Sterl S: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Devillers A: Data Curation, Formal Analysis, Investigation, Software, Validation, Visualization, Writing – Review & Editing; Chawanda CJ: Methodology, Resources; van Griensven A: Funding Acquisition, Writing – Review & Editing; Thiery W: Funding Acquisition, Writing – Review & Editing; Russo D: Conceptualization, Project Administration, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was performed under the project CIREG (Climate Information for Integrated Renewable Electricity Generation), which is part of ERA4CS, an ERA-NET Co-fund action initiated by JPI Climate, funded by BMBF (DE), FORMAS (SE), BELSPO (BE) and IFD (DK) with co-funding from the European Union’s Horizon2020 Framework Program (Grant 690462). Further, the work benefitted from a mobility grant provided to S.S. through EIT InnoEnergy. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Sterl S et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Sterl S, Devillers A, Chawanda CJ et al. A spatiotemporal atlas of hydropower in Africa for energy modelling purposes [version 1; peer review: 1 approved, 1 approved with reservations] Open Research Europe 2021, 1:29 https://doi.org/10.12688/openreseurope.13392.1

First published: 26 Mar 2021, 1:29 https://doi.org/10.12688/openreseurope.13392.1
Plain language summary

Hydropower plants rely on river flow to generate electricity. Since river flows change between different seasons, electricity from hydropower plants will also change from season to season. In this paper, we present a new database that contains calculated profiles of electricity generation from season to season for hundreds of hydropower plants in Africa, both existing and future ones. This database will be helpful to scientists doing research on electricity generation in different African countries.

1. Introduction

To achieve the long-term objectives of the Paris Agreement, it is well-established that electricity supply worldwide will have to decarbonise by mid-century. In this context, it is imperative that the shares of low-carbon resources in power systems increase. Low-carbon resources include solar photovoltaics (PV), concentrated solar power (CSP), wind power, hydropower, geothermal power, ocean power, bioenergy and nuclear power. Among these, the strongest growth rates over the past decade, and the highest drops in price, have been recorded by solar PV and wind power, which are thus seen more and more as potential backbones of future power systems.

Given the dependence of solar PV and wind power generation on meteorological variables, these are classified as “variable renewables”, or VRE. Because of this variability in generation from short (sub-hourly) to long (seasonal and interannual) timescales, increasing the share of VRE in electricity systems will require increased flexibility and storage to solve issues related to mismatches between VRE supply and electricity demand, which must be considered in modelling exercises.

Although solar and wind power have recorded the highest rates of growth among renewable resources in recent years, the most-used renewable electricity resource worldwide is currently still hydropower. This comprises run-of-river hydropower without storage, which is essentially another form of VRE; reservoir hydropower, which can be dispatched flexibly to aid VRE grid integration; and pumped-storage hydropower, which can be used as a “battery” to avoid curtailment of surplus VRE.

To inform long-term planning and modelling of renewable power capacity expansion, it is crucial that reliable resource profiles of VRE and hydropower are available to the modelling community. The inclusion of such resource profiles at high spatiotemporal resolution, from hourly to seasonal and interannual timescales and across geospatial zones of different resource strengths, is crucial to accurately represent modern renewable technologies in energy system models. For this reason, dedicated spatiotemporal databases on solar and wind resource strength and availability have been developed, such as the Global Solar Atlas and the Global Wind Atlas. This African hydropower atlas is hereafter abbreviated by “AHA”.

The data gap is especially problematic for regions where (i) hydropower forms an important backbone of many power systems, (ii) substantial expansions of hydropower generation are still planned, and (iii) precipitation patterns are highly variable on seasonal timescales. All of these apply to the African continent, for which science-based services for the renewable energy sector are in short supply. To close the data gap and improve the resources available for energy modelling on Africa, we present here a new spatiotemporal data atlas for nearly all existing and several hundred future hydropower plants across the African continent, containing (i) geospatial references, (ii) technical characteristics, and (iii) seasonal power plant availability profiles, including uncertainty ranges reflecting interannual hydrological variability. The seasonal availability profiles in the atlas include the effect of reservoir sizes on operational possibilities to shift seasonal availabilities of hydropower dispatch, and of current and future configurations of hydropower plants in a cascade. This African hydropower atlas is hereafter abbreviated by “AHA”.

2. Materials and methods

The AHA, which is herewith made freely available to the research community, is designed to be a comprehensive resource containing technical, spatial, and temporal data on existing and future hydropower plants across Africa. It covers all continental African countries which together constitute the major African Power Pools (respectively the North, West, Central, Eastern, and Southern African Power Pool), as well as the island nation of Madagascar.

The AHA is collated into a single spreadsheet-based file which contains both inputs and results of the calculations carried out to establish the atlas. An overview of the calculation flow performed to obtain the full dataset is provided in Figure 1. Each of the elements of this workflow are described in a separate subsection hereafter.
2.1 Database of technical characteristics of African hydropower plants

The technical information for each hydropower plant includes the rated capacity (in MW), the reservoir size (in million m$^3$ wherever applicable), the multiannual mean discharge of the river section upon which the plant is located (in m$^3$/s), the design discharge wherever known (in m$^3$/s), the earliest expected year of entry into service, and the multiannual average capacity factor of the plant wherever known from previous research (in %). In cases where the latter value was unknown, it was assumed to be 50% based on typical values observed for hydropower plants around the world.

This data was collated from a wide array of available information. Globally, the data sources can be divided into three categories: (i) existing hydro databases, such as the Global Reservoir and Dam (GRanD) database, the FAO’s Dams in Africa dataset, and the West African Renewable Power Database (WARPD); (ii) bespoke information, pertaining to individual hydropower projects, from technical project overview sheets, environmental impact assessments, white papers, scientific papers, and other technical modelling studies; and (iii) online news articles on hydropower projects. The consultation and selection of data sources happened strictly according to the hierarchy (i)-(ii)-(iii), with sources from category (i) forming the default, being supplemented by categories (ii) and (iii) wherever necessary. All used data sources are referenced in the AHA.

The processing of this data to calculate temporal hydropower availability profiles is explained further below, in section 2.5.

The database includes both existing (active) hydropower plants, as well as future plants. The term “future” is relatively broad and may encompass, for example, projects under construction or in the pipeline, projects in need of financing, or projects in the pre-feasibility phase. In many cases, distinguishing between these categories is not straightforward. Based on the above-mentioned data sources, the AHA distinguishes between three categories of future projects in descending order of concreteness: committed, planned, and candidate. For any future plant where no specific information was identified regarding its status (as of the writing of this paper), the categorization was set to “candidate” by default. In those cases, the “first year” parameter was left empty. Projects in this category may either be currently unlikely to obtain financing, have been shelved, or have never gone beyond pre-feasibility studies.

We note that we constrained the entries to the current version of the atlas by the criterion that the data should be available in publicly consultable sources. Thus, the atlas could be improved if presently undisclosed information available in, for example, internal documents of planning agencies were to be made publicly available. We therefore eagerly invite all relevant stakeholders to review and submit corrections and/or missing data to the author team, since the goal is for the database to be regularly updated. This particularly concerns the list of future projects, which can likely be expanded much beyond its current state and of which we do not claim full comprehensiveness.

Currently, the AHA contains a total of 633 entries on hydropower plants, of which 266 are existing, 60 committed, 44 planned and 263 candidates. Their total capacity amounts to 132 GW, of which 24% is existing (approximately 32 GW, lining up well with other statistics on existing plants), 19% committed, 6% planned, and the remaining 51% candidate. The division of the total capacity by category and by country is shown in Figure 2.

We note that hydropower plants have been allocated to the country of their coordinates, notwithstanding that, in some cases, a part of the produced electricity would be allocated for exports (e.g. hydropower plants in some river basins are shared among all riparian countries). In the cases of hydropower plants located on rivers forming country borders (11 cases in total in the AHA), their capacity was allocated equally among the countries in question, thus forming separate entries in the database.

2.2 Database of geospatial coordinates of African hydropower plants

The geo-referencing of hydropower plants was done according to a hierarchy of data choices, depending on the status of each plant. Firstly, all existing plants were georeferenced...
Figure 2. Overview of total capacity of existing, committed, planned, and candidate hydropower plants across Africa as collected in the AHA, for countries where this capacity totals (a) > 5 GW, (b) 1–5 GW, and (c) < 1 GW. DRC = Democratic Republic of the Congo; Congo (Rep.) = Republic of the Congo; CAR = Central African Republic.

A spatial overview of the hydropower plants collected in the AHA is shown in Figure 3.

2.3 River flow dataset for the African continent
To estimate hydropower generation profiles for each of the identified locations under the given technical plant characteristics, estimations of river flow at monthly resolution on the African continent were obtained from dedicated simulations with SWAT+ (Soil and Water Assessment Tool\cite{30}). A previous version of this dataset has been used for hydropower potential assessment in West Africa before (refs. 9,31); the updated version used for this paper is available through the repository in ref. 32. Detailed descriptions of the characteristics of the simulations are provided in refs. 9,33,34; performance metrics of the simulations in comparison to observed data from the Global Runoff Data Centre (GRDC) are described in ref. 34. The most important points from these publications are repeated below.

In SWAT+, watersheds are delineated into sub-basins from which hydrologic response units (HRUs, which are distinct areas of a sub-basin with a unique combination of land use, soil type and slope class) are defined. For the SWAT+ model used for the AHA, sub-basins were delineated using 3,500 km² as...
threshold, yielding 5,700 sub-basins and 461,829 HRUs across the African continent. Input data was obtained from the following sources:

- Digital elevation: A 90 x 90 m Digital Elevation Model (DEM) acquired from the Shuttle Radar Topography Mission;
- Land use: Data from the Land Use Harmonization (LUH2) dataset at 0.25° x 0.25° resolution;
- Soil: Data from the Africa Soil Information Service (AfSIS) dataset resampled at 0.25° x 0.25°;
- Meteorological forcing: Data from the EWEMBI dataset at 0.5° x 0.5°.

Figure 3. An overview of the georeferenced African hydropower plants by category (existing, committed, planned, candidate). Sizes of icons reflect installed capacity as per the legend. The characters (A)–(F) refer to the plants whose temporal power generation profiles are shown in Figure 4. Background: Esri’s World Imagery (see Acknowledgements).
Further, the following methodologies were employed to estimate evapotranspiration and surface runoff and perform flow routing:

- Evapotranspiration: Using the Penman–Monteith method;
- Surface runoff: Using the Soil Conservation Service curve number method;
- Flow routing: Using the variable storage routing method.

Temporally, the simulations were carried out at daily resolution across the 37-year period 1980–2016. For the reposited dataset, results were averaged to monthly timescales to reduce file size. The first eight years of the simulation were considered as spin-up time and left out of the analysis. Spatially, each river section of the modelled river network is designated by a unique identifier (ID) as provided in the reposited dataset, to which hydropower plant coordinates could be mapped (see next section).

2.4 Inflow profiles for each African hydropower plant
The geospatial information described in section 2.2 and the river flow information described in section 2.3 were combined as follows to obtain the river inflow feeding each hydropower plant.

First, the geospatial hydropower plant information (coordinates) was mapped to the river network of the SWAT+ simulations (river sections), such that monthly river flow across the 37-year simulation period could be extracted separately for each hydropower plant. This “snapping” was straightforward in 74% of cases, with hydropower plant coordinates being precisely covered by the SWAT+ river network. In the other 26% of cases, the river stretch most representative for the hydropower plant coordinates was selected according to the following hierarchy. First, if the hydropower plant coordinates were so close to the river source that the modelled SWAT+ network did not extend sufficiently far upstream, the most upstream river section in the modelled network (downstream of the plant coordinates) was selected. Second, if the hydropower plant was located on an affluent not covered by the SWAT+ network at all, the geographically nearest river section in the same river basin (draining into the same main river) was selected. Third, in the extremely rare cases where the entire river basin of the hydropower plant was not covered by the SWAT+ network, but a nearby river basin with the same prevalent precipitation seasonality was covered, the geographically nearest river section of that basin was selected. Note that in all these cases, the objective of this snapping was to infer a reasonable estimate of river flow seasonality and interannual variability for each hydropower plant. The AHA includes the selected SWAT+ river section ID for each identified set of hydropower plant coordinates.

Second, a typical range of years of different “wetness”, spanning the range from very dry to very wet years, was selected as follows. First, the flow profile for a “normal year” was defined as the monthly median of the dataset. Subsequently, the flow profile for “very dry” and “very wet” years was taken to be the “normal year” profile multiplied by a corrective factor, calculated as the ratio of the 5th (very dry) and 95th (very wet) percentile value of average annual flow to the multiannual average flow. To account for the fact that some few hydropower plants with very large reservoirs are capable of buffering water on interannual timescales and thus mitigate interannual variability, an exception in the calculation was made for those plants with a typical filling time of more than one full year. For these plants, instead of the 5th and 95th percentiles, the 10th and 90th percentiles were taken to account for this mitigation of dry and wet extremes on interannual timescales.

Third, the seasonality of river flow for these three types of years (very dry, normal, and very wet, each characterized as a time series of twelve values representing the months of the year) was calculated by dividing each time series by the multiannual average flow. In this way, the (normalized) seasonality was obtained for each plant in the AHA for which a match of geospatial coordinates with SWAT+ simulated river stretches could be performed.

Fourth, wherever possible, the three resulting time series of river inflow to each hydropower plant were additionally bias-corrected (using the simple scaling technique) to the multiannual mean river discharge value collected from existing databases and literature (see section 2.1). This last step could be performed for 60% of cases (380 out of 633 plants).

2.5 Calculation of representative seasonal hydropower availability profiles for energy modelling
The final step in the calculations was to convert the typical river inflow datasets (whether bias-corrected or not) for each reservoir to typical power output profiles. A distinction was made between (i) run-of-river hydropower plants, (ii) reservoir hydropower plants, and (iii) hydropower plants in a cascade. For each of these, typical profiles of outflow (e.g. of turbined water) were calculated from inflow profiles as described below, before these were further converted to typical seasonal capacity factors.

2.5.1 Run-of-river hydropower plants. For run-of-river hydropower plants, the turbined outflow profiles were taken equal to the inflow profiles. Power generation was assumed to be a linear function of the turbined outflow profile, with the exception that maximum power output was assumed to be reached when outflow was equal to or higher than the design discharge reflecting the fact that run-of-river hydropower plants are typically designed to produce at full capacity during several months of the year, not only during the single wettest month).
Typical seasonal capacity factors were thus calculated according to:

\[
\left\langle CF_{\text{hydro}} \right\rangle_{m}^{n,d,w} = \min \left( \frac{\left\langle Q(t) \right\rangle_{m}^{n,d,w}}{Q_{\text{design}}}, 1 \right),
\]

where \(\left\langle CF_{\text{hydro}} \right\rangle_{m}^{n,d,w} \) is the average capacity factor of the hydropower plant in month \(m\) during a normal (\(n\)), very dry (\(d\)) or very wet (\(w\)) year; \(\left\langle Q(t) \right\rangle_{m}^{n,d,w} \) is the average turbined inflow in that month; and \(Q_{\text{design}}\) is the design discharge.

In cases where the design discharge was not known, it was estimated by dividing the multiannual mean river discharge value (used for bias-correction of SWAT+ data) by the multiannual average capacity factor recorded in the AHA (assumed to be 50% unless known otherwise, as mentioned in section 2.1). Thus, for instance, the design discharge of a hydropower plant with an average capacity factor of 50% was assumed to be twice the average discharge. For such cases, the capacity factor was thus calculated according to:

\[
\left\langle CF_{\text{hydro}} \right\rangle_{m}^{n,d,w} = \min \left( \frac{\left\langle Q(t) \right\rangle_{m}^{n,d,w}}{Q_{\text{mean}}} \times CF_{\text{hydro}}^{\text{mean}}, 1 \right),
\]

where \(CF_{\text{hydro}}^{\text{mean}}\) is the assumed multiannual average capacity factor, and \(Q_{\text{mean}}\) the multiannual average river discharge.

In those cases where neither the design discharge \(Q_{\text{design}}\) nor the multiannual mean river discharge \(Q_{\text{mean}}\) were available (the latter meaning that no bias-correction could be performed), it was assumed that the design discharge corresponded to 50% of the maximum flow in a “normal” year. The (non-bias corrected) monthly profiles were then divided by that (non-bias corrected) value, thus obtaining an estimate of typical monthly average capacity factors:

\[
\left\langle CF_{\text{hydro}} \right\rangle_{m}^{n,d,w} = \min \left( \frac{\left\langle q(t) \right\rangle_{m}^{n,d,w}}{0.5 \times \max_{m} \left\{ \left\langle q(t) \right\rangle_{m}^{n,d,w} \right\}}, 1 \right),
\]

where \(q(t)\) represents the flow time series before bias-correction.

All above calculations were performed separately for the months of a normal, very dry, and very wet year. An example of a capacity factor profile calculated for a run-of-river hydropower plant is shown in Figure 4(a).

### 2.5.2 Reservoir hydropower plants

For all reservoir-based plants, the reservoir inflow was separated into a “storable” and a “non-storable” component, based on the typical “filling time” of the reservoir (the time it would take for the average inflow to fill the reservoir). This approach is described in detail in the Supplementary Material of ref. 9 and briefly summarized here.

Essentially, the “storable” component corresponds to the percentage of inflow that, if cumulated across the year, would be precisely enough to fill the reservoir’s live storage volume; this component is assumed to be stored by the reservoir and redistributed equally over the different seasons (see section 3 for a discussion of this assumption). The “non-storable” component, on the other hand, corresponds to the remainder of the inflow which hence cannot be stored (as this would lead to spilling, which is to be minimized in normal reservoir operation schemes); it is therefore assumed to be directly turbined.

For reservoirs with a filling time of more than one year, the non-storable component is equal to zero. Note that the filling time can differ between dry and wet years; thus, a reservoir’s non-storable component may be zero during very dry years (resulting in an unseasonal outflow profile) but finite during very wet years (bringing a seasonal peak into the outflow profile). We assumed live storage volume to be 70% of total reservoir volume in all cases.

The total outflow of the reservoir-based plants was then calculated as the sum of the redistributed “storable” and “non-storable” flow components. Subsequently, the conversion of these outflow profiles to typical monthly average capacity factor profiles was done as described by Equation (1)–Equation (3) in section 2.5.1.

Four examples of capacity factor profiles for reservoir hydropower plants are shown in Figure 4(b)–(e), of which two with less-than-a-year (b–c) and two with more-than-a-year filling time (d–e).

#### 2.5.3 Cascade configurations

For the development of the AHA, the definition of a “cascade” was taken to refer to any one or more run-of-river plants, or plants with relatively small reservoirs, being located downstream of larger reservoir plants on the same river stretch. In such cases, the inflow profile of the first downstream run-of-river plant was taken equal to the calculated outflow profile of the upstream reservoir plant; the inflow profile of the second downstream plant was taken equal to the outflow profile of the first downstream plant; and so forth. Finally, the outflow profiles of each plant were converted to typical monthly average capacity factor profiles as described by Equation (1)–Equation (3) in section 2.5.1.

Since cascade configurations can be time-dependent – for instance, a reservoir plant may be planned or under construction upstream of an existing run-of-river plant – the outcomes of this calculation depend on the year for which the calculations are performed, and whether this is before or after the planned reservoir plant comes online. To differentiate between these cases, the AHA contains results sheets for different example years: 2020, 2030, and “All”, the former two reflecting the hydro fleets of 2020 (present-day) and 2030, respectively, and the latter reflecting the hypothetical case where all hydropower plants, including “candidate” plants, are constructed.
Figure 4. Six demonstrations of the monthly typical capacity factor profiles in the AHA (normal years as well as very dry and very wet years). Showcased are a run-of-river plant (a), two reservoir plants with less-than-a-year storage capacity (b–c), and two reservoir plants with more-than-a-year storage capacity (d–e). Further, the plant in (c) will form part of a cascade with (e) in the future, resulting in profile (f). GERD = Grand Ethiopian Renaissance Dam.

An example of capacity factor profiles for a hydropower plant that is currently not part of a cascade system, but will become so in the future due to upstream construction of a large reservoir plant, is provided in Figure 4(c) & (f).

2.5.4 Data coverage. With these procedures, seasonal availability profiles could be calculated for 550 out of 633 hydropower plant entries in the AHA (87%). For the remaining 83 entries – mostly small existing plants for which the snapping to the simulated river network could not be performed with confidence (see section 2.4), and “candidates” with unclear locations – the profiles could not be calculated from the present version of the AHA. Future iterations of the database and the simulations may make it possible to further close this gap.

3. Use and limitations of AHA data in energy modelling

The data provided in the AHA is aimed at servicing the energy modelling community to enable better representation of seasonal constraints of hydropower availability at a plant-by-plant level. The best way to import these profiles into any model will depend on the specific software used.

However, the general principle of importing and applying the profiles in energy models is as follows. For run-of-river plants, the AHA profiles can be used as-is (i.e. considered fixed), as these plants are not considered to be dispatchable, and cannot ramp up or down in function, for example, of the day-night cycle of solar PV or power demand. These profiles are thus to be used in the same way as would solar PV or wind resource profiles.

For reservoir plants, the profiles denote seasonal availability constraints rather than a fixed curve of power output. Such plants can be dispatched flexibly up to a certain extent, for example, to follow demand or to aid VRE integration, constrained by typical (sub-)hourly ramping rates which are different from case to case. In such cases, the modelling should be set up in such a way as to ensure that the power plants are represented as dispatchable technologies but constrained by average seasonal availability profiles as given by the AHA.

It is important to note that the AHA represents a first attempt at providing a comprehensive, continent-wide spatiotemporal dataset for Africa. As such, it is subject to various limitations which must be considered. The most important limitations are summarised below.

First, the river flow profiles were obtained from simulations representing a historical period. Thus, any potential effects of future climate change on river flow, which may be substantial, have not been taken into account. However, this has been planned for future iterations of the AHA based on SWAT simulations forced by relevant data from climate change scenarios.

Second, for the same reason, the capacity factor calculations were purely based on simulated reservoir inflow and did not consider evaporation and precipitation effects on the reservoir.
surfaces of future reservoirs which do not form part of the hydrological network as simulated. However, the effects of this omission are expected to be relatively minor since inflow is normally by far the dominant component of reservoir water budgets. (A notable exception to this rule is Lake Victoria, a natural lake that was later dammed for hydropower generation at its outlet.)

Third, the calculations did not explicitly model reservoir dynamics and thus do not include the effect of seasonal hydraulic head variations on seasonal capacity factors. While this effect exists, it is typically minor except for reservoir plants with very low heads.

Fourth, the calculations took a strong supply-side view in assuming that the purpose of hydropower reservoirs is to (partly) remove the seasonality and variability of river inflow such as to stabilize power output on seasonal timescales. However, in cases where power demand itself has a strong seasonality, or in cases where other sources in the electricity mix, like solar and wind power, exhibit extremely pronounced seasonalities and these have a major effect on the supply-demand balance, reservoir hydropower may be required to follow these seasonalities rather than fully flattening the “storable” component of river flow. If the load profiles that hydropower should follow are known, corresponding calculations could be straightforwardly carried out by adapting the methodology described in section 2.5.2. However, we note that this is mostly of importance for reservoirs with more-than-a-year storage capacity (7% of entries in the AHA). For such cases, we recommend that specific case studies be undertaken on the hydropower plants in question to elucidate the potential re-introduction of seasonalities under integrated hydro-VRE operation, such as ref. 41.

Fifth, for all hydropower plants, there may be additional constraints not included in the AHA that impact their inclusion in energy modelling exercises. For example, there may be certain environmental outflow constraints that put further limits on monthly hydropower generation42, or certain hydropower plants where power generation needs to be co-optimised with irrigation or other secondary purposes43.

Sixth, in its current form, the AHA covers the African mainland and Madagascar. However, there is potential for small hydropower plants on other, small African island nations such as São Tomé & Príncipe and the Comoros. These are currently not covered by the hydrological simulations used for the AHA. However, these countries will be integrated into the AHA in the future, contingent upon more exhaustive river flow data becoming available.

4. Conclusions and outlook
This paper describes a new African Hydropower Atlas, which marks the first, continent-wide spatiotemporal database of hydropower generation profiles for existing and future hydropower plants. The aim of the AHA is to provide estimates of monthly constraints on capacity factors of hydropower plants to the energy modelling community at a plant-by-plant resolution, taking the differences between moderately dry, normal, and moderately wet years into account. The data set is made freely available in a spreadsheet-based format; in the future, it may be integrated into a web-based interface to allow interactive visualisation of the results and promote more widespread diffusion of the resource.

By helping energy modellers to better represent hydropower plants’ contribution to electricity mixes across Africa, the AHA may support more informed prioritisation of future hydropower projects to be developed. This is important both from a financial and an environmental point of view. On the financial side, using AHA data in energy modelling may help elucidate which hydropower plants would be most suitable to contribute to a cost-optimised configuration of future power mixes, taking into account the seasonal variability of the hydro resource. On the environmental side, we note that it is undesirable that Africa’s full hydropower potential be exploited, such that excessive ecological impacts of river-damming interventions may be avoided44; using AHA data, priority could be allocated to hydropower plants whose contribution to diversified electricity mixes would be most conducive towards low costs and high VRE penetration, allowing to deprioritize and/or shelve plans for other hydropower plants and avoid lock-in to hydro-dependency45.

The main contribution of this work to the existing literature is the collation of large amounts of data and their processing into a single final product. This is not to say that the data sources that have been used are necessarily the best ones available. In the future, we hope that new iterations of hydrological simulations, new knowledge on the effects of climate change, and new knowledge on existing and upcoming hydropower plants as communicated by public documents and stakeholder feedback can be integrated into the AHA to improve its quality.

Data availability
HydroShare: Online repository of materials for an all-Africa hydropower atlas (v1.0). https://doi.org/10.4211/hs.acff23a8fcede4703a7f1f8a3a75b68bd19.

This project contains the following underlying data:
- The AHA provided as a spreadsheet (.XLSX), containing the geospatial references of the hydropower plants and their technical characteristics used in the calculations, as well as their typical monthly capacity factor profiles for normal, dry and wet years
- SWAT+ simulation results used to extract river flow profiles provided as text files (.TXT)
- GIS shapefile of the river sections covered in the SWAT+ simulation

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Code availability
Analysis code available from: https://github.com/VUB-HYDR/2021_Sterl_etal_AHA
Acknowledgements

The authors thank A. Miketa, D. Gielen, R. Roesch, C. Ruiz, I. Gherboudj, M.S. Nababa, Y. Li & T.T. Bastianello (IRENA), S. Kuujala, J. Huhdanmäki, J. Leino & M. Jokinen (Wärtsilä), and A. Campbell, E. Rich, S. Law, O. McCue, & C. Diez Santos (International Hydropower Association) for inspiring comments and discussions.

The map in Figure 3 was created using the Esri Satellite Basemap and thus using the ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used herein under license. Copyright ©Esri. All rights reserved. For more information about Esri® software, please visit www.esri.com.

References


8. IEA Hydropower: Flexible hydropower providing value to renewable energy integration. Reference Source


Characterizing the potential for hydropower production at any location of the hydrological network is obviously key. This requires to estimate the mean resource but also its temporal variability at different temporal scales, including the seasonality and the interannual variability. The work of Sterl et al. aims to provide this characterisation for a large set of locations of the hydrological network in Africa, provided the catchment area of the drainage basin is large enough.

To my knowledge, this characterisation was never proposed before. The characterization is based here on river flow time series reconstructed for a 37 years period (1980-2016) from hydrological simulations. River discharges are simulated with SWAT+ from weather pseudo-observations, that is temperature and precipitation reconstructions from different observational datasets.

As reported by the authors in the discussion, the AHA (African Hydropower Atlas) represents a first attempt to provide a comprehensive continent wide spatio-temporal dataset for Africa. This is obviously a very relevant contribution and the AHA is expected to be an important dataset for practitioners and policy makers. A large amount of data has been also collected to characterize water reservoirs along and beside the river network.

At the present state however, the methodology used to develop the AHA is not described in enough detail and important clarifications are needed. The AHA is also subject to important limitations. The authors list a number of those. Other important limitations are missing and have to be mentioned and perhaps also discussed. It is very likely that the dataset will be widely used in the coming years, by many “non expert” engineers and policy makers; this definitively call for a better description of the AHA and of its quality.

The AHA has definitively to be indexed in a short time because of its high value for engineers and policy makers but, for the different reasons mentioned here, I recommend major corrections / clarification before the indexing of the manuscript.

Meteorological data and hydrological model
From what I can judge, the most critical limitation is related to the quality of hydrological reconstructions. There are here some reasons of potential concern. This has to be clearly mentioned in the discussion section (perhaps also in the abstract and in the introduction).

Hydrological data are scarce in Africa. Hydrological data are therefore here obtained via simulation from SWAT+, a conceptual hydrological model, forced with EWEMBI weather data. EWEMBI are not observations, but estimates of past weather conditions obtained with weather observations from stations, outputs of meteorological models, satellite data (satellite data do not provide observations of precipitation but proxy)... The diversity of data used in EWEMBI have to be acknowledge in the manuscript to highlight the potential complexity of such reconstruction and the potential errors associated. The limitations of EWEMBI have also to be acknowledge. They are expected to be large in regions without or with scarce meteorological stations which is the case almost everywhere in Africa. Some reference with the relevant evaluation should be provided.

The quality of the SWAT+ hydrological model used for the reconstruction is another possible critical point. It not presented in the manuscript and thus unknown. A section should be ideally dedicated to the description / discussion of its quality for Africa. SWAT+ has been used / evaluated for West Africa (2 references are mentioned in the paper for this). What is its performance for the rest of the continent? For the performance metrics, the authors mention the study of Chawanda et al. 2020, but this study only focuses on South Africa. Then, even if we just consider this region, the performance of SWAT+ is rather (very) low and I fear that many hydrologists would not consider the simulations as very useful : the NASH efficiency criterion is indeed negative for more than 35% of evaluated river sites (that is, the model is worst than a simple « constant » model, where the constant is just the « interannual mean observed discharge »).

For South Africa, Chawanda et al. also mention potentially critical limitations due to the low unavailability of reservoir management data; and to the limited information on agricultural management practices. These are expected to strongly modify 1) the water balance and then the available resource and 2) the seasonality of the river discharge downstream the dams. This has likely also to be discussed, at least mentioned.

In a conventional publication for the academic world, the reader will guess how large is the uncertainty obtained with SWAT+ simulations in this scarce data context. The AHA is however to be used by practitioners / policy makers. The limitation associated to the different datasets used / produced here have clearly to be mentioned, especially for Africa, where the density of the meteorological / hydrological network is very low. The reader can understand that there are many paths for improving the models and then the AHA dataset but I recommend that a fair evaluation of the performance of the modelling chain and of the hydrological model is presented.

Finally, the concepts behind SWAT+ are rather simple. Other modelling approaches are possible and other have been proposed in the recent year. One important contribution here is the GloFas-ERA5 operational global river discharge reanalysis 1979-present (Harrigan et al. 2020)¹. The simulations are obtained with LISFLood a hydrological model with a long time development and evaluation worldwide. The authors should also cite this dataset and highly recommend practitioners to consider different “reanalysis” datasets to have an idea of the uncertainty / errors associated to hydrological data and simulations.

**Bias correction.** P7 C2 § 3. The principle / interest / limitation of bias correction have to be
If bias correction is applied on simulated discharges, observations are used therefore. What is then the added value of simulations as the characterisation of hydrological regime can be done with those observations only?

Is bias correction applied on a monthly basis (i.e. a correction function is determined for each month separately as is typically done for bias correction of climate projections)?

If bias correction can be applied to 2 stations with observations on a same river, what is done for the stations that are in between: bias will likely occur also at these intermediate stations; if bias is not corrected there, some discontinuity / inconsistency is expected along the river network. This has likely to be commented.

What are the reference periods in the observation and in the simulation used to calculate the bias correction factor? Are the periods the same for a given location (and then the hypothesis is made that the bias is the same at any time, for any other year of the simulation period (expected to cover a larger period than the observations?)

The application of bias correction or not for a given location is key. It is mentioned in the metadata of the AHA?

Section "reservoir hydropower plants"

I do not understand the way the operation of reservoirs is simulated. The description given P8 section 2.5.2. is not really clear.

The authors mention "For all reservoir-based plants, the reservoir inflow was separated into a "storable" and a "non-storable" component, based on the typical "filling time" of the reservoir (the time it would take for the average inflow to fill the reservoir). This approach is described in detail in the Supplementary Material of ref. 9 and briefly summarized here."

I did not find this description in the SM of ref9. Note that SM is 33 pages with 10 different subsections. The "storable" term is only used once in this SM in the sub-section "Note 5 : REVUB: Reservoir simulation for small hydropower plants" dedicated to small hydropower plants. This does obviously not correspond to the "large reservoir" case considered here. In the SM of ref9, the authors introduce the REVUB model to simulate the operations of reservoirs. Do they use such a simulation model here (the model REVUB is used in the subsection where the "storable" word is used)?

If yes:

- the basics of the model should be given in the present work, at least in a SM specially dedicated to the present manuscript. The ref.9 is indeed not accessible for free, which is a critical limitation as it is mentioned as a key methodological reference for the present work.

- For the same reasons, the figures S2 and S3 of the SM of ref9 could be included in the present work for a comprehensive illustration of the REVUB model.

- The “REVUB model simulates a baseload oriented operation strategy” of the reservoir, i.e. the objective of the operations is the production of a baseload discharge downstream. From what I can understand, this is thus basically just a low pass filtering of the highly variable inflow in the reservoir (where the intensity of the filter is defined by the parameters in figure S3 of SM in ref9). This assumption may be not appropriate in many reservoir
configurations where the operations are produced to best follow the expected (based on past year observations) annual load profile (and its multiscale temporal variability at least from sub-daily to seasonal). What is the simulation performance of such a model if used in the present work. This "performance" issue should be also likely introduced and an estimation for different reservoir contexts should be presented. The limitations of the model have then also likely to be mentioned: e.g. adequacy of the baseload production assumption, + other assumptions such as those on direct precipitation / evaporation in / from the reservoirs.

If the REVUB model is not used in the present work, the reference to the SM of ref9 is probably not more relevant and a special section should likely describe what is done there. In all cases, the "storable component" concept has to be clarified. I do not understand what is described in § 2 col 2 of p.8. How is estimated this "storable volume"; how is it used and for what then ?...

Is the idea to consider that a filling temporal sequence is absolutely required each year? If yes, can you precise why (because of the occurrence of a dry inflow period each year that lead the reservoir level to go down to low and sub-optimal filling rates?)? Why could we not consider that the reservoir is filled all the time? This would allow for the best production efficiency (with the highest possible hydraulic head at any time)... Clarification is required here. An illustration of a given year (with the time series of the inflow, the filling rate, the outflow, the load) for a given reservoir would be welcome to understand the process and what is refered to as "storable" and "non-storable".

The authors say "Essentially, the “storable” component corresponds to the percentage of inflow that, if cumulated across the year, would be precisely enough to fill the reservoir’s live storage volume; this component is assumed to be stored by the reservoir and redistributed equally over the different seasons (see section 3 for a discussion of this assumption)." From which initial storage level do you start to estimate this required volume? The required volume will be obviously different from one year to the other (as mentioned later in the paragraph) depending on how wet or dry the year was and on large or small the demand was in the preceding year. If you have 30 different years, you can estimate 30 different required volumes. What do you do with these different volumes?

On the other hand, in the real real world, institutions in charge of the operations of reservoirs do not know what will be the meteorological conditions for the next weeks / months : they do thus not know what will be the inflow / load demand for the next months and thus they do not know what is / will be the required volume for the current year. In this real world configuration, the institutions in charge can but use a probabilistic approach to fill the reservoir again, defining a risk level for which the reservoir will not be filled at the end of the year (in the configuration of a highly seasonal flow). So, how is estimated this "storable" component here? How do you account for this "uncertainty / predictability" issue in your approach? A deterministic approach is likely limited there. A discussion on this "operational issue" would be welcome. In all cases, one can understand that simplifications of the operation are required provided they mimic the true operations in a relevant way. In this context, the authors have also to mention the existence of studies where more realistic representations of the reservoir operation are developed (typically based on some optimisation process) (e.g. Minville et al. 2009; Turner et al. 2017; Danso et al. 2021).

Note: The term "storable" is to me not really appropriate if you refer to the "amount of water that is required to fill the reservoir for a given year". This required water amount is obviously smaller than the storage capacity of the reservoir. The storable water suggests : “a water amount that can
be stored but this is not mandatory to store it”. And indeed, the amount of water that could be stored over a given year can be much larger than the reservoir storage capacity: in a configuration where you never have overspilling, all the inflow water volume can be stored (at a given time of the year). For me, the "non storable" component of the inflow is just the amount of water that arrives in the reservoir when the reservoir is 100% filled (something that may happen during some (limited) periods of the year).

The authors are thus gently asked to clarify the terminology / description of the methodology here.

Cascade configurations
How was the outflow profile of upstream large reservoirs determined? This is not clear. Did you consider the output of the VURB model for the upstream reservoirs?
As mentionned by Biemans et al. 2011; reservoirs can contribute significantly to irrigation water supply in many regions and significantly modify water resource downstream. A comment on this is likely required.

Metadata for the description of data / reservoirs
The description of what has been done, with what data for which reservoir / location will be key for practitioners. A table with the description of all data / metadata provided for any given location is thus probably to be given in the manuscript.

Is there any strategy to check the metadata already collected and processed in this work? For the reservoir description (P4, §4) for instance: Are the authors in contact the "Water Resource Authorities" of the different countries to check the exhaustivity / characteristics of the dams considered in the work.

Then it would be interesting to precise the strategy retained to update the data / metadata of the AHA. For instance to update the metadata for the description of the different "existing" and future projects? Is there any reference institution that is committed to do this update? What is the contact for this? Do the authors expect to produce / deliver improved releases of the AHA in the coming years?
P7 §3 : "snapping" : is the hierarchy level considered for each plant in the AHA mentioned in the metadata of the AHA (this information is obviously key)?

Data Availability
I am not a typical enduser of such database but the author say that the AHA gives “- SWAT+ simulation results used to extract river flow profiles provided as text files (.TXT).”
The format of the data there is surprising and makes for me the file hard to exploit. All monthly time steps of the 1980 – 2016 period are given in turn, with all 5438 hydrological units used in the SWAT+ model pooled for each time step in 5438 consecutive lines. We do not have thus access to the time series of simulated discharges for each hydrological unit individually. This could be likely improved so that the enduser could access easily to the full time series (and not only to the yearly mean profiles) of each hydrological unit.
I also did not find a note that describes what is contained in each file of the database. Can such a be given as a SM of the present note?

Detailed comments
P4. First §: the multi-annual mean discharges. Can you precise for which period? Based on which
data?

p7. C2 - § 1: "To account for the fact that some few hydropower plants with very large reservoirs are capable of buffering water on interannual timescales and thus mitigate interannual variability, an exception in the calculation was made for those plants with a typical filling time of more than one full year. For these plants, instead of the 5th and 95th percentiles, the 10th and 90th percentiles were taken to account for this mitigation of dry and wet extremes on interannual timescales." I do understand the rationale in the previous paragraph but I would suggest that the first estimate with the 5th and 95th percentiles is given for all reservoirs and that the 10th and 90th percentiles are given as additional information for the very large reservoirs. Is it also mentioned in the AHA what is the mean time required to fill in the reservoir from the zero level?

p7. C2 - § 2: please reformulate "the seasonality of river flow for these three types of years (very dry, normal, and very wet, each characterized as a time series of twelve values representing the months of the year) was calculated by dividing each time series by the multiannual average flow." Did you apply simple scaling on the mean hydrological cycle characterized by 12 monthly mean values? If yes, please clarify.

p8 C1 §3:
"the maximum flow" >> "the maximum monthly flow"
"monthly profiles" >> "annual cycles"

P10. C1 §1. “Inflow is normally by far the dominant component of reservoir water budget”. Yes, but this is not always the case, especially in arid regions. Typical % values from previous literature in different contexts should be likely mentioned there to fix ideas.

P9 C2 – Climate change. This issue is indeed important. Some references to recent work could be mentioned (e.g. Stanzel et al. 2018; Bichet et al. 2020; Moussa et al. 2020).

SWAT+: The description of SWAT+ has to be given in more details. The authors should give at least the references for the Penman-Monteith, the SCS-CN and the routing method used. To my knowledge (but upgraded versions could have been produced in the meantime), the SCS-CN method is suited for "single rainfall-runoff events" simulations, but not for "continuous" hydrological simulation covering multiple years.

References
Abstract | Publisher Full Text

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Partly

Are the datasets clearly presented in a useable and accessible format?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Hydroclimatology, Hydropower, Mutlipurpose Water Reservoir management and modelling, Climate change impacts and regionalisation, Energy transition, Renewable Energy.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 09 April 2021
https://doi.org/10.21956/openreseurope.14462.r26658

© 2021 De Felice M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Matteo De Felice
Joint Research Centre (JRC), European Commission, Petten, The Netherlands

This article gives a clear description of the African Hydropower Atlas (AHA). The authors explain all the steps behind the creation of the data, and in general the methodology...
is very clear. The task of creating a complete and accurate dataset of all current and future hydropower projects in Africa is definitely challenging and I can say that the authors succeeded in this task.
However, I would like to mention a few points that can be addressed or discussed to improve the work:

1. There are some hydropower plants that are built as a joint initiative of multiple countries which agree then to share the electricity generated. A clear example is the Manantali plant which is actually in Mali but its electricity is shared with Senegal (33%) and Mauritania (15%)

   The authors should consider to mention this and the possibility to add this type of information in the future releases because it can be very relevant for countries with a minimal electrical infrastructure.

2. Do the authors have tried to validate the seasonal patterns of the capacity factors? For example, some Transmission System Operators in African countries publish information on their hydropower plants.

3. The data published is very rich but the text files lack metadata. The authors should add: a) information on the meaning of the columns in the txt files and b) information on the linkage between the Excel file and the SWAT data, in other words what the link between items in SWAT_reservoir_mon_EWEMBI_hist.txt and the plants in the Excel file?

   I think the third point is very important to make the data published more usable and accessible. I would also invite the authors of using a more user-friendly format for their data in the future, for example considering creating a Tabular Data Package.

   **Is the rationale for creating the dataset(s) clearly described?**
   Yes

   **Are the protocols appropriate and is the work technically sound?**
   Yes

   **Are sufficient details of methods and materials provided to allow replication by others?**
   Yes

   **Are the datasets clearly presented in a useable and accessible format?**
   Yes

   **Competing Interests:** No competing interests were disclosed.

   **Reviewer Expertise:** Data science, climate analysis, energy modelling.

   I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.