Environmental Research Letters

ACCEPTED MANUSCRIPT • OPEN ACCESS

Monitoring hydropower reliability in Malawi with satellite data and machine learning

To cite this article before publication: Giacomo Falchetta *et al* 2019 *Environ. Res. Lett.* in press https://doi.org/10.1088/1748-9326/ab6562

Monitoring hydropower reliability in Malawi with satellite data and machine learning

Giacomo Falchetta[a,b], Chisomo Kasamba[c,*], Simon C. Parkinson[d,e]

[a] Future Energy Program, FEEM - Fondazione Eni Enrico Mattei
[b] Dept. of Internat. Economics, Institutions and Development, Cattolica University
[c] Dept. of Energy, Ministry of Natural Resources, Energy and Mines of Malawi
[d] Energy Program, International Institute for Applied Systems Analysis (IIASA)
[e] Institute for Integrated Energy Systems, University of Victoria

Abstract

Hydro-climatic extremes can affect the reliability of electricity supply, in particular in countries that depend greatly on hydropower or cooling water and have a limited adaptive capacity. Assessments of the vulnerability of the power sector and of the impact of extreme events are thus crucial for decision-makers, and yet often they are severely constrained by data scarcity. Here, we introduce and validate an energy-climate-water framework linking remotely-sensed data from multiple satellite missions and instruments (TOPEX/POSEIDON. OSTM/Jason, VIIRS, MODIS, TMPA, AMSR-E) and field observations. The platform exploits random forests regression algorithms to mitigate data scarcity and predict river discharge variability when ungauged. The validated predictions are used to assess the impact of hydroclimatic extremes on hydropower reliability and on the final use of electricity in urban areas proxied by night-time light radiance variation. We apply the framework to the case of Malawi for the periods 2000-2018 and 2012-2018 for hydrology and power, respectively. Our results highlight the significant impact of hydro-climatic variability and dry extremes on both the supply of electricity and its final use. We thus show that a modelling framework based on open-access data from satellites, machine learning algorithms, and regression analysis can mitigate data scarcity and improve the understanding of vulnerabilities. The proposed approach can support long-term infrastructure development monitoring and identify vulnerable populations, in particular under a changing climate.

Keywords: Hydroelectricity, vulnerability, extreme hydroclimatic events, energy-climate-water nexus, random forests, remote sensing

*Corresponding author
Email address: chisomokasamba@yahoo.it (Chisomo Kasamba)

List of abbreviations

ATCF: As-the-crow-flies: CHIRPS: Climate Hazards Group InfraRed Precipitation with Station data; DD: Discharge Deviation; EGENCO: Electricity Generation Company Malawi; GADM: Database of Global Administrative Areas G-REALM: Global Reservoirs/Lakes database; HCF: Hydropower capacity factor; HPP: Hydropower plant; KGE: Kling-Gupta Efficiency; ML: Machine learning; MLI: Maximum likelihood; MODIS: Moderate Resolution Imaging Spectroradiometer; MOAIWD: Ministry of Agriculture, Irrigation and Water Development; MV: Medium-voltage; NOAA-CDR: National Oceanic and Atmospheric Administration - Climate Data Record NTL: Nighttime light; NSE: Nash-Sutcliffe Efficiency; P.P.: Percentage point PV: Photovoltaics RF: Random forests; SPEI: Standardised Precipitation-Evapotranspiration Index; SSA: Sub-Saharan Africa; VIIRS-DNB: Visible Infrared Imaging Radiometer Suite Day/Night Band.

Introduction

Developing countries experience recurrent issues in guaranteeing a reliable and secure provision of electricity to satisfy their domestic demand, with significant repercussions on economic growth and development prospects [1, 2, 3, 4, 5] and on the environment [6, 7]. A scarce diversification of the generation mix, the lack of sufficient and affordable back-up options, a limited adaptive capacity, and few international transmission lines represent some of the key underlying issues.

In sub-Saharan Africa, dependency on hydropower represents one of the most critical aspects of sustainable development [8, 9]. A large number of countries lack affordable means for coping with temporary disruptions caused, for instance, by hydro-climatic extremes such as a delayed rainy season or anomalous drought and flood periods. Diesel back-up capacity provided by independent power producers is often prohibitively costly for fully replacing the temporary loss in hydro generation [10]. As a result, load sheddings, brownouts, and blackouts are recurrent. Over the recent years, drought-related disruptions have been reported, for instance, in Kenya, Malawi, Tanzania, Ghana, Zimbabwe and Zambia, with frequent outages, power rationing, adverse business experience and competitiveness loss during precipitation anomalies [11].

Hydrological measurements and electricity supply and use data are affected by scarcity, quality, or inaccessibility issues [12]. This represents a great barrier to performing effective integrated assessment studies, developing modelling frameworks, and recommending policies for resilience building. While a multitude of studies have been carried out at the basin or global scale in terms of assessing the projected long-term impacts of climate change on hydropower generation potential [13, 14, 15, 16, 17, 18] and on the discharge of rivers [19], only few have assessed the impact of hydro-climatic extreme events on power supply reliability [20, 21, 22] and the related impacts on electricity consumption. Moreover, researchers (see [23]) have recently highlighted the necessity of reconciling top-down and bottom-up approaches to climate and energy-related assessment, indicating that novel methodologies - including the use of earth observation data [24, 25] - are required.

Here, we propose a novel framework based on open-access satellite-derived observations and their coupling with and validation against limited field data. The approach is applied to the case of Malawi, a country almost entirely dependent on hydropower [26] and currently lacking international transmission interconnections [27]. Input data (see Table SI1) include: remotely-sensed measurements on Lake Malawi water level (from the G-REALM database) [28], VIIRS-DNB product nighttime lights [29] (as a proxy of the local monthly urban electricity use in the country, see [30, 31, 32, 33], and thus also of outages [34, 35]), and climate conditions (including the SPEI drought index [36], precipitations [37], temperature [38], and soil moisture [39]). These datasets are modelled and validated against daily gauge data for discharge in the Shire River (between 2000 and 2018) and power generation at HPPs (between 2012 and 2018).

Our contribution shows that a modelling framework exploiting open climate and remotely-sensed data can reconstruct discharge measurements in situations of data scarcity and thus evaluate the impact of extreme hydro-climatic events on hydropower reliability. In turn, it provides a proof-of-concept for the use of nighttime satellite measurements of electric light radiance as a proxy to observe urban power consumption responses to hydrological shocks, underpinning the challenges stemming from a dependency on hydropower. This is a particularly relevant finding given the forecasted intensification of extreme hydrological events in East Africa [40].

Study area: contextualising the case of the Shire River Basin in Malawi

Fig. 1 depicts the nighttime light radiance and the MV distribution grid (panel A) - which provide a snapshot of the current local electricity access and use situation -, the georeferenced population density of Malawi for year 2018 [41] (panel B), and the hydrological basin modelled in this study (panel C), including the location of the hydropower stations currently operating in the country and the hydrological gauge stations. The population is distributed, with high density settlements concentrated in the center-south of the country, around the cities of Lilongwe and Blantyre. Lake Malawi, the third largest in Africa by extent, delimits a large part of the eastern border of the country. The Shire River is an outlet of Lake Malawi, and along it the bulk of the installed hydropower capacity is concentrated. Table SI2 lists the technical specifications of each of those generation plants. Figure 1D provides a profile view of the topography of the Shire River, including the location of dams, gauge stations, and tributary rivers.

Previous hydrological studies have assessed the trends and relationships between the water level in Lake Malawi [44, 45], the discharge in the Shire River [46],

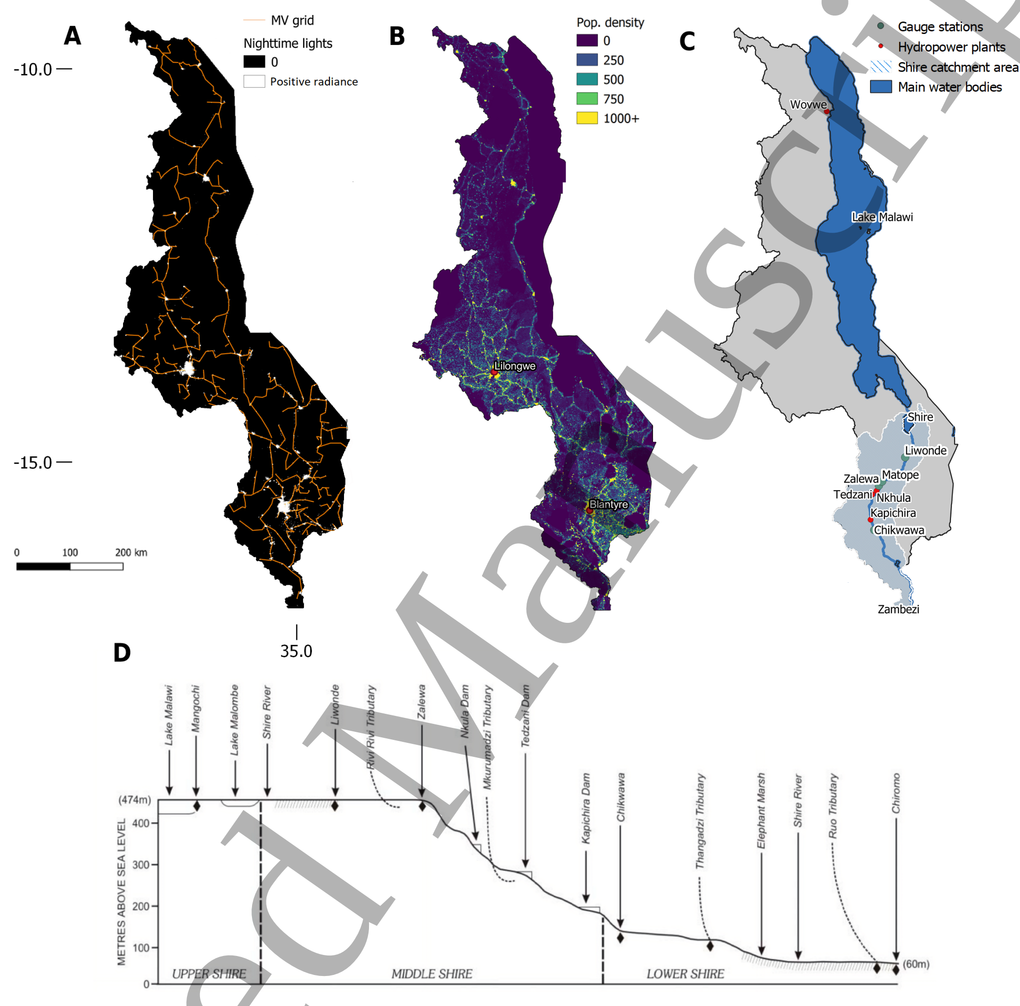

Figure 1: Maps of Malawi, representing: (A) a snapshot of the nighttime lights radiance and the MV distribution grid (using data from year 2018. Sources: [29, 27]); (B) the local population density in 2017 (data from [42, 43]); (C) the Shire River basin and Lake Malawi, and the location of hydropower plants operating in Malawi and of the discharge measurement stations considered in this study; (D) the topography of the Shire River, including the location of hydropower schemes, gauge stations, and tributary rivers.

and the observed and potential impacts of climate change on the local hydrology [47, 48] and hydropower generation [49], as well as the perceived risks and potential adaptation options under climatic and socio-economic uncertainty in the Shire River Basin [50]. The literature has also highlighted that the lake level is highly sensitive to climate variability [46], with cyclic fluctuations in levels being largely subject to annual rainfall patterns and seasonal precipitation and temperature variables anticipating lake level changes by approximately two months. According to modelling studies based on downscaled climate projections [48], a warmer climate will likely contribute to a further decrease in the water balance. Concerning discharge in the Shire River, local precipitations and temperature have been found to anticipate river flow surges by 2 days [46]. In general, long-lived hydrological flood and drought events in the Shire River basin are influenced by the large-scale atmospheric circulation and rainfall in the surrounding highlands. Hence, impact assessment tools should consider satellite and radar coverage of the entire basin.

Materials and methods

Fig. 2 presents a schematic of the integrated framework here proposed and tested empirically. The acquisition and processing procedure for each specific dataset and variable is fully described in the Detailed Materials and Methods section in the SI, which also includes the explicit regression equations, random forest parameters and set-up, and GIS algorithms operated at each stage of the platform. Monthly-invariant factors are always included to control for the role of seasonality in hydrological, climate, and power supply and demand. The Data Availability Section presents a repository integrating R, Python and Google Earth Engine API code that enable the replication of the modelling framework and results.

First, a random forests algorithm (run using the *caret* R package [51] with 250 trees, a 10-dimensional parameter tuning length, and a 10-fold cross-validation) assesses the predictive power of open-data measured over the entire Shire River Basin for precipitations, temperature, soil moisture, and the SPEI index (Standardised Precipitation Evapotranspiration Index, see Table SI3 for the definition and classification of its values) at multiple scales [36] over the water level measured by satellites at Lake Malawi [28]. This step evaluates the consistency among remotely-sensed and field gauge observations.

Lake Malawi's level measurements are then combined with climate control variables to account for precipitations and evapotranspiration over the entire river basin and used to evaluate the predictive accuracy over the discharge (measured in $m^3 \cdot s^{-1}$) in the Shire River at three gauging stations: Liwonde, 36.5 km ATCF south of Lake Malawi; Matope, 50 km ATCF south of Liwonde, which is itself 16 and 23 km ATCF north of Nkula A&B and Tedzani run-of-river HPP, respectively; and Chikwawa, 10 km ATCF south of Kapichira Dam. The approach is essential to fill the sporadic discharge time series and thus improve the subsequent assessment of the impact of hydrological extremes on hydropower. It also tests the potential of remotely-sensed data to largely replace ground measurements. Both hydrological modelling steps are carried out at a daily temporal resolution.

A measure of Discharge Deviation, defined as the difference between the daily observed discharge and the long-term mean discharge for the month m in the

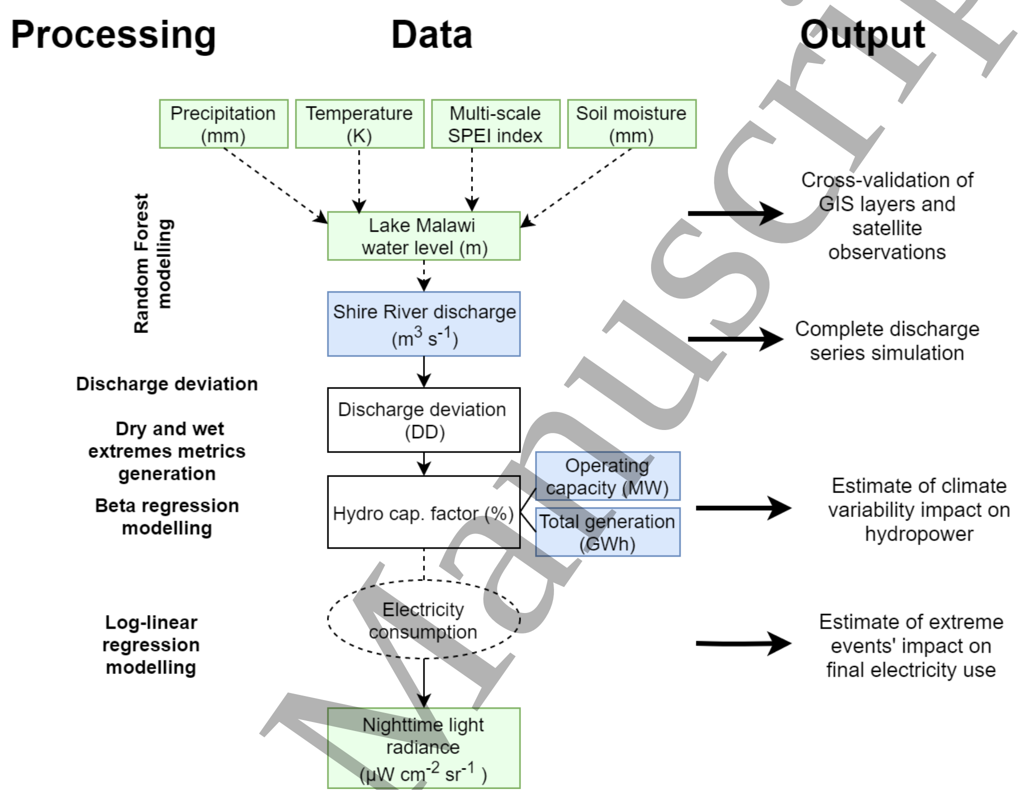

Figure 2: Schematic of the modelling and validation framework for (i) assessing the impact of hydro-climatic variables on the water level at Lake Malawi; (ii) estimating the daily discharge at different gauge stations in the Shire River; (iii) assessing the impact of deviations in the discharge and of extreme discharge deviations, respectively, on (a) the hydropower capacity factor and (b) the detected NTL radiance over urban areas. A green shading denotes remotely-sensed or geoprocessed openly-available and regularly updated datasets; a blue shading denotes field gauge variables, which are used to validate the model; a white shading refers to variables calculated through combination of the data input sources. Dashed arrows denote a simulation process is carried out to fill gaps in the time series, while dashed boxes represent unobserved, proxied variables.

river at each gauging station g in each day d, (as in Eq. 1) is introduced. Here, ΔD represents discharge deviation and $\bar{D}$ is the long-run mean discharge (based on data between 2000 and 2018) in month m (the corresponding month of belonging of day d).

$$\Delta D_d^g = D_d^g - \bar{D_m^g} \quad (1)$$

The impact of deviations in the discharge on the capacity factor of each individual hydropower plant i and on the total operating capacity in the country is assessed using a beta regression model via maximum likelihood (MLI) (see Materials and Methods SI). The capacity factor is defined as the effective output as a share of the total installed capacity operating at each day d, as in Eq. 2. Here, T is the constant adjusted to 24 and 720 for the daily and monthly capacity factors, respectively. The operating capacity is defined based on a broad assessment of online industry and technical reports providing the construction and rehabilitation works at individual schemes over the period of time covered in the analysis (Table SI2).

$$HCF_d^i = \frac{P_d^i}{C_d^i \cdot T} \tag{2}$$

P is total generation (in MWh), C is operating capacity (in MW), T is time (in h), i defines each hydropower plant and d identifies each day. Metrics for assessing dry and wet extremes (throughout the paper we adopt this definition because there are no standard thresholds to define drought and flood events) is also generated, as in Eq. 3a-3b, which classifies dry and wet extremes as the discharge deviation events below the 5^{th} and above the 95^{th} percentile of the distribution, respectively.

$$dryextreme_d = \begin{cases} 1, & \text{if } \Delta D_d < 5^{th} percentile \Delta D_d \\ 0, & \text{otherwise} \end{cases} \tag{3a}$$

$$wetextreme_d = \begin{cases} 1, & \text{if } \Delta D_d > 95^{th} percentile \Delta D_d \\ 0, & \text{otherwise} \end{cases} \tag{3b}$$

To link the supply and demand-side, the relationship between the incidence of extreme hydrological events and the satellite-detected nighttime light radiance [52] both throughout urban areas of Malawi and in each specific province is evaluated through a log-linear OLS (ordinary least squares, the standard statistical regression framework model). In this case, the relationship is assessed at a monthly scale, the native temporal resolution of the VIIRS DNB product. NTL radiance at month m in province p is defined as the sum of radiance in each pixel n within province p, conditional on the pixel having a population density greater than 250 inhabitants $\cdot$ km^{-2} (using gridded population data from [41]).

$$NTL_m^p = \sum_1^n NTL_n^p \tag{4}$$

Here, NTL is nighttime light radiance (in $\mu W \cdot cm^{-2} \cdot sr^{-1}$) in month m in each province p. Thus, the total NTL is given by the sum of NTL throughout all provinces p at month m:

$$NTL_m^{Country} = \sum_i^p NTL_m^i \tag{5}$$

Results

The results of the analysis (detailed in the next paragraphs) highlight that - with a proper modelling framework - open-access data can be leveraged to assess climate-induced power generation fluctuations and disruptions, as well as implications for electricity use in urban areas. We provide evidence of the strong effectiveness of remotely-sensed, open-data in complementing limited field gauge observations in energy-climate-water nexus modelling. Our empirical estimates of the impact of hydroclimatic extremes for hydropower reliability suggest average declines of 9.4 percentage points (in absolute terms) in the monthly HCF in Malawi during dry extreme events compared to the long-run average value recorded in the same month. Yet, we find no evidence of an adverse impact of wet extreme events on HCFs. Finally, we show that unmet urban demand and outages driven by declines in hydropower generation can be successfully detected via changes in the detected nighttime lights (with average decreases of 31 p.p. in the monthly urban NTL radiance during a dry extreme event). This also reveals substantial heterogeneity in the province-level responses, where both policy and electricity access and use levels play a role in determining exposure.

Hydrological response to climate: predicting the water level at Lake Malawi and concurrent discharge in the Shire River

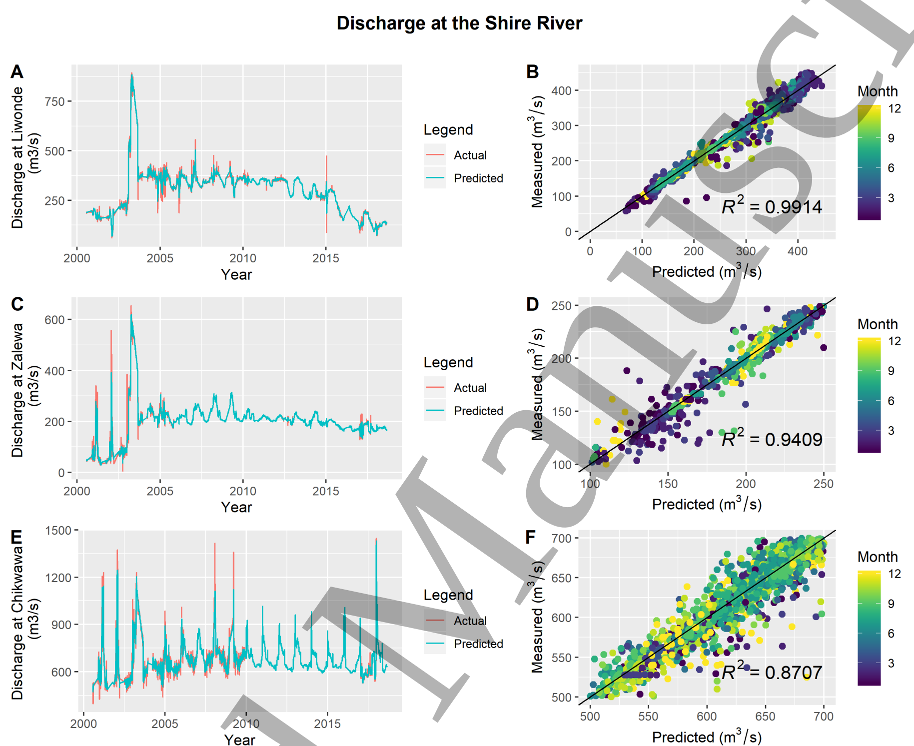

Figure 3: Comparison and statistical evaluation of modelled and gauged discharge at Liwonde (a,b), Matope (c,d), and Chikwawa (e,f). Left hand side panels: time-series representation. Right hand side panels: scatter-plot and R^2 results from the random forests regression analysis.

While Lake Malawi's level measurement series is complete and does not necessitate missing values imputation, we begin by operating a RF regression to evaluate the consistency of the remotely-sensed variables implemented in the next steps. Fig. SI1 depicts the predicted vs. the satellite-measured lake level at a daily temporal resolution. The predicted values are obtained from the random forests regression model described in the corresponding SI section, while the RF model statistics and diagnostics are reported in Table SI4 and Fig. SI3. The results show that the geospatial open-data have nearly full explanatory power over the daily water level at Lake Malawi as measured by TOPEX/POSEIDON and Jason satellites. In particular, 10-fold cross-validated training accuracy and test accuracy values are both above 0.99. The variable importance metric (discussed in [53] and depicted in Fig. SI4) shows that the long-term scale SPEI48 index is the most significant predictor, followed by SPEI24 and SPEI06, and by the average temperature over the previous three months and the monthly seasonality. Only a fraction of the total variance remains unexplained (as seen in Fig. SI1). We thus find evidence of a strong consistency across variables from different source, which encourages their use in the following hydrological modelling.

In a second step, we evaluate the precision with which satellite-derived lake level and the upstream (predicted) gauges can model the discharge (in $m^3 \cdot s^{-1}$ units) at Liwonde, Matope, and Chikwawa discharge gauge stations, the geographical position of which is reported in Fig. 1C. The SPEI index at several scales, precipitations, and temperature are added as covariates to account for precipitations falling directly on the riverbed and on tributaries. The random forests regression models (see Figs. SI5 to SI10 for parameters and variable importance metrics) yield cross-validated training accuracy and validation accuracy values of 98.7% and 99% at Liwonde, of 97.3% and 94.5% at Matope, and of 90.6% and 89% at Chikwawa. We calculated KGE and NSE metrics for hydrological model validation - which are particularly suitable to cope with extreme values (see Material and Methods SI) - and obtained values of 0.99 (for both KGE and NSE), 0.98 and 0.97 (for KGE and NSE, respectively), and 0.96-0.94 (for KGE and NSE, respectively) at Liwonde, Matope, and Chikwawa, respectively.

This result is of great importance to the purposes of our analysis, because it shows the capability of the model to accurately reproduce sporadic discharge measurements from the field with the exclusive use of satellite and other open geospatial data. In the original time series for the 2000-2018 period examined, 29%, 72% and 61% of daily observations are ungauged at Liwonde, Matope, and Chikwawa, respectively. The RF modelling allows accurately filling these large gaps and thus performing a more precise impact assessment. Fig. 3 shows the goodness-of-fit at each gauge station. The approach is thus found to be appropriate to fill the massive gaps in the gauge time-series. Notably, as seen from Fig. 3, the model fails to reproduce certain extreme spikes - which could either be instrument measurement errors or extreme events. Thus, our estimates of the impact of extreme events are most safely interpreted as lower-bound values.

Linking discharge deviations and extreme events to hydropower output

Once the complete discharge series is simulated, we calculate metrics of discharge deviations and extreme events (as defined in Eqs. 1-3b) using discharge values at the nearest upstream gauge station to the bulk of the installed run-of-river hydropower capacity, i.e. Matope. We tested a statistical relationship with the hydropower capacity factor, defined as the electricity generated as a share of the maximum technical generation potential in each unit of time. The complete regression specifications are illustrated in Eqs. SI7-8 in the Detailed Materials and Methods SI. As seen from Fig. SI2, in theoretical terms the most widespread kind of turbines run optimally when the relative discharge lies between the 80-90% interval, and yet that efficiency is little responsive to changes in relative discharge up to a level of about 30%, after which efficiency sinks. Fig. 4A plots the relationship between discharge in the Shire River at the Matope gauge station (located 16 km upstream of Nkula Dam) and the daily total HCF, while Fig. 4B shows the estimated range of effect of days classified as dry and wet extremes on the daily total HCF.

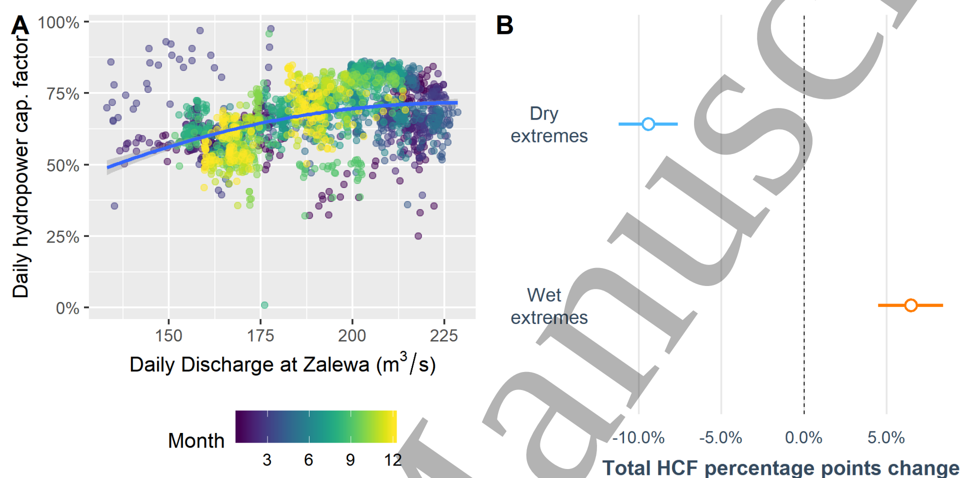

Figure 4: (A) Empirical relationship between the discharge in the Shire River measured at Matope gauge station and daily total HCF; (B) Error bars of the impact of days classified as dry and wet extremes on the total HCF

Our regression results (see Table SI5) are consistent with the theoretical relationship illustrated by efficiency curves. When considering the model developed independently of month and year, on average, a 1 m^3/s decrease in the standard deviation of long-run discharge results in a 0.2 p.p. decline in the total daily hydropower capacity factor ($P < 0.01$). The result reflects the low sensitivity of hydropower turbines to discharge deviation, as long as discharge remains in the normal fluctuation range. Yet, when restricting the measurement of the impact to days classified as dry extremes, we highlight that these determine an average decline by 9.4 p.p. in the total hydropower capacity factor compared to the average value for the same month ($P < 0.01$), as seen from Fig. 4B. This result is reported in Table SI9. On the other hand, we find no evidence of an adverse effect of wet extreme events, which are in fact associated with slightly higher than average HCFs. (Table SI10). The impact of dry extremes is also tested for the capacity factors of each individual hydropower plant (to assess the heterogeneity in the vulnerability of each facility). The corresponding regressions results are reported in Tables SI6 to SI8.

Measuring final power use responses with nighttime lights

Due to the lack of official sub-yearly and sub-national urban electricity consumption data, we exploit the observed NTL radiance in urban areas with a density > 250 inhabitants $\cdot$ km^{-2} at both the country and at the province level as a proxy variable for estimating the effect of extreme hydrological events. A growing stream of literature has shown that nighttime light data are able of capturing spatio-temporal electricity use variation (and in particular outages and disaster-related disruptions) [30, 54, 32, 34, 35]. Malawi is relying largely on hydropower and it has little backup options. At the same time, the country is constrained by the inability to import power from abroad. Thus periods of climate-induced reduced domestic supply determine a reduced consumption potential. From a statistical point of view, both power supply and consumption are endogenously determined by an array of unobserved external factors (such as costs, policies, industrial activity, etc.) and they simultaneously affect each other. Yet, river discharge is assumed to be exogenous to power consumption when no storage reservoir is available, as it is the case in Malawi. Specifications consider month fixed-effects to account for seasonality in the regressors, i.e. recurrent seasonal patterns in nighttime light radiance (i.e. electric power use) and hydropower capacity factor.

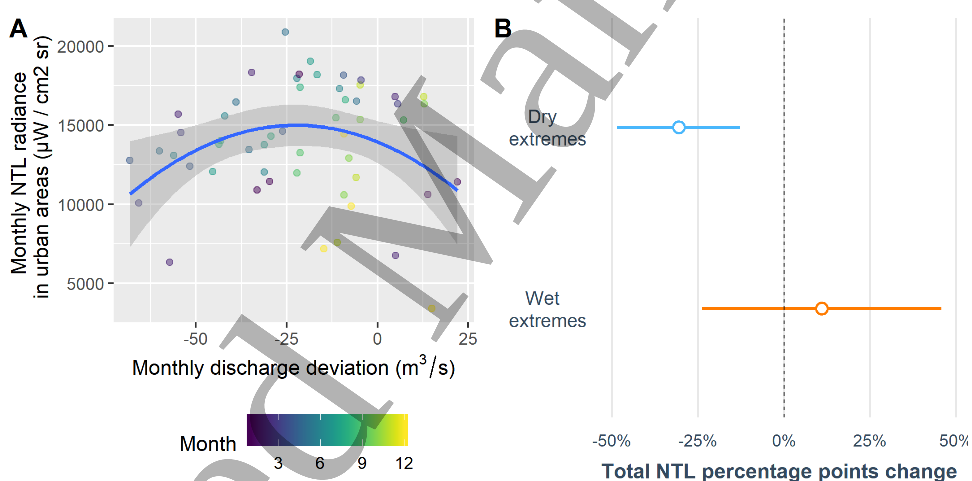

Figure 5: Empirical relationship between the monthly discharge deviation from long-run mean value in the same month and the sum of NTL radiance in urban areas of Malawi in the same month.

Fig. 5A depicts the relationship between the monthly deviation from long-run average discharge in the same month - DD_m - and the monthly sum of NTL radiance in urban areas of Malawi. When limiting the assessment to the effect of extreme discharge events on the total NTL radiance of urban areas (Fig. 5B), country-level results (see Table SI11) suggest a very significant negative effect. This is quantified at an average value of 31 p.p. in response to a negative shock during the average dry extreme month (with $P < 0.01$), while no significant effect on the detected NTL level is found for wet extremes (Table SI12).

A further related question concerns the sub-national heterogeneity in the fluctuation of NTL radiance during months affected by extreme events. Fig. 6 plots the effect of an extreme hydrological event on the detected NTL radiance at the different provinces. The results reveal a heterogeneous picture, with some provinces showing declines in the total NTL radiance of more than 150% compared to non dry extreme months, and some other province where no significant effect is found. For those provinces where a statistically significant (P¡0.05) effect of extreme events on NTL radiance is found, we observe a moderate positive correlation ($\rho = 0.2$) between the magnitude of the average NTL decline and the local electricity access level reported by the 2015-16 Demographic Health Survey, [55]), i.e. the fraction of households exposed to supply disruptions. This provides evidence of the decline in NTL being associated with the share of households with electricity access in each province,

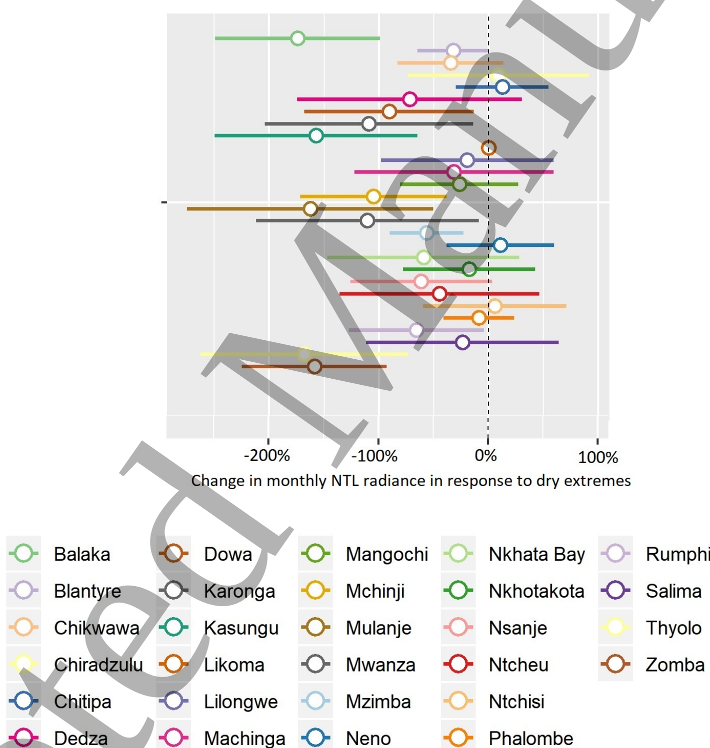

Figure 6: Heterogeneity in effect of dry extremes on NTL across provinces of Malawi. The figure illustrates the average effect on the change in the NTL radiance (in p.p.) compared to non dry extreme months.

Yet, the effect is found to be statistically insignificant in the two largest cities of Malawi, i.e. Lilongwe and Blantyre, reflecting the fact that the main centers are less targeted by load-sheddings during supply shortages (see the SI for a description of ESCOM's load shedding policy in different provinces of Malawi) and that some diesel-fired backup capacity is available locally. The strongest (proxied) consumption declines as a result of dry extremes are localised in Mulanje, Salima, and Dedza provinces, located in Central and Southern Malawi, in the proximity of the two largest cities, Lilongwe and Blantyre.

Main limitations and uncertainty

The results presented in this research letter provide a relevant proof-of-concept of how satellite data can be modelled to improve the prediction of different interrelated trends in the water-climate-energy nexus where data are infrequently gauged or not publicly accessible. Yet, an explicit statement of the limitations and the main sources of uncertainty encapsulated at the different stages of the framework are necessary.

Firstly, the remotely sensed input data - in particular high spatio-temporal resolution climatic observations - are the result of calibration and interpolation techniques and can thus bear an error component. This is of particular relevance in a data-sparse region like that examined in this study, where field validation is likely to be very limited. Secondly, the authors are aware that there is a certain and unavoidable degree of arbitrariness in the classification of extreme events, and this has a direct impact on their measured impact (as discussed in the relevant literature [56, 57]). Finally, the framework introduced and modelled does not encapsulate a hydropower operation model capable to assess human decisions relative to power plants operation and optimise them. The objective of this study is in fact offering a schematic approach to measure and understand the observed water-energy nexus relationships to render them more explicit to scientists and decision-makers. Our framework can serve as a concept for the development or improvement of impact assessment systems through the use of earth observation data. At the same time, the estimated sensitivity parameters (such as that of HCF to the SPEI drought index variance) can support the development and calibration of hydrological and energy-climate-water integrated assessment models.

We encourage further work to elaborate on similar frameworks and adapting them to account for dam operation dynamics in contexts of grater complexity. This could enable an integration into predictive models with projected climate parameters under different global warming scenarios.

Discussion

In developing regions, including SSA, data scarcity is a strong barrier to strategic environmental and socio-economic assessments. Building on a modelling framework exploiting open climate and remotely-sensed data, we have shown that hydropower generation in Malawi is mildly sensible to discharge deviations but it is particularly affected by extremes, which reduce HCFs by an average of 9.4 p.p. compared to the usual level in the same month. We found that this translates into average 31 p.p. decreases in the monthly NTL radiance during extreme hydrological events and it plummets by more than 150% in specific exposed provinces of Malawi. This is a particularly relevant finding given the forecasted intensification of extreme hydrological events in East Africa [40].

In this context, power mix diversification and transboundary electricity transmission infrastructure development represent crucial policy actions to increase supply reliability. Water supply is likely to grow due to both climatic stressors and increasing consumptive demand from the agricultural sector and other human uses. Currently, 548 MW of new hydropower capacity distributed across three dams (more than half of which on the Shire River, as reservoir dams) are expected to be delivered by 2025, while the only non-hydro expansion projects currently announced is the 300 MW coal-fired Khammwamba Power Station, expected for 2022, with coal imported from Mozambique via rail.

This is despite the fact that variable renewable sources of energy (VRE) are widely available in Malawi. Throughout the country, the solar photovoltaics (PV) generation potential is above 1600 kWh/kWp , with peaks of 1800 kWh/kWp per year [58] which imply significant potential for utility-scale PV parks. A recent study highlighted that currently 60 MW of techno-economic potential are available in the country [59], mostly localised in sites in the southern part of the country [60]. The co-integration of VRE with hydro, in particular when reservoir-based schemes will become operational, offers potential benefit in terms of increasing resilience of the power sector and balancing both renewables intermittency and hydropower output fluctuations, as evidenced in the literature [61, 62, 63, 64, 65]. This would also alleviate the need for importing coal from Mozambique, or, possibly, natural gas from Tanzania, and thus guarantee supply self-sufficiency and a cleaner power sector.

Conclusion

In this study, we have shown that a modelling framework exploiting open climate and remotely-sensed data can (i) reconstruct discharge measurements in situations of data scarcity and thus (ii) evaluate the impact of extreme hydroclimatic events on hydropower reliability. In turn, (iii) nighttime lights data can be used to observe power consumption responses to hydrological shocks in urban areas at a monthly scale, underpinning the challenges stemming from a dependency on hydropower.

Supplementary information

The Supplementary Information accompanying the paper includes an extensive description of the Materials and Methods, result tables, and figures.

Data availability

The R code for processing the workflow (including Python scripts for interacting with the Google Earth Engine API and processing remotely-sensed data), and the required input data for replication are stored at the following repository: `https://github.com/giacfalk/hydropower_remotesensing`.

Acknowledgements

Financial support from the MIUR (Italian Ministry of University and Research) through the Catholic University of Milan and to Fondazione Eni Enrico Mattei is gratefully acknowledged. The authors are grateful to the Ministry of Agriculture, Irrigation and Water Development of Malawi for sharing the Shire River discharge data and to the Electricity Generation Company of Malawi (EGENCO) for providing the hydropower generation time-series at individual facilities. Furthermore, the first author would like to thank the participants to the EGU 2019 General Assembly for their feedback during and after the oral presentation of a preliminary version of this paper. The authors would like to thank Sebastian Sterl and Edward Byers for their comments on the manuscript.

Author contributions

G.F. processed the satellite data, developed the modelling framework, and wrote the paper; C.K. processed and provided the field gauge data, contributed to the analysis with local knowledge, and wrote the paper; S.C.P. provided useful input to the hydrological modelling and contributed to writing the paper.

Declaration of interest

Declarations of interest: none.

References

[1] Cole, M. A., Elliott, R. J., Occhiali, G. & Strobl, E. Power outages and firm performance in sub-Saharan Africa. *Journal of Development Economics* **134**, 150–159 (2018).

[2] Nduhuura, P., Zerga, A. & Garschagen, M. Power outages in Africa – An assessment based on regional power pools. *PAUWES Research-2-Practice Forum 2018* (2018).

[3] Andersen, T. B. & Dalgaard, C.-J. Power outages and economic growth in Africa. *Energy Economics* **38**, 19–23 (2013).

[4] Oseni, M. O. Power outages and the costs of unsupplied electricity: evidence from backup generation among firms in Africa. *Working Paper, Judge Business School, University of Cambridge* (2012).

[5] Kaseke, N. & Hosking, S. G. Sub-Saharan Africa Electricity Supply Inadequacy: Implications. *Eastern Africa Social Science Research Review* **29**, 113–132 (2013).

[6] Farquharson, D., Jaramillo, P. & Samaras, C. Sustainability implications of electricity outages in sub-Saharan Africa. *Nature Sustainability* **1**, 589 (2018).

[7] Tanner, A. M. & Johnston, A. L. The impact of rural electric access on deforestation rates. *World Development* **94**, 174–185 (2017).

[8] Falchetta, G., Gernaat, D., Hunt, J. & Sterl, S. Hydropower dependency and climate change in sub-Saharan Africa: A nexus framework and evidence-based review. *Journal of Cleaner Production* **231**, 1399–1417 (2019).

[9] Cole, M. A., Elliott, R. J. & Strobl, E. Climate change, hydro-dependency, and the African Dam boom. *World Development* **60**, 84–98 (2014).

[10] Karekezi, S., Kimani, J., Onguru, O. & Kithyoma, W. Energy Security and Adaptation to Climate Change in East Africa and the Horn of Africa: Large Scale Hydropower vs. Decentralized Renewables. *AFREPREN/FWD Occasional Papers Series* (2012).

[11] Gannon, K. E. *et al.* Business experience of floods and drought-related water and electricity supply disruption in three cities in sub-Saharan Africa during the 2015/2016 El Niño. *Global Sustainability* **1** (2018).

[12] Larsen, M. *et al.* Challenges of data availability: Analysing the water-energy nexus in electricity generation. *Energy Strategy Reviews* **26**, 100426 (2019).

[13] Sridharan, V. *et al.* Resilience of the Eastern African electricity sector to climate driven changes in hydropower generation. *Nature communications* **10**, 302 (2019).

[14] Van Vliet, M. T., Wiberg, D., Leduc, S. & Riahi, K. Power-generation system vulnerability and adaptation to changes in climate and water resources. *Nature Climate Change* **6**, 375 (2016).

[15] van Vliet, M. T. H. *et al.* Multi-model assessment of global hydropower and cooling water discharge potential under climate change. *Global Environmental Change* **40**, 156–170 (2016).

[16] Cervigni, R., Liden, R., Neumann, J. E. & Strzepek, K. M. *Enhancing the Climate Resilience of Africa's Infrastructure* (Washington, DC: World Bank, 2015). DOI: 10.1596/978-1-4648-0466-3.

[17] Turner, S. W. D., Ng, J. Y. & Galelli, S. Examining global electricity supply vulnerability to climate change using a high-fidelity hydropower dam model. *The Science of the Total Environment* **590-591**, 663–675 (2017).

[18] Hamududu, B. & Killingtveit, A. Assessing climate change impacts on global hydropower. *Energies* **5**, 305–322 (2012).

[19] Döll, P. & Schmied, H. M. How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis. *Environmental Research Letters* **7**, 014037 (2012).

[20] van Vliet, M. T., Sheffield, J., Wiberg, D. & Wood, E. F. Impacts of recent drought and warm years on water resources and electricity supply worldwide. *Environmental Research Letters* **11**, 124021 (2016).

[21] Kull, D. Connections between recent water level drops in Lake Victoria, dam operations and drought (2006).

[22] Bekoe, E. O. & Logah, F. The impact of droughts and climate change on electricity generation in Ghana. *Meteorology and Energy Security: Simulations, Projections, and Management* 163–178 (2016).

[23] Conway, D. *et al.* The need for bottom-up assessments of climate risks and adaptation in climate-sensitive regions. *Nature Climate Change* (2019).

[24] Leibrand, A., Thomas, A., Sadoff, N. & Maslak, T. Using Earth Observations to Help Developing Countries Improve Access to Reliable, Sustainable and Modern Energy. *Frontiers in Environmental Science* **7**, 123 (2019).

[25] Gleason, C. J., Wada, Y. & Wang, J. A hybrid of optical remote sensing and hydrological modeling improves water balance estimation. *Journal of Advances in Modeling Earth Systems* **10**, 2–17 (2018).

[26] EIA. International Energy Statistics (2019).

[27] Gershenson, D., Roher, B. & Lerner, A. Predictive model for accurate electrical grid mapping (2019).

[28] USDA. Satellite Radar Altimetry: Global Reservoir and Lake Elevation Database (2018). [Online; accessed 2018-02-15].

[29] Elvidge, C. D., Baugh, K., Zhizhin, M., Hsu, F. C. & Ghosh, T. VIIRS night-time lights. *International Journal of Remote Sensing* **38**, 5860–5879 (2017).

[30] Stokes, E. C. *et al.* Urban Applications of Nasa's Black Marble Product Suite. In *2019 Joint Urban Remote Sensing Event (JURSE)*, 1–4 (IEEE, 2019).

[31] Xiao, H. *et al.* Spatio-temporal simulation of energy consumption in China's provinces based on satellite night-time light data. *Applied energy* **231**, 1070–1078 (2018).

[32] Hu, T. & Huang, X. A novel locally adaptive method for modeling the spatiotemporal dynamics of global electric power consumption based on DMSP-OLS nighttime stable light data. *Applied Energy* **240**, 778–792 (2019).

[33] Shi, K. *et al.* Detecting spatiotemporal dynamics of global electric power consumption using DMSP-OLS nighttime stable light data. *Applied energy* **184**, 450–463 (2016).

[34] Wang, Z. *et al.* Monitoring disaster-related power outages using NASA black marble nighttime light product. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci* 1853–1856 (2018).

[35] Román, M. O. *et al.* Satellite-based assessment of electricity restoration efforts in Puerto Rico after Hurricane Maria. *PloS one* **14**, e0218883 (2019).

[36] Beguería, S. & Vicente-Serrano, S. M. *SPEI: Calculation of the Standardised Precipitation-Evapotranspiration Index* (2017). [Online; accessed 2018-05-15].

[37] Funk, C. *et al.* The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data* **2**, 150066 (2015).

[38] Wan, Z., Hook, S. & Hulley, G. MOD11A2 MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V006. *NASA EOSDIS Land Processes DAAC* **10** (2015).

[39] Bolten, J. & Crow, W. Improved prediction of quasi-global vegetation conditions using remotely-sensed surface soil moisture. *Geophysical Research Letters* **39** (2012).

[40] Conway, D., Dalin, C., Landman, W. A. & Osborn, T. J. Hydropower plans in eastern and southern Africa increase risk of concurrent climate-related electricity supply disruption. *Nature Energy* **2**, 946 (2017).

[41] Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C. & Worley, B. A. LandScan: a global population database for estimating populations at risk. *Photogrammetric engineering and remote sensing* **66**, 849–857 (2000).

[42] Pekel, J.-F., Cottam, A., Gorelick, N. & Belward, A. S. High-resolution mapping of global surface water and its long-term changes. *Nature* **540**, 418 (2016).

[43] World Resources Institute. WRI's Major Watersheds of The World (2006).

[44] Makwinja, R., Phiri, T., Kosamu, I. B. & Kaonga, C. C. Application of stochastic models in predicting Lake Malawi water levels. *International Journal of Water Resources and Environmental Engineering* **9**, 191–200 (2017).

[45] Kaunda, P. *Investigating the impacts of climate change on the levels of lake Malawi*. Ph.D. thesis, University of Nairobi (2016).

[46] Jury, M. R. Malawi's shire river fluctuations and climate. *Journal of Hydrometeorology* **15**, 2039–2049 (2014).

[47] Warnatzsch, E. A. & Reay, D. S. Temperature and precipitation change in Malawi: Evaluation of CORDEX-Africa climate simulations for climate change impact assessments and adaptation planning. *Science of the Total Environment* **654**, 378–392 (2019).

[48] G Kumambala, P. & Ervine, A. Water balance model of Lake Malawi and its sensitivity to climate change. *The Open Hydrology Journal* **4** (2010).

[49] Kaunda, C. S. & Mtalo, F. Impacts of environmental degradation and climate change on electricity generation in Malawi. *International Journal of Energy & Environment* **4** (2013).

[50] Bhave, A., Bulcock, L., Dessai, S., Jewitt, G. & Conway, D. Malawi's vulnerability to threshold behaviour of Lake Malawi: A lake-basin modelling study for informing adaptation decision making under uncertainty. In *EGU General Assembly Conference Abstracts*, vol. 20, 14755 (2018).

[51] Kuhn, M. *et al.* Building predictive models in R using the caret package. *Journal of statistical software* **28**, 1–26 (2008).

[52] Román, M. O. *et al.* NASA's Black Marble nighttime lights product suite. *Remote Sensing of Environment* **210**, 113–143 (2018).

[53] Archer, K. J. & Kimes, R. V. Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis* **52**, 2249–2260 (2008).

[54] Jasiński, T. Modeling electricity consumption using nighttime light images and artificial neural networks. *Energy* **179**, 831–842 (2019).

[55] USAID. MEASURE DHS STATcompiler. (2009).

[56] Lloyd-Hughes, B. The impracticality of a universal drought definition. *Theoretical and Applied Climatology* **117**, 607–611 (2014).

[57] Wilhite, D. A. & Glantz, M. H. Understanding: the drought phenomenon: the role of definitions. *Water international* **10**, 111–120 (1985).

[58] World Bank. *Solar Resource and PV Potential of Malawi: Solar Resource Atlas*. 133649 (2018).

[59] Sundu, Y. Should Malawi look to wind for energy solution? *The Nation Online* (2017). URL `https://mwnation.com/malawi-look-wind-energy-solution/`.

[60] DTU Technical University of Denmark. Global Wind Atlas (2018). Funded by Danish Energy Agency EUDP 11-II, Globalt Vind Atlas J.nr. 64011-0347.

[61] Francois, B. *et al.* Integrating Hydropower and Intermittent Climate-Related Renewable Energies: A Call for Hydrology. *Hydrol. Process* **28**, 5465—-5468 (2014).

[62] Sterl, S., Liersch, S., Koch, H., Lipzig, N. P. M. V. & Thiery, W. A New Approach for Assessing Synergies of Solar and Wind Power: Implications for West Africa. *Environmental Research Letters* (2018).

[63] Rogeau, A., Girard, R. & Kariniotakis, G. A Generic Gis-Based Method for Small Pumped Hydro Energy Storage (Phes) Potential Evaluation At Large Scale. *Applied Energy* **197**, 241–253 (2017).

[64] Sterl, S., Van Lipzig, N., Van Griensven, A. & Thiery, W. Flexible Hydro-Solar-Wind Mixes for Reliable Power Supply in west Africa. In *Geophysical Research Abstracts*, vol. 21 (2019).

[65] Barasa, M., Bogdanov, D., Oyewo, A. S. & Breyer, C. A cost optimal resolution for Sub-Saharan Africa powered by 100% renewables in 2030. *Renewable and Sustainable Energy Reviews* **92**, 440–457 (2018).