Check for updates

DATA NOTE

REVISED A spatiotemporal atlas of hydropower in Africa for energy modelling purposes

[version 3; peer review: 2 approved, 1 approved with reservations]

Sebastian Sterl[1-4], Albertine Devillers[4,5], Celray James Chawanda[1], Ann van Griensven[1,6], Wim Thiery[1], Daniel Russo[4]

[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

v3 **First published:** 26 Mar 2021, **1**:29
https://doi.org/10.12688/openreseurope.13392.1
Second version: 22 Dec 2021, **1**:29
https://doi.org/10.12688/openreseurope.13392.2
Latest published: 29 Mar 2022, **1**:29
https://doi.org/10.12688/openreseurope.13392.3

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

Open Peer Review

Approval Status ✓ ? ✓

	1	2	3
version 3 (revision) 29 Mar 2022			✓ view
version 2 (revision) 22 Dec 2021	✓ view		? view
version 1 26 Mar 2021	✓ view	? view	

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

2. **Benoît Hingray**, Université Grenoble Alpes, Grenoble, France

3. **Jean-Philippe Vidal**, INRAE, Lyon, France

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the Horizon 2020 gateway.



This article is included in the Energy Systems Modelling collection.

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: © 2022 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 3; peer review: 2 approved, 1 approved with reservations]** Open Research Europe 2022, **1**:29 https://doi.org/10.12688/openreseurope.13392.3

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

REVISED **Amendments from Version 2**

The article has been improved based on suggested clarifications by referee #3. Several phrasings have been clarified and various additional disclaimers and/or explanations have been provided through additional text and references. In particular, the discussion section has been updated with additional disclaimers related mainly to the bias-correction process of the data, clarifying some of the shortcomings of this approach related to a lack of a harmonised observational dataset that would have allowed for a spatiotemporally fully streamlined bias-correction. We have also clarified several other concepts, e.g. the definition of "capacity factor" and "storable component", in the manuscript text.

Any further responses from the reviewers can be found at the end of the article

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[1]. 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[2], which are thus seen more and more as potential backbones of future power systems[3].

Given the dependence of solar PV and wind power generation on meteorological variables, these are classified as "variable renewables", or VRE[4]. 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[5].

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[2]. This comprises run-of-river hydropower without storage, which is essentially another form of VRE[6]; reservoir hydropower, which can be dispatched flexibly to aid VRE grid integration[4,7–11]; and pumped-storage hydropower, which can be used as a "battery" to avoid curtailment of surplus VRE[12].

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[13]. 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[14] and the Global Wind Atlas[15] or the reanalysis-based web interface "*renewables.ninja*"[16]. Such resources typically allow the user to select locations on the world map and extract representative resource profiles for VRE from hourly to seasonal and interannual timescales, which can then be used in energy modelling exercises.

The picture is different for hydropower. Comprehensive and integrated databases of hydropower resources are currently unavailable to the modelling community at the required level of detail[17]. This is a consequence of the challenge of accurately modelling river flows across a wide range of river basins with different hydrometeorological conditions within a single model framework[18], as well as the wide disparity in individual hydropower plants' technical characteristics[19]. A consequence of this comparative disparity vis-à-vis solar and wind power, and the resulting lack of comprehensive hydropower databases, is that hydropower plants – which are more and more considered to be an important lever to support VRE uptake thanks to their flexibility of dispatch (for reservoir plants) and potential seasonal synergy with VRE (for run-of-river plants) – are often represented coarsely and without the warranted spatiotemporal detail in energy models[9]. For instance, many studies lump hydropower plants in a region together as one single technology without detail on individual plants (e.g. 3,20), do not consider interannual variability of river flows (e.g. 21), or do not contain information on seasonally constrained availabilities of hydropower (e.g. 22).

This 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[23–25], for which science-based services for the renewable energy sector are in short supply[26]. 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 %). Here, the capacity factor of any power plant is taken to refer to the electricity that it generated over a certain time period (e.g. hour, day, season, year...) divided by the power that would have been generated if the plant had run at full capacity over that entire period. 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[2].

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[27], the FAO's Dams in Africa dataset[28], and the West African Renewable Power Database (WARPD)[9]; (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[29]), 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

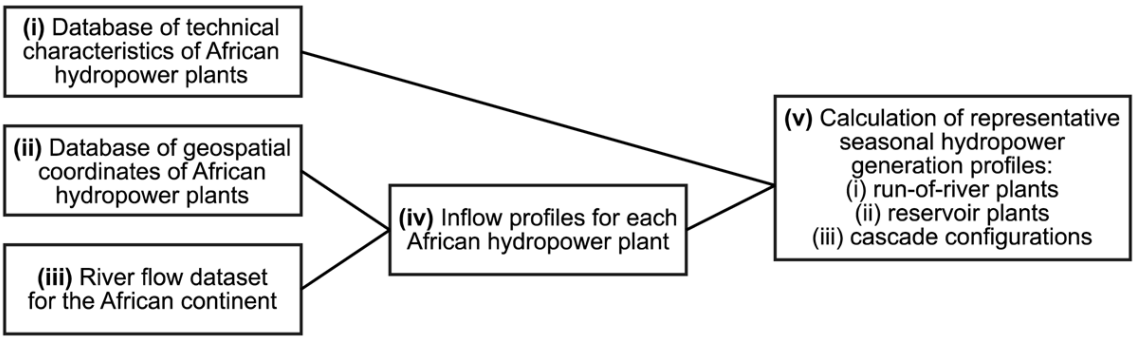


Figure 1. Schematic overview of the various inputs, intermediate results, and outputs of the calculations performed to create the African Hydropower Atlas.

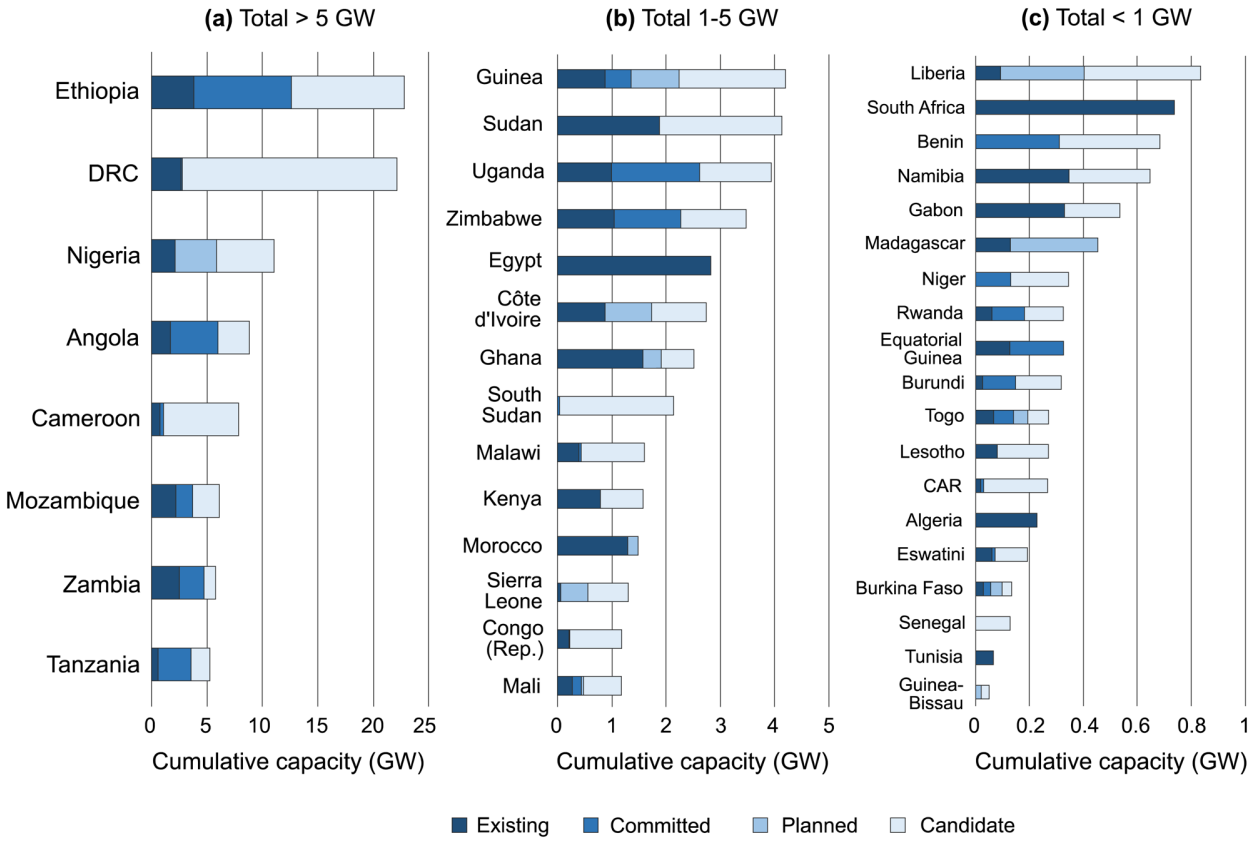


Figure 2. Overview of total capacity of existing, committed, planned, and candidate hydropower plants across Africa as collected in the AHA, for countries where this capacity totals (**a**) > 5 GW, (**b**) 1-5 GW, and (**c**) < 1 GW. DRC = Democratic Republic of the Congo; Congo (Rep.) = Republic of the Congo; CAR = Central African Republic.

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 using satellite imagery; the coordinates given in the AHA correspond to the location of the dam and/or powerhouse as identifiable via Google Maps. Secondly, all hydropower plants that are not yet servicing the grid but are clearly identifiable as being under construction on satellite imagery, were similarly georeferenced. Thirdly, the locations of all other committed, planned and candidate hydropower plants were identified as best possible from specific project information available in any of the consulted sources.

This last category of data could take on a variety of specificity: in some cases, georeferenced coordinates of the intended location of the planned plant were provided in the consulted document(s) as referenced in the AHA; in others, the information remained less precise (e.g. "*the plant will be constructed about 50 km downstream of location A, about 100 km west of city B*"). In the latter case, satellite imagery was consulted to roughly identify the river section corresponding to the description, and a "best guess" location (e.g. where whitewater reveals the presence of rapids, showing a relatively steep head drop) was selected on the river section. We note that, as long as the river section is identifiable at the spatial resolution of the river flow data that is used (see section 2.3), this approximation is unproblematic for the analysis.

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[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

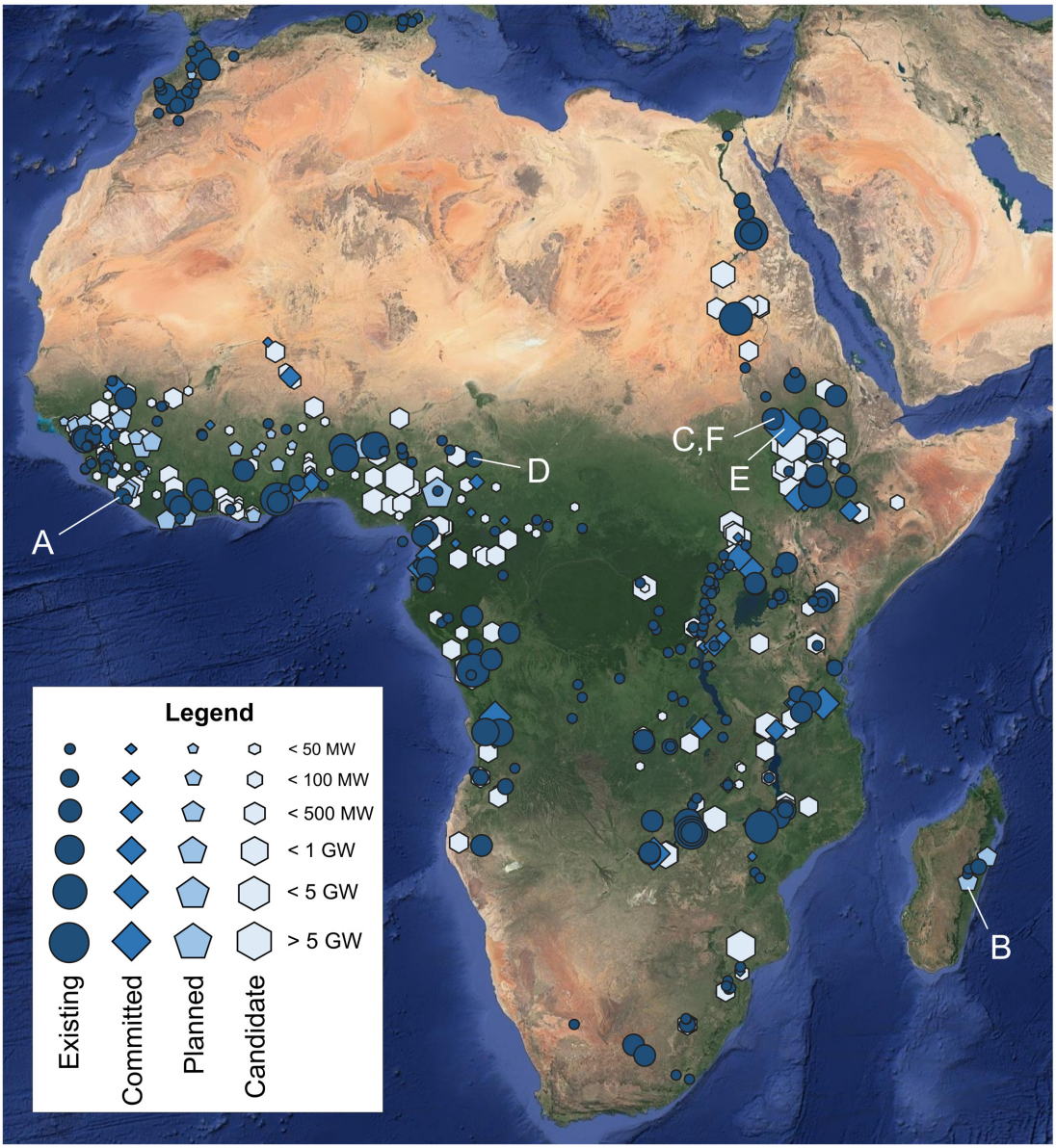


Figure 3. An overview of the georeferenced African hydropower plants by category (existing, committed, planned, candidate). Sizes of icons reflect installed capacity as per the legend. The characters (**A**)–(**F**) refer to the plants whose temporal power generation profiles are shown in Figure 4. Background: Esri's World Imagery[32] (see Acknowledgements).

ref. 33. Detailed descriptions of the characteristics of the simulations are provided in refs. 9,34,35; performance metrics of the simulations in comparison to observed data from the Global Runoff Data Centre (GRDC) are described in ref. 35. 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 km^2 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 × 90 m Digital Elevation Model (DEM) acquired from the Shuttle Radar Topography Mission[36];
- Land use: Data from the Land Use Harmonization (LUH2) dataset[37] at 0.25° × 0.25° resolution;
- Irrigated areas were obtained from Food and Agriculture Organisation (FAO) data at 0.083° × 0.083° resolution[38]. Irrigation modelling was implemented as explained in ref. 35.

- Soil: Data from the Africa Soil Information Service (AfSIS) dataset[39] resampled at 0.25° × 0.25°;
- Meteorological forcing: Data from the EWEMBI dataset[40] at 0.5° × 0.5°.

Further, the following methodologies were employed to estimate evapotranspiration and surface runoff and perform flow routing:

- Evapotranspiration: Using the Penman–Monteith method[41];
- Surface runoff: Using the Soil Conservation Service curve number method[42];
- Flow routing: Using the variable storage routing method[43].

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 seasonal profiles 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[9] 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 seasonal river flow profile thus obtained (for very dry, normal, and very wet years), each characterised as a time series of twelve values representing the months of the year, was normalised to dimensionless values by dividing each time series by the simulated multiyear average flow as simulated (a single number). 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 dimensionless time series of normalised river inflow to each hydropower plant were multiplied again with a bias-correction factor (simple scaling[44]), equal to the multiannual mean river discharge value collected from existing databases and literature (again, a single number; see section 2.1). Bias-correction was only performed for cases where either (i) mean discharge values were available for the hydropower site in question (e.g. from environmental impact assessments where measurements directly at the hydropower plant construction site were invoked), or (ii) mean discharge values were available from gauging stations located at or very close to the hydropower plant. (Note that we did not work with a specific reference period for the bias-correction, which was meant rather as an estimation to be compared to design discharge than as a highly accurate measure of absolute flow rates, since this would have been near-impossible to apply across the board given the disparity in observational data sets.) In this way, bias-correction could be applied to 60% of cases (380 out of 633 plants). We note that this step is not crucial for the data product and serves mostly for providing a dimension to the river flow seasonality (expressing it in m^3/s instead of dimensionless, which allows a comparison to e.g. design discharges of hydropower plants), as the monthly availability curves of hydropower plants could still be estimated without performing the bias-correction step, i.e. for the remaining 40% of cases (see section 2.5.1).

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:

$$\left\langle CF_{hydro} \right\rangle_m^{n,d,w} = \min\left(\frac{\langle Q(t) \rangle_m^{n,d,w}}{Q_{design}}, 1 \right), \quad (1)$$

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

$$\left\langle CF_{hydro} \right\rangle_m^{n,d,w} = \min\left(\frac{\left\langle Q(t)_m^{n,d,w} \right\rangle}{Q_{mean}} \times CF_{hydro}^{mean}, 1 \right), \quad (2)$$

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

In those cases where neither the design discharge Q_{design} nor the multiannual mean river discharge Q_{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 monthly flow in a "normal" year. The (non-bias corrected) annual cycles were then divided by that (non-bias corrected) value, thus obtaining an estimate of typical monthly average capacity factors:

$$\left\langle CF_{hydro} \right\rangle_m^{n,d,w} = \min\left(\frac{\left\langle q(t)_m^{n,d,w} \right\rangle}{0.5 \times \max_m[\left\langle q(t)_m^{n} \right\rangle]}, 1 \right), \quad (3)$$

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 therefore equal to zero.

In terms of volume, the "storable component" of the total yearly volume inflow thus equals the live reservoir volume, and the "non-storable component" is the part of total yearly volume inflow that exceeds this live reservoir volume. As an example, let us say a hypothetical hydropower reservoir has a (live) volume of 2, but a river carries an annual discharge volume of 8 in a "normal" year. In this case, the "storable component" of river flow is equal to 2 (25%, or 2 out of 8), and the "non-storable component" of river flow is the remainder, i.e. 6 (75%).

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)[9]. 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.

The above calculation is performed independently of the initial storage level and of the demand profile. Principally, the methodology amounts to assuming that the amount of water that *can* be stored and flexibly turbined across the year without seasonality effect, *will* be stored and flexibly turbined in this way, since the idea of large reservoirs is precisely to mitigate flow seasonalities (see also the discussion in section 3). It is to be noted that seasonalities in electricity demand, although they exist, tend to be much less pronounced than those in river inflow (cf. ref. 45).

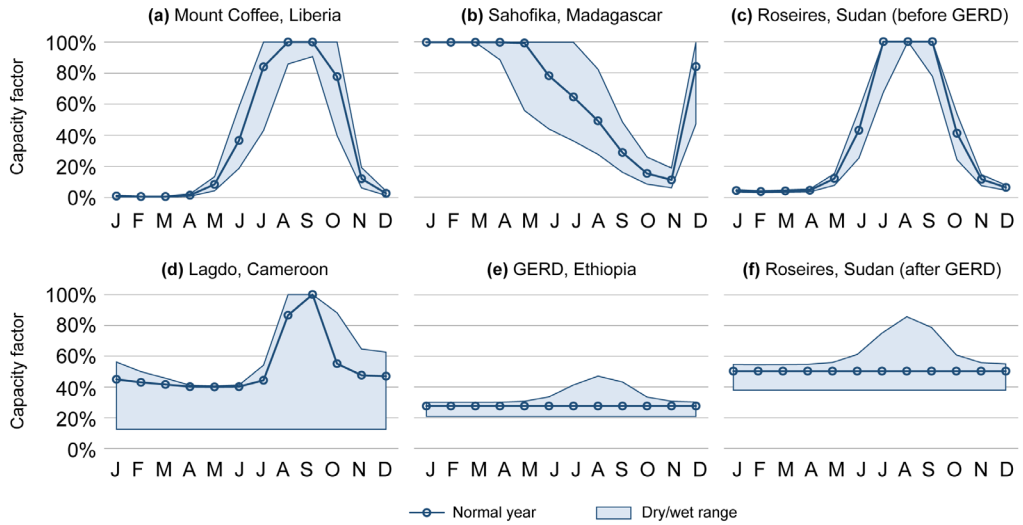


Figure 4. Six demonstrations of the monthly typical capacity factor profiles in the AHA (normal years as well as very dry and very wet years). Showcased are a run-of-river plant (**a**), two reservoir plants with less-than-a-year storage capacity (**b**–**c**), and two reservoir plants with more-than-a-year 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.

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.

By default, this methodology assumes that this outflow profile (as seen from the point of view of downstream run-of-river plants) does not change significantly before arriving at the downstream plant. We deem this a reasonable assumption as cascade configurations typically consist of several plants situated relatively close together geographically.

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.

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[9], 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 presented in this paper were obtained from simulations representing a historical period. However, potential effects of future climate change on river flow, may be substantial[46]. A first iteration of the AHA under climate change scenarios has been included in its version 2.0 (see "Data availability"). These results were obtained by forcing SWAT+ with simulation results obtained from four bias-adjusted global climate models (GFDL-ESM2M, HadGEM2-ES, ISPL-CM5A-LR and MIROC5). The hydropower availability profiles were calculated separately for each model in the ensemble for the historical period 1975–2005 and for the future period 2075–2100, under three RCP-SSP combinations (Representative Concentration Pathways and Shared Socioeconomic Pathways, respectively): SSP1-RCP2.6, SSP4-RCP6.0, and SSP5-RCP8.5, whereby the land cover scenario is taken from the Land Use Harmonization version 2 data[37] and the global climate models are forced with greenhouse gas concentration trajectories associated with each RCP[47]. We then calculated the relative difference between the results (in terms of monthly availability of hydropower plants) from the future and historical period, and imposed this relative difference on the results obtained for the historical period forced with EWEMBI (see above).

Second, 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. (Two notable exceptions to this rule are Lake Victoria, a natural and mostly rain-fed lake that was later dammed for hydropower generation at its outlet, and Lake Nasser, an artificial lake in the Egyptian desert whose evaporation losses are untypically substantial in comparison to the total water budget.)

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[9].

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. 45.

Fifth, it is to be noted that our approach is statistical rather than deterministic. Instead of modeling actual reservoir dynamics from hour to hour, such as in e.g. refs. 48–50, we used statistics of modelled time series of river flow to infer "typical" seasonal profiles of hydropower availability. The implicit assumption here is that reservoir operation may typically follow a probabilistic approach based on historical experiences.

Sixth, 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 generation[51], or certain hydropower plants where power generation needs to be co-optimised with irrigation or other secondary purposes[52]. Our simulations already account for (existing) reservoir management and irrigation practices, but given the lack of data to constrain the SWAT+ simulations on a site-by-site basis, these practices were modelled using generalised rules on such management. Despite these limitations, we note that this is the first time that an Africa-scale calibrated hydrological model has been generated and run on climate change time scales which includes such considerations of management of dams and irrigation schemes, albeit modelled in a generalised manner.

Seventh, 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.

Eighth, we note that hydrological data are here obtained via simulation with a hydrological model forced with EWEMBI model forcing data, which may introduce additional uncertainties. The EWEMBI dataset was compiled to support the bias-correction of climate input data for the impact assessments carried out in phase 2b of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP)[47] which contributed to the 2018 IPCC special report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, and has as such been widely accepted by the

scientific community. EWEMBI is based on the meteorological dataset ERA-Interim[53], which provides estimates of past weather conditions through numerical weather prediction models, constrained by observations from meteorological or hydrological stations, satellite data, and other sources. While uncertainties in the EWEMBI dataset are indeed larger over the African continent due to sparser observational networks as compared to elsewhere, such reanalyses are still well-constrained by the physical equations that govern weather dynamics, especially for the variable we are interested in, i.e. river discharge, which integrates precipitation over large regions and multiple weeks[47,54]. We also note that the EWEMBI dataset has meanwhile been superseded by the WFDE5 dataset, produced for phase 3 of ISIMIP[55].

Ninth, we note that ideally, all river flow profiles extracted from the SWAT+ simulations would be bias-corrected to data covering the same time period as the simulation period (cf. section 2.4), which is currently not the case. However, we note that an observational data product covering a substantial amount of the locations and a large part of the modelling period considered in this study does, to the authors' knowledge, not exist at this point. This led the authors to consult a wide range of data sources for the bias-correction, whose time periods are not necessarily consistent with one another. Thus, it is important to note that due to the spatiotemporal limitations of the bias-correction as it was performed, this translates into related spatiotemporal inconsistencies in the quality of the AHA as well. We hope that this can be resolved in future iterations of the database.

Related to the above, we note here that other river flow datasets exist which may be used as alternative or as complement to SWAT+. Relying on a single river flow dataset may have limitations, especially for the African continent where the density of the meteorological and hydrological network is relatively low. For this reason, the authors of this paper recommend practitioners to also consider such datasets as input to AHA-type assessments to more comprehensively clarify the uncertainty associated to hydrological data and simulations. An example of a complementary dataset may be the the GloFas-ERA5 operational global river discharge reanalysis (1979-present)[56].

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 visualization 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 avoided[19]; 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-dependency[23].

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 (v2.0). https://www.hydroshare.org/resource/5e8ebdc3bfd24207852539ecf219d915[33].

This project contains the following underlying data:

- The AHA (v2.0) 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). The historical runs based on EWEMBI observations are entitled "SWAT+_channel_mon_EWEMBI_hist" and "SWAT+_reservoir_mon_EWEMBI_hist". We refer to the SWAT+ output documentation (accessible through https://swatplus.gitbook.io/docs/download-docs) for further metadata on the columns included in this .txt file.
- SWAT+ simulation results based on runs from an ensemble of global climate models (GCMs). The channel and reservoir .txt files are given in the zipped folders "SWAT+_simulations_GCM_historical.rar" and "SWAT+_simulations_GCM_ssp_rcp.rar".
- 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).

The data/metadata included in the AHA spreadsheet is summarised in the table below.

Parameter	Explanation	Unit
Country	The country in which the hydropower plant is located.	
Unit Name	Name of the hydropower plant as recorded.	
Status	Existing, Committed, Candidate or Planned	
Latitude	Self-explanatory	°N
Longitude	Self-explanatory	°E
River Name	The name of the river/tributary on which the hydropower plant is located	
River Basin	The name of the river basin that the hydropower plant is part of, in the widest sense of the term "river basin" (encompassing all streams whose discharge eventually drain at the same outlet at the ocean or at an endorheic lake).	
Spill From	For cascade systems : the nearest upstream plant	
River Channel ID	The SWAT+ simulation ID of the river channel on which the hydropower plant is built	
Reservoir ID	The SWAT+ simulation ID of existing reservoirs or future plants directly downstream of existing plants	
Reservoir Size	Full (live + dead) storage volume of the hydropower plant	million m^3
Mean Annual Discharge	Recorded mean annual discharge values at or very near the hydropower plant site used for bias-correction	m^3/s
Filling Time	Time it would take the mean annual inflow to completely fill the reservoir storage	days
Design Discharge	Discharge at which the hydroturbines would operate at maximum capacity (assuming constant head)	m^3/s
Availability	Wherever recorded: multiannual average capacity factor. Default taken to be 50%.	fraction
First Year	Year of entry into service (for future plants : assumed *earliest possible* year)	
Capacity	Rated capacity	MW
Size Type	Small (<10 MW), Medium (<100 MW) or Large (≥100 MW)	
Shared Plant	Describes whether the electricity generated by the plant is shared between different countries.	
Shared Plant Split	In case the electricity generated by the plant is shared between different countries, this column describes whether the plant has been allocated to a single country or split across multiple countries in the database (e.g. whether it is physically located in only one or in multiple countries). In case of the latter, a separate entry is included for each of the *N* involved countries with a fraction *1/N* of the plant's capacity allocated to each country.	
Single Source	Source in cases where a single source was consulted for all recorded data. If not applicable, the following columns (Source_status, source_location, etc.) were filled with the respective sources.	

Code availability

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

Archived analysis code at time of publication: https://zenodo.org/record/5777129[57].

License: MIT

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 authors further thank A. Carlino (Politecnico di Milano & IIASA) for constructive criticism that helped to improve the database.

The map in Figure 3 was created using the Esri Satellite Basemap[32] 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

1. Kuramochi T, Höhne N, Schaeffer M, *et al.*: **Ten key short-term sectoral benchmarks to limit warming to 1.5°C.** *Clim Policy.* 2018; **18**(3): 287–305.
Publisher Full Text
2. IRENA: **Renewable Power Generation Costs in 2019.** International Renewable Energy Agency, Abu Dhabi, 2020; Accessed: Sep. 10, 2020.
Reference Source
3. Bogdanov D, Farfan J, Sadovskaia K, *et al.*: **Radical transformation pathway towards sustainable electricity via evolutionary steps.** *Nat Commun.* 2019; **10**(1): 1077.
PubMed Abstract | Publisher Full Text | Free Full Text
4. Engeland K, Borga M, Creutin JD, *et al.*: **Space-time variability of climate variables and intermittent renewable electricity production – A review.** *Renew Sustain Energy Rev.* 2017; **79**: 600–617.
Publisher Full Text
5. IRENA: **Innovation landscape for a renewable-powered future: Solutions to integrate variable renewables.** International Renewable Energy Agency, Abu Dhabi, 2019.
Reference Source
6. Gils HC, Scholz Y, Pregger T, *et al.*: **Integrated modelling of variable renewable energy-based power supply in Europe.** *Energy.* 2017; **123**: 173–188.
Publisher Full Text
7. Hirth L: **The benefits of flexibility: The value of wind energy with hydropower.** *Appl Energy.* 2016; **181**: 210–223.
Publisher Full Text
8. IEA Hydropower: **Flexible hydropower providing value to renewable energy integration**. 2019.
Reference Source
9. Sterl S, Vanderkelen I, Chawanda CJ, *et al.*: **Smart renewable electricity portfolios in West Africa.** *Nat Sustain.* 2020; **3**(9): 710–719.
Publisher Full Text
10. Jacobson MZ, Delucchi MA, Cameron MA, *et al.*: **Low-cost solution to the grid reliability problem with 100% penetration of intermittent wind, water, and solar for all purposes**. *Proc Natl Acad Sci U S A.* 2015; **112**(49): 15060–5.
PubMed Abstract | Publisher Full Text | Free Full Text
11. Schillinger M, Weigt H, Hirsch PE: **Environmental flows or economic woes-Hydropower under global energy market changes.** *PLoS One.* 2020; **15**(8): e0236730.
PubMed Abstract | Publisher Full Text | Free Full Text
12. IRENA: **Innovation landscape brief - Innovative operation of pumped hydropower storage**. 2020.
Reference Source
13. Welsch M, Deane P, Howells M, *et al.*: **Incorporating flexibility requirements into long-term energy system models – A case study on high levels of renewable electricity penetration in Ireland.** *Appl Energy.* 2014; **135**: 600–615.
Publisher Full Text
14. World Bank Group: **Global Solar Atlas**. 2020.
Reference Source
15. World Bank Group: **Global Wind Atlas**. 2020.
Reference Source
16. Pfenninger S, Staffell I: **Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data.** *Energy.* 2016; **114**: 1251–1265.
Publisher Full Text
17. Mattsson N, Verendel V, Hedenus F, *et al.*: **An autopilot for energy models – Automatic generation of renewable supply curves, hourly capacity factors and hourly synthetic electricity demand for arbitrary world regions.** *Energy Strategy Rev.* 2021; **33**: 100606.
Publisher Full Text
18. Linke S, Lehner B, Dallaire CO, *et al.*: **Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution.** *Sci Data.* 2019; **6**(1): 283.
PubMed Abstract | Publisher Full Text | Free Full Text
19. Moran EF, Lopez MC, Moore N, *et al.*: **Sustainable hydropower in the 21st century.** *Proc Natl Acad Sci U S A.* 2018; **115**(47): 11891–11898.
PubMed Abstract | Publisher Full Text | Free Full Text
20. Barasa M, Bogdanov D, Oyewo AS, *et al.*: **A cost optimal resolution for Sub-Saharan Africa powered by 100% renewables in 2030.** *Renew Sustain Energy Rev.* 2018; **92**: 440–457.
Publisher Full Text
21. Oyewo AS, Aghahosseini A, Ram M, *et al.*: **Transition towards decarbonised power systems and its socio-economic impacts in West Africa.** *Renew Energy.* 2020; **154**: 1092–1112.
Publisher Full Text
22. IRENA: **Planning and prospects for renewable power: West Africa.** International Renewable Energy Agency, Abu Dhabi, 2018.
Reference Source
23. Falchetta G, Gernaat DEHJ, Hunt J, *et al.*: **Hydropower dependency and climate change in sub-Saharan Africa: A nexus framework and evidence-based review.** *J Clean Prod.* 2019; **231**: 1399–1417.
Publisher Full Text
24. Cole MA, Elliott RJR, Strobl E: **Climate Change, Hydro-Dependency, and the African Dam Boom.** *World Dev.* 2014; **60**: 84–98.
Publisher Full Text
25. Buontempo C, Booth B, Moufouma-Okia W: **The climate of the Sahel**. 2012; 58–72.
Publisher Full Text
26. Vogel C, Steynor A, Manyuchi A: **Climate services in Africa: Re-imagining an inclusive, robust and sustainable service.** *Clim Serv.* 2019; **15**: 100107.
Publisher Full Text
27. Lehner B, Liermann CR, Revenga C, *et al.*: **High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management.** *Front Ecol Environ.* 2011; **9**(9): 494–502.
Publisher Full Text
28. FAO: **Geo-referenced database on dams in Africa**. 2016.
Reference Source
29. IRENA: **Renewable Capacity Statistics 2020**. International Renewable Energy Agency, Abu Dhabi, 2020; Accessed: Apr. 06, 2018.
Reference Source
30. Bieger K, Arnold JG, Rathjens H, *et al.*: **Introduction to SWAT+, A Completely Restructured Version of the Soil and Water Assessment Tool.** *JAWRA J Am Water Resour Assoc.* 2017; **53**(1): 115–130.
Publisher Full Text
31. Chawanda CJ, van Griensven A, Thiery W, *et al.*: **SWAT+ simulation result used in 'Smart renewable electricity portfolios in West Africa**. 2019.
Reference Source
32. ESRI: **Satellite World Imagery [basemap]**. 2009.
33. Sterl S, Chawanda CJ: **Online repository of materials for an all-Africa hydropower atlas (v2.0)**. HydroShare, 2021.
https://www.hydroshare.org/resource/e8ebdc3bfd24207852539ecf219d915
34. Chawanda CJ, George C, Thiery W, *et al.*: **User-friendly workflows for catchment modelling: Towards reproducible SWAT+ model studies.** *Environ Model Softw.* 2020; **134**: 104812.
Publisher Full Text
35. Chawanda CJ, Arnold J, Thiery W, *et al.*: **Mass balance calibration and reservoir representations for large-scale hydrological impact studies using SWAT+.** *Clim Change.* 2020; **163**: 1307–1327.
Publisher Full Text
36. Farr TG, Rosen PA, Caro E, *et al.*: **The Shuttle Radar Topography Mission.** *Rev Geophys.* 2007; **45**(2).
Publisher Full Text
37. Hurtt GC, Chini LP, Frolking S, *et al.*: **Harmonization of land-use scenarios for the period 1500-2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands.** *Clim Change.* 2011; **109**(1): 117.
Publisher Full Text
38. Siebert S, Frenken K: **Irrigated areas Atlas of African agriculture research & development.** 2014; 18–19.
Reference Source
39. Leenaars JGB: **Africa Soil Profiles Database, version 1.1: a compilation of georeferenced and standardised legacy soil profile data for Sub-Saharan Africa (with dataset)**. Africa Soil Information Service (AfSIS) project, ISRIC. 2013.
Publisher Full Text
40. Lange S: **EartH2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI).** GFZ Data Services, 2016.
Publisher Full Text
41. Allen RG, Pereira L, Raes D, *et al.*: **Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements. FAO Irrigation and Drainage Paper 56.** Rome, Italy: Food and Agriculture Organisation, 1998.
Reference Source
42. Mishra SK, Singh VP: **SCS-CN Method.** In: *Soil Conservation Service Curve Number (SCS-CN) Methodology*. Water Science and Technology Library. Springer, Dordrecht, 2003; **42**: 84–146.
Publisher Full Text
43. Williams JR: **Flood routing with variable travel time or variable storage coefficients.** *Transactions of the ASAE.* 1969; **12**(1): 100–103.
Publisher Full Text
44. Chen J, Brissette FP, Chaumont D, *et al.*: **Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America.** *Water Resour Res.* 2013; **49**(7): 4187–4205.
Publisher Full Text
45. Sterl S, Fadly D, Liersch S, *et al.*: **Linking solar and wind power in eastern Africa with operation of the Grand Ethiopian Renaissance Dam.** *Nat Energy.* 2021; **6**: 407–418.
Publisher Full Text
46. Conway D, Dalin C, Landman WA, *et al.*: **Hydropower plans in eastern and**

southern Africa increase risk of concurrent climate-related electricity supply disruption. *Nat Energy.* 2017; **2**: 946–953.
Publisher Full Text

47. Frieler K, Lange S, Piontek F, *et al.*: **Assessing the impacts of 1.5 °C global warming – simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b).** *Geosci Model Dev.* 2017; **10**(12): 4321–4345.
Publisher Full Text

48. Minville M, Brissette F, Krau S, *et al.*: **Adaptation to Climate Change in the Management of a Canadian Water-Resources System Exploited for Hydropower.** *Water Resour Manage.* 2009; **23**(14): 2965–2986.
Publisher Full Text

49. Danso DK, François B, Hingray B, *et al.*: **Assessing hydropower flexibility for integrating solar and wind energy in West Africa using dynamic programming and sensitivity analysis. Illustration with the Akosombo reservoir, Ghana.** *J Clean Prod.* 2021; **287**: 125559.
Publisher Full Text

50. Turner SWD, Ng JY, Galelli S: **Examining global electricity supply vulnerability to climate change using a high-fidelity hydropower dam model.** *Sci Total Environ.* 2017; **590–591**: 663–675.
PubMed Abstract | Publisher Full Text

51. Jägermeyr J, Pastor A, Biemans H, *et al.*: **Reconciling irrigated food production with environmental flows for Sustainable Development Goals implementation.** *Nat Commun.* 2017; **8**: 15900.
PubMed Abstract | Publisher Full Text | Free Full Text

52. Zeng R, Cai X, Ringler C, *et al.*: **Hydropower versus irrigation—an analysis of global patterns.** *Environ Res Lett.* 2017; **12**(3): 034006.
Publisher Full Text

53. Dee DP, Uppala SM, Simmons A, *et al.*: **The ERA-Interim reanalysis: configuration and performance of the data assimilation system.** *Q J R Meteorol Soc.* 2011; **137**(656): 553–97.
Publisher Full Text

54. Lange S: **Bias correction of surface downwelling longwave and shortwave radiation for the EWEMBI dataset.** *Earth Syst Dynam.* 2018; **9**(2): 627–645.
Publisher Full Text

55. Cucchi M, Weedon GP, Amici A, *et al.*: **WFDE5: bias-adjusted ERA5 reanalysis data for impact studies.** *Earth Syst Sci Data.* 2020; **12**(3): 2097–2120.
Publisher Full Text

56. Harrigan S, Zsoter E, Alfieri L, *et al.*: **GloFAS-ERA5 operational global river discharge reanalysis 1979–present.** *Earth Syst Sci Data.* 2020; **12**(3): 2043–2060.
Publisher Full Text

57. Sterl S, Devillers A: **Archived code for the paper "A spatiotemporal atlas of hydropower in Africa for energy modelling purposes"**. *Zenodo*. 2021.
https://zenodo.org/record/5777129

Open Peer Review

Current Peer Review Status: ✔ ? ✔

Version 3

Reviewer Report 07 April 2022

https://doi.org/10.21956/openreseurope.15833.r28953

© **2022 Vidal J.** 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.

✔

Jean-Philippe Vidal
UR RIVERLY, INRAE, Lyon, France

The authors have in my opinion satisfactorily responded to the comments made during my first revision.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Hydroclimatogy, climate variability, climate change, drought and low-flows, reservoir management

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.

Version 2

Reviewer Report 24 February 2022

https://doi.org/10.21956/openreseurope.15526.r28473

© **2022 Vidal J.** 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.

?

Jean-Philippe Vidal
UR RIVERLY, INRAE, Lyon, France

Overall comments

This data note describes an atlas of hydropower in Africa. This atlas, compiling existing databases on existing and planned hydropower plants, and continental-scale hydrological modelling, provides information on the hydropower generation profiles at all these plant/reservoir locations. It is therefore of high interest for continental-scale energy modelling. The revised version of the manuscript has been much improved following the first-round comments of reviewers, and future projections are welcome. However, there are still places in the text and specific points that lack clarity, and some methodological issues would also require clarifications and additions to the text and the database. They are detailed below.

Main issues

- *Section 2.3, quality of SWAT+ simulations*: The responses to comments by Benoît Hingray on this specific issues are in my view not satisfying. Hydrological modelling is in this study used to get reasonable estimates of seasonal and interannual variability of streamflow across the continent. I therefore strongly believe that an appropriate assessment of such properties is definitely required for one to be able to gain confidence in the AHA results. One may think of monthly NSE or monthly NSE with respect to the average regime, to check the ability of SWAT+ to get interannual variability right. This assessment should be included in the AHA database where available, i.e. where bias-correction has been done. More on the bias-correction below.

- *Section 2.4, "a typical range of years"*: It should be specified here what kind of years are considered here. Civil years? Hydrological years (preferably)? If hydrological years, what is the starting date and is it homogeneous over the continent? This choice is rather critical as it may have strong implications and it should be necessary justified and commented in the discussion part.

- *Section 2.4, "First, the flow profile..."*: here, in spite of modifications made after the first round of comments, I am still unclear on how the final monthly regimes of normal, dry and wet are obtained. More precisely it is not clear what is the "monthly median of the dataset". I would tend to understand that you took the median interannual value for each month of the year and use it to build a "normal" 12-value regime. And in parallel, you considered the distribution of annual-averaged values and computed the ratio between the median value and the 5/95 percentiles to get the correction factors. And that you subsequently applied these correction factors to all 12 values of the "normal" regime. Am I right? But then the following paragraph with the scaling made me once more hesitating on what is really done. I guess a schematic of the time series processing is strongly required here readability and reproducibility purposes. The text should definitely be clear on that, as this way of taking account of inter-annual and intra-annual variability relies on strong hypotheses that would not necessarily be shared by all users.

- *Section 2.4, bias-correction*: I believe that the answers made to Benoît Hingray on this issue are only partly satisfying. Indeed, while I understand and approve the need for bias-correcting (for this specific application), there are two points that require actions on the manuscript. First, the bias-correction (scaling) is made to existing data on multi-annual mean river discharge, but nothing is said on the period used for this bias-correction, and specifically whether, for a given location, the period for comparing observed and simulated multi-annual averages is the same. This is particularly important in (some parts of) Africa where the multidecadal variability is very high and could therefore lead to unwanted bias-

correction effects. Clarifications on the methodology in the text about this point are essential, and the corresponding period should be added to the database. Second, the consequences of this bias-correction for only specific locations should be discussed in the manuscript as it generates de facto possible longitudinal (but limited as I understand form the responses) but clearly spatial inconsistencies in the quality of simulations and thus in corresponding AHA products. This comment should be added to the manuscript.

- *Section 2.5.2, storable component*: I am sorry to say that in spite of my careful reading of the responses to Benoît Hingray and the modifications made to the manuscript, the intra-annual behaviour of "storable component" hypothesized here still quite unclear. Here again, as asked by the previous reviewer, a schematic would be greatly appreciated. This is all the more important for the end-users of this database, for the hypotheses (which are very strong here for this separation between storable and non-storable components) to be correctly understood and apprehended. This is all the more important that the article referred to by the authors on this point is not in Open Access, and not available to my institution (Sterl *et al*., 2020[1]).

- *Section 3, irrigation modelling*: I was quite surprised to learn here that the hydrological simulations take (in some way) irrigation into account while nothing is said about it in the appropriate section 2.3! Such an important information should be given (as well as appropriate references) in the description of the river flow dataset.

Minor issues

- Section 2.1, “multiannual average capacity factor”: the capacity factor should be clearly defined here.

- Section 2.3, “Meteorological forcing”: it should be mentioned (probably in the discussion part) that the EWEMBI dataset (Lange, 2016) used here has now been superseded by WFDE5 (Cucchi *et al*., 2020[2]) produced for phase 3 of ISIMIP (https://www.isimip.org/protocol/3/).

- Table, entry “first year”: → “earliest possible year”.

References

1. Sterl S, Vanderkelen I, Chawanda C, Russo D, et al.: Smart renewable electricity portfolios in West Africa. *Nature Sustainability*. 2020; **3** (9): 710-719 Publisher Full Text
2. Cucchi M, Weedon G, Amici A, Bellouin N, et al.: WFDE5: bias-adjusted ERA5 reanalysis data for impact studies. *Earth System Science Data*. 2020; **12** (3): 2097-2120 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?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Hydroclimatogy, climate variability, climate change, drought and low-flows, reservoir management

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.

Author Response 22 Mar 2022

Sebastian Sterl

This data note describes an atlas of hydropower in Africa. This atlas, compiling existing databases on existing and planned hydropower plants, and continental-scale hydrological modelling, provides information on the hydropower generation profiles at all these plant/reservoir locations. It is therefore of high interest for continental-scale energy modelling. The revised version of the manuscript has been much improved following the first-round comments of reviewers, and future projections are welcome. However, there are still places in the text and specific points that lack clarity, and some methodological issues would also require clarifications and additions to the text and the database. They are detailed below.

Author Response: We thank the reviewer for agreeing to provide feedback on our revised manuscript, and for their positive comments and constructive criticism. We provide point-by-point replies to the issues raised by the reviewer below.

Section 2.3, quality of SWAT+ simulations: The responses to comments by Benoît Hingray on this specific issues are in my view not satisfying. Hydrological modelling is in this study used to get reasonable estimates of seasonal and interannual variability of streamflow across the continent. I therefore strongly believe that an appropriate assessment of such properties is definitely required for one to be able to gain confidence in the AHA results. One may think of monthly NSE or monthly NSE with respect to the average regime, to check the ability of SWAT+ to get interannual variability right. This assessment should be included in the AHA database where available, i.e. where bias-correction has been done. More on the bias-correction below.

Author Response: We appreciate the concern of the reviewer on this point. As we mentioned in the response to the comments by Benoît Hingray, we currently have a study in preparation which evaluates the used SWAT+ modelling results for the entire African continent (Chawanda *et al*., 2022, in preparation). These results are already part of a successfully defended PhD thesis at the Vrije Universiteit Brussel (VUB) in Brussels, Belgium (thesis entitled "*Modelling the Impact of Climate Change & Land-Use/Management Changes on African Water Resources*" by Dr. Celray James Chawanda, 2021); this chapter will become available on the university's repository webpage in the coming months. One of the indicators for quality here is, indeed, the monthly NSE as suggested by this reviewer. Here, on an all-Africa scale, the monthly NSE is positive for roughly 60% of river sites on all-Africa scale. Once the above-mentioned study is published, we agree that the information on monthly NSE on a plant-by-plant basis could ideally be included in a future version of the AHA. We thank the reviewer for this useful suggestion. We acknowledge the fact that the

cited 60% may not be considered as sufficiently high by some hydrologists, but we again would like to reiterate the fact that, as mentioned in our previous response, we find it encouraging that these 60% of positive NSE on an all-Africa scale can be obtained using the hydrological mass-balance calibration (HMBC) approach (Chawanda *et al*., 2020) which focuses on reproducing climatological mean water balance components (e.g. annual mean evaporation) and avoids the need of having to do time-series calibration. This avoids an overfitting of the model, which would yield higher NSEs but not necessarily result in a meaningful model at the continental scale.

Section 2.4, "a typical range of years": It should be specified here what kind of years are considered here. Civil years? Hydrological years (preferably)? If hydrological years, what is the starting date and is it homogeneous over the continent? This choice is rather critical as it may have strong implications and it should be necessary justified and commented in the discussion part.
Author Response: We thank the reviewer for this observation. In fact, the sentence was poorly worded. It should have read "*a typical range of seasonal profiles*". This change has now been made. (The range between what we call "very dry" and "very wet" years spans is based on 5^{th} and 95^{th} percentile values of average annual flow, as the section explains, which thus represents dryness/wetness statistically occurring once every 20 years.)

Section 2.4, "First, the flow profile…": here, in spite of modifications made after the first round of comments, I am still unclear on how the final monthly regimes of normal, dry and wet are obtained. More precisely it is not clear what is the "monthly median of the dataset". I would tend to understand that you took the median interannual value for each month of the year and use it to build a "normal" 12-value regime. And in parallel, you considered the distribution of annual-averaged values and computed the ratio between the median value and the 5/95 percentiles to get the correction factors. And that you subsequently applied these correction factors to all 12 values of the "normal" regime. Am I right? But then the following paragraph with the scaling made me once more hesitating on what is really done. I guess a schematic of the time series processing is strongly required here readability and reproducibility purposes. The text should definitely be clear on that, as this way of taking account of inter-annual and intra-annual variability relies on strong hypotheses that would not necessarily be shared by all users.
Author Response: The interpretation of the reviewer is 100% correct. However, we appreciate the fact that the discussion on scaling was not fully clear, so we have changed the sentence "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 (simple scaling)." to "Third, the seasonal river flow profile thus obtained (for very dry, normal, and very wet years), each characterised as a time series of twelve values representing the months of the year, was normalised to a dimensionless time series by dividing each time series by the simulated multiyear average flow (a single number)." Additionally, the opening sentence of the following paragraph, "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)." was changed to "Fourth, wherever possible, the three resulting dimensionless time series of normalised river inflow to each hydropower plant were multiplied again with a bias-correction factor (simple scaling), equal to the multiannual mean river discharge value collected from existing

databases and literature (again, a single number; see section 2.1)."

Section 2.4, bias-correction: I believe that the answers made to Benoît Hingray on this issue are only partly satisfying. Indeed, while I understand and approve the need for bias-correcting (for this specific application), there are two points that require actions on the manuscript. First, the bias-correction (scaling) is made to existing data on multi-annual mean river discharge, but nothing is said on the period used for this bias-correction, and specifically whether, for a given location, the period for comparing observed and simulated multi-annual averages is the same. This is particularly important in (some parts of) Africa where the multidecadal variability is very high and could therefore lead to unwanted bias-correction effects. Clarifications on the methodology in the text about this point are essential, and the corresponding period should be added to the database. Second, the consequences of this bias-correction for only specific locations should be discussed in the manuscript as it generates de facto possible longitudinal (but limited as I understand form the responses) but clearly spatial inconsistencies in the quality of simulations and thus in corresponding AHA products. This comment should be added to the manuscript.

Author Response: We appreciate this comment from the reviewer, and we agree with them. It is important to note here that a bias-correction approach that would allow (i) **all** locations in the dataset to undergo bias-correction, for (ii) the **same** time period in each case, simply does not exist, this due to the strongly varying periods for which observations are available across the African reference stations. A bias-correction for a limited subset of data points, using time periods that are not necessarily all of the same length and not necessarily consistent with the modelling period, was considered by the authors to be the "next best option", despite clearly not being ideal. It is important to understand that the sources consulted to obtain these data were not necessarily from scientific literature. In many cases, the consulted sources were technical project documents or articles which simply state an average flow rate "as is", without providing the specific period to which this pertains. We agree that this is not ideal, but it is also the "best we have", in many cases. For this reason, we do not provide the "corresponding period" explicitly in the database. However, all the data sources consulted for obtaining the average flow rate for bias-correction are documented in the database. We have adapted the sentence "We note that this step is not crucial for the data product and serves mostly for refinement of final numbers, as monthly availability curves of hydropower plants could be readily estimated without performing the bias-correction step, i.e. for the remaining 40% of cases (see section 2.5.1)." in the paragraph preceding section 2.5 to "We note that this step is not crucial for the data product and serves mostly for providing a dimension to the river flow seasonality (expressing it in m^3/s instead of dimensionless, which allows a comparison to e.g. design discharges of hydropower plants), as the monthly availability curves of hydropower plants could still be estimated without performing the bias-correction step, i.e. for the remaining 40% of cases (see section 2.5.1)." We have also added the following paragraph to the discussion: "Ninth, we note that ideally, all river flow profiles extracted from the SWAT+ simulations would be bias-corrected to data covering the same time period as the simulation period (cf. section 2.4), which is currently not the case. However, we note that an observational data product covering a substantial amount of the locations and a large part of the modelling period considered in this study does, to the authors' knowledge, not exist at this point. This led the authors to consult a wide range of data sources for the bias-correction, whose time periods are not necessarily consistent with one another. Thus, it is important to note that due to the spatiotemporal limitations of the bias-correction as it was

performed, this translates into related spatiotemporal inconsistencies in the quality of the AHA as well. We hope that this can be resolved in future iterations of the database."

Section 2.5.2, storable component: I am sorry to say that in spite of my careful reading of the responses to Benoît Hingray and the modifications made to the manuscript, the intra-annual behaviour of "storable component" hypothesized here still quite unclear. Here again, as asked by the previous reviewer, a schematic would be greatly appreciated. This is all the more important for the end-users of this database, for the hypotheses (which are very strong here for this separation between storable and non-storable components) to be correctly understood and apprehended. This is all the more important that the article referred to by the authors on this point is not in Open Access, and not available to my institution (Sterl et al., 2020).
Author Response: Let us say a hypothetical hydropower reservoir has a (live) volume of 2, but a river carries an annual volume of 8 in a "normal" year. In this case, the "storable component" of river flow is equal to 2 (25%, or 2 out of 8), and the "non-storable component" of river flow is the remainder, i.e. 6 (75%). We have added this example to the text to clarify this idea. We hope that this wording sufficiently clarifies this concept and that a schematic is therefore not needed. In response to the comment that the referenced article is not available to the reviewer's institution, we are happy to report that the manual of the REVUB software, which was used in that work, contains the same discussion of "storable" and "non-storable" components, and is available in open-access on https://github.com/VUB-HYDR/REVUB/blob/master/manual/REVUB_manual.pdf (see "Note 5").

Section 3, irrigation modelling: I was quite surprised to learn here that the hydrological simulations take (in some way) irrigation into account while nothing is said about it in the appropriate section 2.3! Such an important information should be given (as well as appropriate references) in the description of the river flow dataset.
Author Response: The implementation of existing irrigation schemes in the SWAT+ model is fully described in the reference (Chawanda et al., 2020), and section 2.3 was only meant to repeat some of the most important elements of the simulations as described in the latter paper. However, we appreciate the reviewer's concern for clarity on the implementation of irrigation, and have added the following bullet point to the list of data sources cited in section 2.3: "- Irrigated areas were obtained from Food and Agriculture Organisation at 0.083° × 0.083° resolution (FAO; Siebert and Frenken 2014). Irrigation was implemented as explained in Chawanda et al. (2020)."

Section 2.1, "multiannual average capacity factor": the capacity factor should be clearly defined here.
Author Response: We have added the following text: "Here, the capacity factor of any power plant is taken to refer to the electricity that it generated over a certain time period (e.g. hour, day, season, year…) divided by the power that would have been generated if the plant had run at full capacity over that entire period."

Section 2.3, "Meteorological forcing": it should be mentioned (probably in the discussion part) that the EWEMBI dataset (Lange, 2016) used here has now been superseded by WFDE5 (Cucchi et al., 20202) produced for phase 3 of ISIMIP (https://www.isimip.org/protocol/3/). Thank you for this suggestion.
Author Response: We have added the following text to the eighth point of the discussion: "

We also note that the EWEMBI dataset has meanwhile been superseded by the WFDE5 dataset, produced for phase 3 of ISIMIP" with the appropriate reference.

Table, entry "first year": → "earliest possible year".
Author Response: We prefer to keep it as "first year", since it refers also to existing plants, whose first year of service is usually well-defined. However, we noticed a spelling error in the table caption (possibe à possible), which has been corrected.

Competing Interests: No competing interests were disclosed.

Reviewer Report 19 January 2022

https://doi.org/10.21956/openreseurope.15526.r28223

© 2022 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

The revised version has addressed all my previous remarks.

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.

Version 1

Reviewer Report 10 June 2021

https://doi.org/10.21956/openreseurope.14462.r26660

© 2021 Hingray B. 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.

Benoît Hingray
Grenoble-INP, IGE, Université Grenoble Alpes, Grenoble, France

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 clarified.

- 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 obvioulsy 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 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[2]; Turner *et al*. 2017[3]; Danso *et al.* 2021[4]).

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[5]; reservoirs can contribute significantly to irrigation water supply in many regions and significanly 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 sothat 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[9] of more than one full year. For these plants, instead of the 5^{th} and 95^{th} percentiles, the 10^{th} and 90^{th} 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 95h 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[6]; Bichet *et al*. 2020[7]; Moussa *et al*. 2020[8]).

SWAT+ : The description of SWAT+ has to be given in more details. The authors should give at least the references for the Penman-Montheith, 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

1. Harrigan S, Zsoter E, Alfieri L, Prudhomme C, et al.: GloFAS-ERA5 operational global river discharge reanalysis 1979–present. *Earth System Science Data*. 2020; **12** (3): 2043-2060 Publisher Full Text
2. Minville M, Brissette F, Krau S, Leconte R: Adaptation to Climate Change in the Management of a Canadian Water-Resources System Exploited for Hydropower. *Water Resources Management*. 2009; **23** (14): 2965-2986 Publisher Full Text
3. Danso D, François B, Hingray B, Diedhiou A: Assessing hydropower flexibility for integrating solar and wind energy in West Africa using dynamic programming and sensitivity analysis. Illustration with the Akosombo reservoir, Ghana. *Journal of Cleaner Production*. 2021; **287**. Publisher Full Text
4. Turner SWD, Ng JY, Galelli S: Examining global electricity supply vulnerability to climate change using a high-fidelity hydropower dam model.*Sci Total Environ*. 2017; **590-591**: 663-675 PubMed Abstract | Publisher Full Text
5. Biemans H, Haddeland I, Kabat P, Ludwig F, et al.: Impact of reservoirs on river discharge and irrigation water supply during the 20th century. *Water Resources Research*. 2011; **47** (3). Publisher Full Text
6. Stanzel P, Kling H, Bauer H: Climate change impact on West African rivers under an ensemble of CORDEX climate projections. *Climate Services*. 2018; **11**: 36-48 Publisher Full Text
7. Bichet A, Diedhiou A, Hingray B, Evin G, et al.: Assessing uncertainties in the regional projections of precipitation in CORDEX-AFRICA. *Climatic Change*. 2020; **162** (2): 583-601 Publisher Full Text
8. Sidibe M, Dieppois B, Eden J, Mahé G, et al.: Near-term impacts of climate variability and change on hydrological systems in West and Central Africa. *Climate Dynamics*. 2020; **54** (3-4): 2041-2070 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.

Author Response 13 Dec 2021

Sebastian Sterl

(Author Responses in italics)
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.
Author Response: *We thank the reviewer for their positive words about our manuscript, and are grateful for the constructive and very extensive criticism which we have done our best to take into account.*

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 acknowledged. 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.

Author Response: *We thank the reviewer for this comment. We have added an eighth point of discussion in the "Discussion" section, which acknowledges the diversity of data of EWEMBI but notes also the wide scientific acceptance of the EWEMBI dataset. We note here also a few additional points as to why we consider the EWEMBI dataset to be the go-to dataset for this undertaking, specifically the fact that (i) despite the relative scarcity of observational stations on the African continent, reanalyses are well-constrained by the physical equations that govern weather dynamics, and (ii) the variables we are interested in integrate precipitation data over large regions and multiple weeks.*

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 »).

Author Response: *We thank the reviewer for this comment. In classical hydrology, one would use observational time series for calibration. For computational reasons, this approach is not feasible on continental-scale applications such as the AHA. This is why we developed hydrological mass-balance calibration (HSMBC, see Chawanda et al. 2020), which focuses on reproducing long-term mean water balance terms (e.g., evapotranspiration, baseflow) and which is showcased for South Africa in the cited paper. Even though this may give lower quality (expressed e.g. with the NASH efficiency criterion) than a time-series calibration, it has the major advantage that it is applicable on continental scales without overfitting the hydrological model everywhere and at the same time allowing for considerable computational benefits. Given this comparatively simple approach to hydrological model calibration, we consider it encouraging that the NASH efficiency criterion here is positive for roughly 60% of river sites on all-Africa scale. We currently have a study in preparation which evaluates and applies this method to the whole of Africa (Chawanda et al., in prep.). In addition, for the specific purposes of the AHA, we note that we performed an additional bias-correction of river discharge at hydropower / reservoir sites based on observational time series, as*

described in the paper (section 2.4).

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.
Author Response: *Thank you for this suggestion. Our SWAT+ simulations currently account for (existing) reservoir management and irrigation practices in a simplified way. Given the lack of data to constrain the SWAT+ simulations on a site-by-site basis, these practices were modelled using generalised rules on such management. Despite these limitations, we note that this is the first time that an Africa-scale calibrated hydrological model has been generated and run on climate change time scales which includes such considerations of management of dams and irrigation schemes, albeit modelled in a generalised manner. More details are provided in Chawanda et al. 2020. We have included this explanation in section 3 of the revised paper.*

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.
Author Response: *We thank the reviewer for this useful addition. In the revised version of the manuscript, we now cite the proposed datasets as potential alternatives to the SWAT+ dataset used in the present study, including the recommendation to undertake similar analyses in the future based on these datasets for a more comprehensive uncertainty assessment and the rationale behind this (low density of met/hydro station network on the African continent). We also note that an upcoming publication (Chawanda et al., in prep.) analyses the SWAT+ results in details and includes an uncertainty assessment. While too late to be cited in the current AHA study, we hope that this information will be available to the scientific community in due course.*

Bias correction. P7 C2 § 3. The principle / interest / limitation of bias correction have to be clarified. 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?
Author Response: *The added value of the simulations is to provide data on (i) seasonality and (ii) interannual variability of river flow, since the bias correction is based only on observed/cited long-term mean discharges. Clearly, one needs observations on seasonalities and interannual*

variabilities to infer long-term mean discharges, but the simple fact is that the latter are widely available in literature and public datasets, whereas the former are usually not. We note here that our methodology, as explained in the paper, is also able of estimating typical capacity factor profiles on seasonal level even for hydropower plants where no long-term mean discharges are available to bias-correct the simulated data. Effectively, the bias-correction is just a "final step", wherever possible, to refine somewhat the simulated patterns.

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)?
Author Response: *Please see our response to the above comment. Bias-correction is done on the basis of multi-annual mean discharges (see p.7, right column, third section). Monthly river flow time series are not publicly available to an extent that would have permitted monthly basis bias-correction (note, in this context, that we attempted to rely solely on publicly available information to safeguard the open-access character of the hydropower atlas, so we did not consider restricted or even commercial acquisition of river flow data to be an option).*

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.
Author Response: *Bias-correction was only performed for cases where either (i) mean discharge values were available for the hydropower site in question (e.g. from environmental impact assessments where measurements directly at the hydropower plant construction site were invoked), or (ii) mean discharge values were available from gauging stations located very close by. We realise that "very close by" contains a degree of subjectivity, but in practice, it means that for cases cited by the reviewer, bias-correction was abstained from for such "intermediate stations" since no accurate measurements on long-term mean discharge would have been available. A comment on this has been included in the manuscript (end of section 2.4).*

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?)
Author Response: *We did not work with a specific reference period for the bias-correction, since it is practically impossible to find data on multiannual mean discharge covering exactly the same period for the hundreds of locations surveyed. The reference period of the simulation is 1980-2016 (37 years) and the bias-correction factors (which should be seen more as an estimation, to be in the right order of magnitude as compared to the plant's design discharge, than exact measures of mean discharge) may be from within this period or outside it, depending on the data source that was consulted for each. We have clarified in the paper that the bias-correction is meant rather as an estimation to be compared to design discharge than as a highly accurate measure of absolute flow rates (end of section 2.4).*

The application of bias correction or not for a given location is key. It is mentioned in the metadata of the AHA?
Author Response: *Yes. Wherever a multiannual mean discharge value was available as the basis for bias-correction, it is recorded in a separate column in 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)?
Author Response: *The REVUB model was not used in the present paper. It is indeed the sub-section to which the reviewer refers that contains the relevant information. Note that “small” in the referenced SM means “having less-than-a-year storage capacity”. In this regard, the “storable” and “non-storable” components indeed are not relevant for large reservoirs (defined as having “more-than-a-year storage capacity”), where all inflow is “storable”. Please see also our responses further below.*

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.
Author Response: *Since the REVUB model is not used in the present work, we refer to the other responses.*

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 ?...
Author Response: *The storable volume represents simply the volume of the hydropower reservoir,*

minus the "dead storage" component (that is, the volume that can be refilled and drawn down each year). For hydropower plants with less-than-a-year storage capacity, the total yearly volume inflow exceeds this storable volume. The "storable component" of the total yearly volume inflow thus equals the live reservoir volume, and the "non-storable component" is the part of total yearly volume inflow that exceeds this reservoir volume. We have made the language somewhat clearer in order to prevent confusion with the reader (section 2.5.2).

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 referred to as "storable" and "non-storable".
Author Response: *Indeed a filling sequence is required in many cases and is typical in practice. Many rivers across the African continent are highly seasonal in character. The purpose of constructing reservoirs is to allow to flatten out these seasonalities (since they are typically much stronger than the seasonalities in electricity demand) by storing some of the incoming water in the wet season and using it to produce electricity in the dry season. For highly seasonal rivers with pronounced dry and wet seasons, insisting that a reservoir have stable water levels throughout the year basically would amount to filling the reservoir once (after commissioning of the plant) and subsequently always turbining the water as it comes in (e.g. turbining almost nothing in the dry season, turbining a lot in the wet season). This would mean that plant operation would be basically imitating a run-of-river plant, which would serve no general purpose since then one could simply have left out the reservoir. This is more important than ensuring that the hydraulic head is always constant (because what good is hydraulic head if you can't turbine water anyway during the dry season?). We note here that the calculations performed are estimations of the seasonal character of hydropower availability, fully reproducible according to equations 1-3 and the explanations in 2.5.1-2.5.2. These equations do not "step forward in time" without knowing the future. Rather, they use already available time series of reservoir inflow to estimate monthly turbined flow and converting the latter to capacity factors. The concepts of "storable" and "non-storable" have been elaborated in the revision (section 2.5.2).*

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?
Author Response: *This calculation is performed independently of the initial storage level and of the demand. Basically, our methodology amounts to assuming that the amount of water that *can* be stored and flexibly turbined across the year is roughly equal to the live storage volume of the reservoir (we deem this a reasonable assumption, since this is what is physically possible, and the*

idea should be to flatten the seasonality as much as possible, in the sense that demand seasonality is usually much smaller than river flow seasonality, see e.g. Sterl et al. 2021: https://www.nature.com/articles/s41560-021-00799-5). If one had not wished to flatten seasonality to the extent implied by this methodology, one could have conceived a smaller reservoir and inundated a smaller area. We have included a comment in the manuscript that summarises the above (section 2.5.2). Insofar as demand is concerned, this issue is already addressed on p. 10, left column, S3.

On the other hand, in the 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).
Author Response: *We thank the reviewer for this comment. It is to be noted that our approach is statistical rather than actually deterministic – we do not model actual reservoir dynamics from hour to hour; rather, we use statistics (medians, percentile ranges) of modelled time series of river flow to infer "typical" profiles of hydropower generation. The implicit assumption here is indeed that reservoir operation follows a probabilistic approach based on historical experiences. We have added a comment in the discussion section (3) to clarify this point, and also cited some of the literature cited by the reviewer.*

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.
Author Response: *We believe that this question has already been answered in our response to the previous comments and the issue clarified in the new version of the manuscript.*

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? The outflow profiles of reservoir plants (whether large or small) were

determined as described in section 2.5.2, assuming (as mentioned in the comments above) that the full range of reservoir live storage is used to flatten out seasonality as much as possible by redistributing the storable component of river inflow equally across the months, and subsequently adding the non-storable component (whose seasonality is retained).
Author Response: *It is true that this methodology assumes that this outflow profile (as seen from the point of view of downstream run-of-river plants) does not change significantly before arriving at the downstream plant. We deem this a reasonable assumption as cascade configurations typically consist of several plants situated relatively close together geographically, but this was not clear in the previous version of the manuscript. We have included a comment in the manuscript to clarify this point (section 2.5.3).*
As mentioned 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.

Author Response: *A comment on the issue of irrigation water requirements is already mentioned in the discussion section: "there may be (...) certain hydropower plants where power generation needs to be co-optimised with irrigation".*

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.
Author Response: *We thank the reviewer for this observation. A table with all data / metadata provided for any location is now given in the manuscript (section "data availability").*

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.
Author Response: *Unfortunately, this is not yet the case, as it fell outside of temporal and budgetary possibilities available to the authors. We are interested in pursuing such a strategy in the future contingent upon budgetary possibilities.*

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?
Author Response: *Yes, it is expected that improved releases of the AHA will be published in the coming years as new information gets included in the database. The institution committed to do this update is the International Renewable Energy Agency (IRENA) in collaboration with the Vrije Universiteit Brussel. The contact, at this moment, is the lead author of this paper.*
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)? *Yes, the AHA includes a column "source location" which describes the source of the coordinates as included.*

Data Availability I am not a typical end-user 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 end-user could access easily to the full time series (and not only to the yearly mean profiles) of each hydrological unit.
Author Response: *This is the default output format provided by the SWAT+ simulations, but it is straightforward to regroup the data to obtain the simulated discharges for each unit individually. In fact, the Python code written to analyse the data does precisely this before moving on to the other elements of the analysis, and this code has been provided open-source along with the paper. It should thus be straightforward for a potential data user to obtain the time series cited by the reviewer (the variable "flow_extract" in the code). We refer to the SWAT+ output documentation (accessible through https://swatplus.gitbook.io/docs/download-docs) for further metadata on the columns included in the SWAT+ output .txt file.*

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?
Author Response: *In case the reviewer refers to the SWAT+ results when referring to the database, this information is given in the section "Data Availability". In case the reviewer refers to the AHA itself, this information is provided on the first worksheet "0 - Overview" of the spreadsheet file available on HydroShare.*

Detailed comments P4. First § : the multi-annual mean discharges. Can you precise for which period? Based on which data?
Author Response: *See our comment above (fourth response in the section "bias correction").*

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 time9 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 95h 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?
Author Response: *We thank the reviewer for this comment. While we see the rationale of the reviewer's suggestion, we prefer to leave the database as it is without adding additional percentile ranges, as this is likely to lead to confusion as to what our data suggests is representative of "very dry" and "very wet" years. However, it should be noted that, since the code and the AHA are open-access and open-source, any interested reader could re-run the code and extract the 5th and 95th percentiles for all plants. The only necessary adaptation would be changing* range_pct_dry_res *and* range_pct_wet_res *to the desired values in the code "load_flows.py" (e.g. to 5 and 95, instead of 10 and 90).*

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.
Author Response: *Correct. Basically, each of 12 monthly mean values (separately for dry, normal and wet years) was divided by the mean multiannual flow, such as to obtain a "normalised seasonality" including the information of whether a year is dry or wet compared to the long-term average. We have clarified this in the manuscript by adding a note that this refers to a simple scaling.*

p8 C1 §3 : "the maximum flow" >> "the maximum monthly flow" "monthly profiles" >> "annual cycles" *Corrected with thanks.* 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.
Author Response: *We thank the reviewer for this comment and have added a sentence in the discussion to clarify this.*

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).
Author Response: *We fully agree that this issue is important. During the revision of the paper, we developed the next version of the AHA (version 2.0), which now includes a first iteration of climate change scenario results. Dedicated SWAT+ simulations were performed under three RCP-SSP combinations and the AHA results were recalculated based on this. Thus, the AHA v2.0 contains now three extra sheets for these scenarios. We hope that this will add much additional value to the dataset and be to the reviewer's satisfaction. The following text was added to the revised paper: "A first iteration of the AHA under climate change scenarios has been included in its version 2.0 (see "Data availability"). These results were obtained by forcing SWAT+ with simulation results obtained from four bias-adjusted global climate models (GFDL-ESM2M, HadGEM2-ES, ISPL-CM5A-LR and MIROC5). The hydropower availability profiles were calculated separately for each model in the ensemble for the historical period 1975-2005 and for the future period 2075-2100, under three RCP-SSP combinations (Representative Concentration Pathways and Shared Socioeconomic Pathways, respectively): SSP1-RCP2.6, SSP4-RCP6.0, and SSP5-RCP8.5, whereby the land cover scenario is taken from the Land Use Harmonization version 2 data (Hurtt et al., 2020) and the global climate models are forced with greenhouse gas concentration trajectories associated with each RCP (Frieler et al., 2017). We then calculated the relative difference between the results (in terms of monthly availability of hydropower plants) from the future and historical period, and imposed this relative difference on the results obtained for the historical period forced with EWEMBI (see above)."*

SWAT+ : The description of SWAT+ has to be given in more details. The authors should give at least the references for the Penman-Montheith, 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.
Author Response: *We thank the reviewer for this comment. We have included references for the*

Penman-Monteith, SCS-CN and variable storage routing methods. To the authors' knowledge, continuous hydrological simulations using the SCS-CN method are well-documented in literature (e.g. Cho et al. 2018, https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.11463; Geetha et al. 2008, https://link.springer.com/article/10.1007/s11269-006-9149-5).

Competing Interests: No competing interests were disclosed.

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.

Author Response 13 Dec 2021

Sebastian Sterl

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. *We thank the reviewer for their positive words about our manuscript and are grateful for the constructive criticism which we have done our best to take into account.*
However, I would like to mention a few points that can be addressed or discussed to improve the work:

- 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.

Author Response: *We agree with the reviewer that this is important information to include in the database. We have released a new version of the database which includes this information as metadata (columns « SharedPlant » and « SharedPlantSplit » on worksheet no. 1 of the database).*

- 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.

Author Response: *In general, there is ample data available to validate annual average capacity factors (see the « Availability » column in the AHA results sheets), but seasonal patterns are hard to come by from publicly available information. Therefore, while such an exercise could be performed for some individual plants for which such information is publicly available (for a limited set of years), the authors of this paper currently lack temporal and budgetary resources to obtain such data from all African TSOs to provide a comprehensive validation. However, the authors are interested in*

performing such an exercise, and hope that this paper may stimulate data sharing to facilitate this.

- 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.

Author Response: *We thank the reviewer for this comment. We take the opportunity to clarify that the link between the items in the .txt files and the plants in the Excel file is clarified by columns "River Channel" and "Reservoir ID". Each hydropower plant location is linked to a unique River Channel ID and (if existing reservoir present) Reservoir ID. The AHA code, deposited on GitHub, reads the River/Reservoir ID from the Excel file and extracts the corresponding river flow dataset from SWAT+ in the .txt file. It thus links each hydropower plant to its corresponding river flow. Concerning the data format, we note that the .txt files are provided as input to run the simulations on and given in a format that allows to seamlessly integrate SWAT+ output with AHA processing (the SWAT+ model provides output in .txt format which is subsequently read in by the AHA code). The data product of this paper is the Excel file that represents the African Hydropower Atlas. We note that the .txt files contain too many rows to be presented in tabular (.csv) format. We refer to the SWAT+ output documentation (accessible through https://swatplus.gitbook.io/docs/download-docs) for further metadata on the columns included in this .txt file.*

Competing Interests: No competing interests were disclosed.