Water Resources Management (2020) 34:1725–1741
https://doi.org/10.1007/s11269-020-02524-5

Identifying the Links Among Poverty, Hydroenergy and Water Use Using Data Mining Methods

Fuyou Tian[1,2] · Bingfang Wu[1,2] · Hongwei Zeng[1,2] · Shukri Ahmed[3] · Nana Yan[1] · Ian White[4] · Miao Zhang[1] · Alfred Stein[5]

Received: 4 September 2019 / Accepted: 8 March 2020 /
Published online: 25 March 2020

Abstract

Water is fundamental to human well-being, social development and the environment. Water development, particularly hydropower, provides an important source of renewable energy. Water development is strongly affected by poverty, but only few attempts have been made to understand the links between water development and poverty from a global water development point of view. In this work, this linkage was explored using reservoir construction, hydroenergy and water use data along with six derived indicators. We used association rule mining and classification and regression trees (CART) to identify the links. Random forests were employed to search for factors sensitive to poverty. This study shows that the reservoir density is significantly related to poverty, and reservoir densities are lower in countries with higher poverty rates. Countries with a higher use of small hydropower (SHP) systems are generally more prosperous as follows: an SHP utilization

Shukri Ahmed the information and views expressed are the authors' and do not necessarily represent FAO's views, positions, strategies or opinions.

Highlights

1. We investigated the water and poverty links from a global water development view of poverty alleviation.
2. Data mining methods was applied to explore the links between water development and poverty rates in a world-wide context.
3. We find that the ratio of water utilization, water availability per capita and reservoir density contributed most to predict poverty class and related to poverty alleviation.

✉ Bingfang Wu
wubf@radi.ac.cn

1 State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China

2 University of Chinese Academy of Sciences, Beijing 100049, China

3 UN FAO, Rome, Italy

4 Fenner School of Environment & Society, Australian National University, ACT, Canberra 0200, Australia

5 Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7514AE, Enschede, The Netherlands

rate above 27% corresponds to a poverty rate below 4.9%. The ratio of water utilization, water availability per capita (WAPC) and reservoir density were essential for the prediction of the poverty class. All three ratios could be related to poverty alleviation as they enable the identification of the potential for water resource development and their constraints. This study concludes that water development in poor countries needs to receive more attention.

Keywords Poverty · Association analysis · Water resource development · Sustainable development

1 Introduction

Poverty, which is the first target of the Sustainable Development Goals (SDGs), is a multi-faceted phenomenon (Watmough et al. 2016). Poverty is constrained by the presence of natural resources and environmental and energy dimensions (Alkire and Santos 2014). In developing countries, the poor largely rely on natural resources, such as forests, agriculture (Pingali 2012) and water, as basic energy sources. Water is essential for human subsistence, social development and poverty eradication (Mehta 2014) and is a central component of the 17 Sustainable Development Goals (Hanjra and Ferede 2009).

Water is a key factor in providing cheap and clean electricity and promoting food production (Fig. 1a) (Chen et al. 2016). Several studies investigating the linkages between poverty and water have been carried out. The linkage between investments in water and poverty reduction in Africa is explicit, and investments in water for agriculture is key to breaking the poverty trap (Hanjra and Ferede 2009). Nevertheless, few attempts have been made to understand the links between water and poverty from a global water development perspective to address poverty alleviation. Kemp-Benedict found that the connection among poverty, water and agriculture varies among basins based on an investigation of ten river basins (Kemp-Benedict et al. 2011).

Exploring the connections at a scale beyond basins is useful for better understanding the links between water development and poverty alleviation in a global context. This approach can provide an alternative perspective and improve the understanding of the role of water development in poverty alleviation. The potential of water usage, such as for the growth of subsistence crops, is linked to poverty alleviation. However, the linkage between poverty and water is complex and poorly understood thus far (Kemp-Benedict et al. 2011). The stages of development of water-supply basins are important for any analysis of the links between water and poverty (Mehta 2014). This research investigates the links between poverty and water development from a global perspective.

The development of water resources involves the direct usage of water and hydroenergy (Fig. 1a). Water use involves both agricultural and non-agricultural sectors (Hejazi et al. 2014b). Hydroenergy is mainly produced by spinning turbines to produce hydroelectricity. The withdrawal of water refers to the quantity of water that humans consume from natural sources, such as rivers and lakes (Alcamo et al. 2003) and is considered the proportion of water use directly serving dwellings. Among all sectors of water use, the agriculture sector is the largest, accounting for more than 70% of the total water withdrawn (Chen et al. 2016). Scholars tend to agree that the development of agriculture is essential for alleviating poverty, especially during Asia's Green Revolution

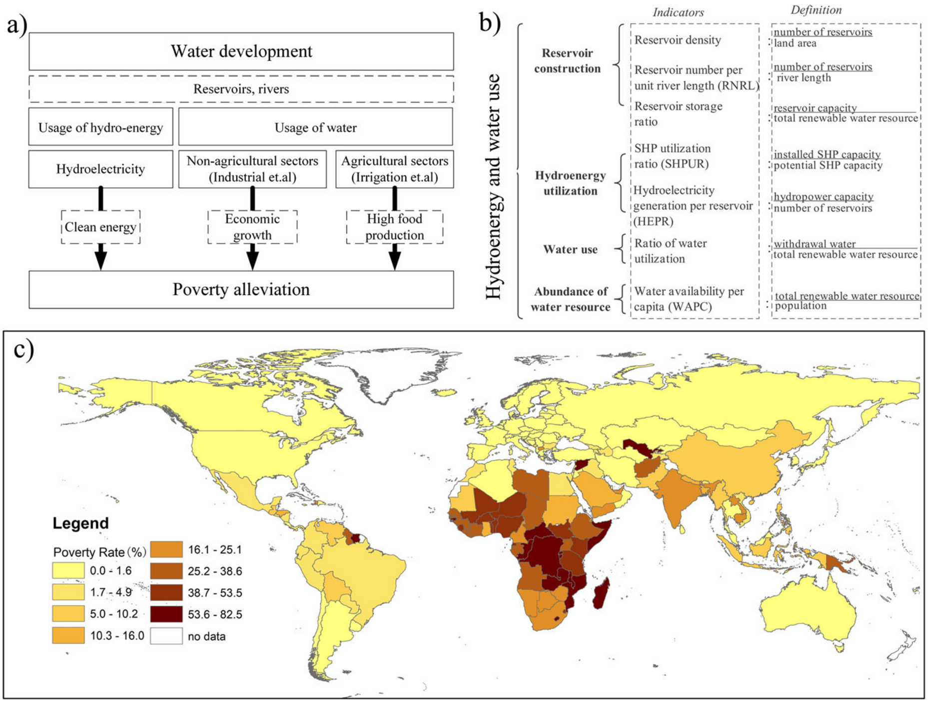


Fig. 1 Water development and poverty alleviation diagram (**a**), and the definitions of the indicators (**b**). Poverty rate as a percentage of the population per country worldwide (**c**)

(Diao et al. 2010), because irrigation is helpful for increasing specific yields (Huang et al. 2006; Hussain et al. 2006). The volume of withdrawn water used in the industry is related to the level of industrial development. Higher industrial development can provide more jobs and boost economic growth, representing an effective path to increasing local household income and reducing the poverty rates (Page and Shimeles 2015).

The development of hydroenergy, i.e., hydropower, provides inexpensive hydroelectricity and increases the access of poorer people to clean energy (Manorom et al. 2017). Hydropower is a cardinal renewable energy source accounting for approximately 71% of the total renewable energy in 2016 (Moran et al. 2018). Starting from the emergency of hydro-turbines in the late nineteenth century, developed countries have built substantial hydropower dams and benefitted from clean and low-cost energy during the industrial revolution (Moran et al. 2018). In the United States, the proportion of hydropower accounted for 40% during the previous century (Moran et al. 2018), but according to the World Bank database (World Bank), 11.13% of the total population, i.e., approximately 0.85 billion people, still lived without electricity in 2018. Free accessibility to electricity is an infrastructure component fundamental to a better quality of life, reduced poverty and higher education and is the pillar of the information revolution based on computers and the Internet.

In this study, we selected several indicators of water development to explore the links between poverty and water development using the data mining method. Similar to the indicators used by Jalilov (Jalilov et al. 2013), water development was assessed based on hydroenergy, water use, and reservoir construction, which is considered the most essential infrastructure for water development (Jalilov et al. 2013; Lehner et al. 2011a). Data mining, i.e., several methods used for knowledge mining from data (Han et al. 2011), was employed, including clustering, classification, and frequency pattern mining. Association rules mining has been applied to urban poverty changes and landcover changes, and several interesting rules have been identified (Mennis and Liu 2005). Due to the vast amount of hydrological data available, data mining methodologies can be used to extract knowledge from a sea of hydrological data (Teegavarapu 2009).

The objectives of this paper were to (1) assess the current status of water development in each country; (2) explore and review the links between poverty and water development and identify the factors associated with global poverty; and (3) determine whether countries with a high poverty rate should continue to develop water resources.

2 Materials and Methods

2.1 Data Source

2.1.1 Poverty rate data

The poverty rate data used in this study were derived from the open websites of the World Bank (World Bank, 2019) and Central Intelligence Agency (CIA) (CIA The world Factbook 2018) of 161 countries worldwide with published poverty rates at the World Bank or the Internet. We consider data from 2016. The poverty rates in Somalia,[1] Saudi Arabia,[2] Oman[3] and Qatar[4] were obtained from news reports or webpages. In the case any 2016 data were absent, the average poverty rates over the last 5 years (2010–2015) were used.

In 2013, the average global extreme poverty rate was 10.7%, indicating that 746 million people live in extreme poverty (Roser and Ortiz-Ospina 2017). Extreme poverty is defined as people living on less than \$US 1.9 day^{-1} (Roser and Ortiz-Ospina 2017). The per country poverty rates are shown in Fig. 1c. Africa is the continent with the largest number of people living in poverty, i.e., 383 million, followed by 327 million people in Asia and 19 million people in South America (Roser and Ortiz-Ospina 2017).

2.1.2 Hydropower potential and development

To evaluate the development of hydropower, total hydropower capacity and small hydropower installation data were used in this research. The total hydropower capacity data were obtained from *World Energy Resources 2016*. The data of small hydropower (SHP) with an installed capacity of less than 50,000 KW were obtained from the *World Small Hydropower Development Report 2016 (United Nations Industrial Development Organization, 2016)*.

[1] Poverty rate derived from https://www.borgenmagazine.com/10-facts-poverty-in-somalia/
[2] Poverty rate derived from http://english.alarabiya.net/en/business/economy/2013/11/03/Kingdom-has-tenth-lowest-poverty-rate-worldwide-says-World-Bank.html
[3] Poverty rate derived from http://timesofoman.com/article/78972
[4] Poverty rate derived from https://en.wikipedia.org/wiki/Economy_of_Qatar

2.1.3 Reservoir construction

The main purposes for constructing reservoirs include hydro-electricity, accounting for 39%, irrigation (accounting for 29%), flood control, and water supply (Lehner et al. 2011a). The Global Reservoir and Dam database (GRanD) (Lehner et al. 2011b) contains 6862 records of reservoirs that have a cumulative storage capacity of 6197 km^3 and includes abundant attributes, such as the dam and reservoir names, construction year, surface area, dam height, main purpose, elevation, spatial coordinates and storage capacity (Lehner et al. 2011b). Although this database mainly focuses on dams and reservoirs with a storage capacity of more than 0.1 km^3, the capacity recorded in the database accounts for approximately 85.2% of all global reservoirs (Lehner et al. 2011a). The dam capacity in 2015 per country was obtained from the FAO AQUASTAT programme.

2.1.4 Water abundance and water use

The relative water uses of a country can be measured by the ratio of the withdrawal of water to the total renewable water. The water withdrawals of 85 countries can be obtained from the FAO AQUASTAT database. The withdrawals by the other 76 countries were estimated by Tethys (Li et al. 2018), which is a Python package for estimating global water withdrawals using the Global Change Assessment Model (GCAM).

The global river data were derived from the World Wildlife Fund's (WWF) HydroSHEDS drainage direction layer and a stream network layer. The raster stream network was determined using the HydroSHEDS flow accumulation grid with a threshold of approximately 100 km^2 upstream area. In addition, the river lengths in Norway, Iceland, Finland, Cape Verde, Comoros, Malaysia, and the Maldives were interpolated based on the land area and the world average national river length of 77.7 m km^{-2}.

The total renewable water resources (TRWR), which were obtained from the FAO AQUASTAT database (Peña-Barragán et al. 2011), represent the sum of the internal renewable water resources and external renewable water resources and correspond to the theoretical maximum annual water available per country at a given time. The TRWR from 2013 to 2017 were used in this research. Other auxiliary socioeconomic data, including the population, GDP, and land area of each country, were obtained from the open Word Bank database (World Bank).

2.2 Indicators of Water Resource Development

We selected six indicators of water resource development among the three categories of constructed reservoirs, hydroenergy utilization and water use. To assess the level of hydropower use, we employed the SHPUR and HEPR. The ratio of annual withdrawal water to total annual renewable water and the water availability per capita (WAPC) indicate water use and availability, respectively (Fig. 1b).

2.3 Association Rules

Association rule learning is a non-parameter rule-based machine learning method used to discover interesting relations or frequent patterns between variables in a database (Agrawal et al. 1993). As introduced by Agrawal (Agrawal et al. 1993), let $I = \{I_1, I_2, \cdots, I_m\}$ be an item set, and let $D = \{t_1, t_2, \cdots, t_n\}$ be the task-relevant data or a set of database transactions. Each transaction in D is a nonempty subset of items I. A rule is defined as an implication of the following form:

$$A \Rightarrow B, \quad \text{where } A \subseteq I, B \subseteq I, A \neq \emptyset, B \neq \emptyset \text{and } A \cap B = \emptyset \tag{2.1}$$

Each rule consists of two different sets of items, A and B, where A is the antecedent or left-hand-side, and B is the consequent or right-hand-side. The rule $A \Rightarrow B$ holds in transaction set D with *support* s, where s is the percentage of transactions in D that contain $A \cup B$ equalling the probability $P(A \cup B)$. The rule $A \Rightarrow B$ has *confidence* c in transaction set D, where c is the percentage of records in D including A that also contain B, which equals the conditional probability $P(B|A)$. The *lift* value, which is an indicator used to filter interesting association rules, is the fraction between the confidence of this rule and the support of the consequent as follows:

$$support(A \Rightarrow B) = P(A \cup B) \tag{2.2}$$

$$confidence(A \Rightarrow B) = P(B|A) \tag{2.3}$$

$$lift(A \Rightarrow B) = \frac{confidence(A \Rightarrow B)}{support(B)} = \frac{P(B|A)}{P(B)} \tag{2.4}$$

A minimum support threshold and minimum confidence threshold are used to filter strong association rules among all rules (Han et al. 2011). The Apriori algorithm (Agarwal and Srikant 1994), which was introduced to find recurrent patterns in a dataset, was employed to mine the interesting relations between water development and poverty. The Apriori algorithm was obtained from GitHub (https://github.com/asaini/Apriori).

Originally, association rule mining was designed to explore categorical data, i.e., not consecutive data. Natural breaks (Jenks 1967), which aims to minimize each class's average deviation from the class mean while maximizing each class's deviation from the means of the other groups (Jenks 1967), was adopted in this research and performed using ArcGIS. The results are shown in Fig. 2g.

2.4 Estimating the Poverty Level with Water Development Metrics

To better understand the nexus between poverty and water development, a classification and regression tree (CART) is performed to estimate the poverty level of each country using water development metrics (Breiman et al. 1984). CART divides the samples at each node to obtain more information. CART acquires a pure subset and enables a fit for a non-linear relationship (Han et al. 2011). Essentially, association rule mining is an unsupervised method used to find frequency rules in a dataset, while CART is a supervised method used to perform classifications or regressions with specific rules. In our study, we intended to discover a direct relationship between water development and poverty in the CART classification rules.

Another useful data mining method is random forest, which is an ensemble of classification trees introduced by Breimen (Breiman 2001). As the name suggests, the random selection of variables and samples was conducted during the model training to avoid the problem of overfitting (Watmough et al. 2016). Random forest selects samples with a bootstrap approach, and only two-thirds of the samples are involved in the model training. The remaining sample is called out-of-bag (OOB) data accounting, and approximately one-third of the samples are used to test the accuracy, representing an unbiased estimation of the overall accuracy. The random selection of variables solves the curse of the dimensionality problem, whereas the variable importance is measured according to the change in accuracy in the OOB data when a permutation of variables is selected as input.

CART and random forest were implemented in *R 3.5.1* using the *tree* (Breiman et al. 1984) and *random forest* (Breiman 2001) packages, respectively. In total, 110 samples were used to train the models after excluding the records with absent data. For CART, a 10-fold cross-validation approach was used to optimize the cost-complexity parameter with the *cv.tree* function, and the tree was pruned with the *prune.misclass* function.

3 Results

3.1 Poverty and Reservoir Construction

3.1.1 Reservoir densities and poverty

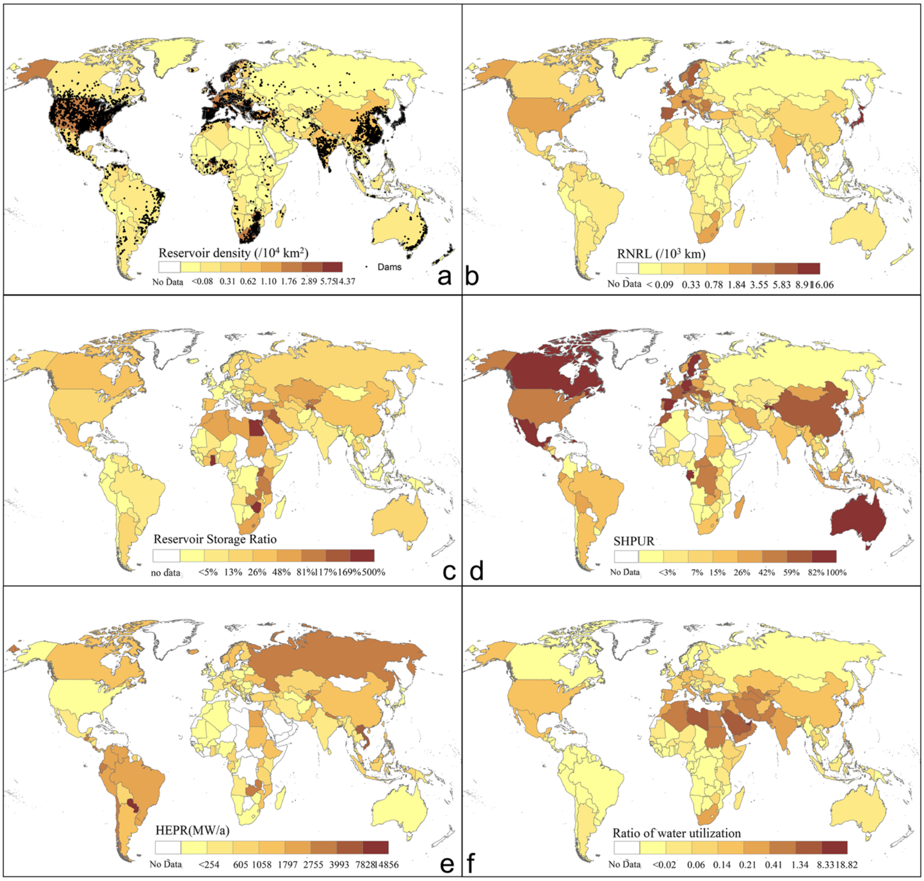
The reservoir density and dams by country are shown in Fig. 2a. Japan (14.37/10^4 km^2), Switzerland (9.20/10^4 km^2), and Portugal (5.75/10^4 km^2) are the three countries with the highest reservoir densities. The association rules are obtained by applying the Apriori algorithm to the poverty rate, reservoir density and water availability per capita with a minimum support of 5% and a minimum confidence of 80% (Table 1).

Based on the association rules between poverty and reservoir density, we observed that the reservoir density in countries with a high poverty rate is low. When the poverty rate exceeds 5%, the reservoir density is generally less than 0.31/10^4 km^2 by integrating rules 1 to 3 (support = 38.5%, confidence = 83.1%), which is the case in Pakistan, Botswana, the Democratic Republic of the Congo, and Kenya (rules 1–3). In contrast, when the reservoir density is high, the poverty rate of countries is generally low. Specifically, when the reservoir density is greater than 1.11/10^4 km^2, the national poverty rate is lower than 4.9% (support = 20.2%, confidence = 83.8%), which is summarized in rules 4 and 5. Japan, the United States of America, Germany, France, Belgium, Italy and other developed countries fall under these 2 rules. According to the three correlation coefficients (Fig. 3a), the reservoir density is significantly related to the poverty rate ($p < 0.01$).

The reservoir density is also related to water abundance. In countries with abundant water resources, such as Brazil and Zambia, the reservoir density is generally low (rule 6). In poor countries with abundant water resources, such as Uzbekistan, Mozambique, and Tanzania, the reservoir density is very low. In more developed countries with abundant water resources, the reservoir density may also be low (rules 7–8). When water availability is relatively abundant and the reservoir density is generally high, the country may be more prosperous (rule 9).

3.1.2 RNRL and poverty

Figure 2b shows the RNRL per country. Cyprus (16.06 /10^3 km), Japan (13.20 /10^3 km), and Switzerland (12.84 /10^3 km) were the top 3 countries with the highest reservoir numbers per unit river length, whereas this indicator was low in Africa, South America and Southeastern and Southern Asia. When the poverty rate, RNRL, and WAPC were used for the mining



g) Result of natural breaks

Class	Poverty rate (%)	Ratio of water utilization	Reservoir storage ratio	SHPUR(%)
0	0-4.9	0-0.06	0-0.13	0-7
1	5-16	0.07-0.21	0.14-0.48	8-26
2	16.1-38.6	0.22-1.00	0.49-1.17	27-59
3	38.7-82.5	1.01-18.82	1.18-5	60-100
Class	**HEPR(MW/a)**	**WAPC ($\times 10^3$ m^3)**	**Reservoir density (/10^4 km^2)**	**RNRL (/10^3 km)**
0	0.0-605.3	0-5.4	0-0.31	0.00-0.33
1	605.4-1796.7	5.5-34.5	0.32-1.10	0.34-1.84
2	1796.8-3993.0	34.6-99.1	1.11-2.89	1.85-5.83
3	3993.1-14856.3	99.2-506.8	2.90-14.37	5.84-16.06

Fig. 2 Reservoir density and dams (**a**), RNRL (**b**), reservoir storage ratio (**c**), SHPUR (**d**), HEPR (**e**), and ratio of water utilization (**f**) by country. Figure (**g**) shows the break points derived from the natural breaks method

association rules, a minimum support of 5% and a minimum confidence of 70% were selected, and the rules produced are shown in Table 1.

The following conclusions can be drawn based on the rules between the poverty rate and RNRL shown in Table 1. In a country with relatively abundant water resources, the RNRL is generally low. When water availability is between 34.6 and 99.1 thousand m^3/cap, the RNRL is less than 0.33/10^3 km (support = 8.7%,

Table 1 Rules between the poverty rate and indicators produced by the association rule mining method

Index	Rules Between PR and RD	Support	Confidence	Lift	Index	Rules between RWU and PR	Support	Confidence	Lift
1	[PR = 3] ⇒ [RD = 0]	13.00%	80.00%	1.33	1	[RWU = 3] ⇒ [WAPC = 0]	4.40%	100.00%	1.93
2	[PR = 2] ⇒ [RD = 0]	16.40%	86.20%	1.43	2	[RWU = 2] ⇒ [WAPC = 0]	16.90%	100.00%	1.93
3	[PR = 1] ⇒ [RD = 0]	9.10%	82.40%	1.37	3	[RWU = 1] ⇒ [WAPC = 0]	15.60%	80.60%	1.56
4	[RD = 3] ⇒ [PR = 0]	8.50%	86.70%	1.62	4	[RWU = 2] ⇒ [PR = 0]	13.10%	77.80%	1.48
5	[RD = 2] ⇒ [PR = 0]	11.70%	81.80%	1.62	5	[PR = 3] ⇒ [RWU = 0]	13.80%	81.50%	1.37
6	[WAPC = 2] ⇒ [PR = 0]	9.80%	88.20%	1.47	6	[WAPC = 3] ⇒ [RWU = 0]	2.50%	100.00%	1.69
7	[WAPC = 1 & PR=3] ⇒ [RD = 0]	5.20%	100.00%	1.66	7	[WAPC = 2] ⇒ [RWU = 0]	2.50%	10.000%	1.69
8	[WAPC = 2 & PR = 0] ⇒ [RD = 0]	5.90%	100.00%	1.66	8	[WAPC = 1] ⇒ [RWU = 0]	31.30%	89.30%	1.51
9	[WAPC = 1 & RD = 1] ⇒ [PR = 0]	5.80%	88.80%	1.66					
Index	Rules between RNRL and PR	Support	Confidence	Lift	Index	Rules between SHPUR and PR	Support	Confidence	Lift
1	[PR = 1] ⇒ [RNRL = 0]	9.30%	83.30%	1.39	1	[PR = 2] ⇒ [SHPUR = 0]	11.00%	65.20%	1.85
2	[PR = 2] ⇒ [RNRL = 0]	16.80%	87.10%	1.45	2	[PR = 3] ⇒ [SHPUR = 0]	8.80%	57.10%	1.62
3	[PR = 3] ⇒ [RNRL = 0]	13.80%	81.40%	1.36	3	[SHPUR = 2] ⇒ [PR = 0]	14.70%	66.70%	1.18
4	[WAPC = 2] ⇒ [RNRL = 0]	8.70%	82.40%	1.37	4	[SHPUR = 3] ⇒ [PR = 0]	19.10%	89.70%	1.58
5	[RNRL = 1] ⇒ [PR = 0]	12.50%	76.90%	1.46	5	[SHPUR = 1] ⇒ [WAPC = 0]	12.50%	58.60%	1.24
6	[RNRL = 2] ⇒ [PR = 0]	15.70%	83.30%	1.59	6	[PR = 3] ⇒ [WAPC = 0]	9.50%	61.90%	1.31
7	[RNRL = 3] ⇒ [PR = 0]	4.40%	87.50%	1.67					
Index	Rules between RSR and PR	Support	Confidence	Lift	Index	Rules between HEPR and PR	Support	Confidence	Lift
1	[RSR = 2] ⇒ [WAPC = 0]	9.40%	87.00%	1.73	1	[WAPC = 0] ⇒ [HEPR = 0]	30.80%	65.50%	1.32
2	[WAPC = 2] ⇒ [RSR = 0]	9.40%	93.00%	1.58	2	[HEP = 2] ⇒ [PR = 0]	11.10%	76.50%	1.24
3	[WAPC = 2 & PR = 0] ⇒ [PR = 0]	5.10%	87.50%	1.48	3	[PR = 2] ⇒ [HEPR = 0]	8.50%	66.70%	1.35
4	[WAPC = 1 & RSR = 1] ⇒ [PR = 0]	7.20%	90.90%	1.60					

PR poverty rate, *RD* reservoir density, *RSR* reservoir storage ratio, *RWU* ratio of water utilization, *RNRL* reservoir number per unit river length, *SHPUR* small hydropower utilization rate, *WAPC* water availability per capita,*HEPR* hydroelectricity generation per reservoir

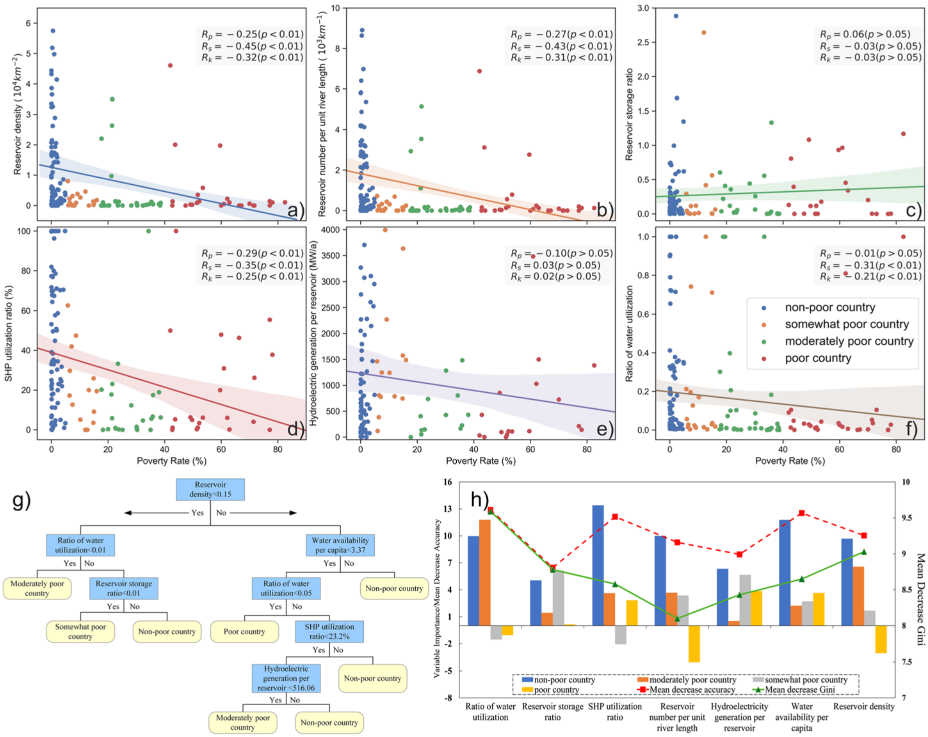


Fig. 3 The relationship between the poverty rate and reservoir density (**a**), RNRL (**b**), reservoir storage ratio (**c**), SHPUR (**d**), HEPR (**e**), and ratio of water utilization (**f**). R_p: Pearson correlation coefficient, R_S: Spearman's rank correlation coefficient, and R_k: Kendall tau rank correlation coefficient. The corresponding p-values are shown in the figure. The decision tree was derived from CART with the "*tree*" package in *R 3.5.1* (**g**). The importance of the variables in each poverty class is derived from random forest (**h**)

confidence = 82.4%), which is the case in Bolivia, Brazil, Canada, Chile, Colombia, and the Lao People's Democratic Republic (rule 4).

In countries with a high poverty rate, the RNRL is generally low. When the poverty rate is greater than 5%, the RNRL is usually less than $0.33/10^3$ km (support = 39.9%, confidence = 84.2%), which is the case in Bangladesh, Cambodia, the Lao People's Democratic Republic, Malawi, and Tanzania, as revealed by rules 1 to 3 and verified in Fig. 3b. However, in countries with a high RNRL, the poverty rate is relatively low. Integrating rules 5 to 7 implies that the poverty rate is generally less than 4.9% when the RNRL is greater than $0.34/10^3$ km (support = 32.6%, confidence = 81.2%), which is the case in France, Germany, the United States and the United Kingdom. According to rules 5 to 7, the confidence of this association increases from 76.9% to 87.5% as the class of RNRL changes from class 1 to class 3. Overall, the RNRL is strongly associated with the poverty rate ($p < 0.01$) (Fig. 3b). Additionally, if the poverty rate is high in a country, the RNRL is generally small.

3.1.3 Reservoir storage ratios and poverty

Zimbabwe (499.65%), Egypt (288.51%), and Ghana (264.23%) were the top 3 countries with the highest reservoir storage ratio worldwide (Fig. 2c). Assuming that the global reservoir storage was 8070 cubic km (Lehner et al. 2011a), the global reservoir storage ratio was approximately 14.88%. Combined with the poverty rate and WAPC, the reservoir storage ratio was imported into the Apriori algorithm with a minimum support of 5% and a minimum confidence of 70%. The output of the association rules is shown in Table 1.

The reservoir storage ratio was mainly affected by the abundance of water resources and had no significant association rule with poverty (Fig. 3c). In countries with a high reservoir storage ratio, water resources are relatively scarce (rules 1–2). In contrast, if a country is rich in water resources, the reservoir storage ratio is relatively low.

3.2 Poverty and Hydroenergy Utilization

3.2.1 SHPUR and poverty

Small hydropower (SHP) has matured into an economically feasible technology with a minimal impact on the environment (United Nations Industrial Development Organization and International Center on Small Hydro Power). The SHPUR estimates the development of hydropower by country. Some countries, such as Denmark, Tajikistan, Mexico, Switzerland, Sweden, Australia and Canada, have exploited almost all potential SHP (Fig. 2d). The globally installed SHP capacity was estimated at 78 GW in 2016, accounting for approximately 36% of the total global SHP potential. The results of the Apriori algorithm are listed in Table 1.

In a country where the poverty rate is greater than 6.1%, the SHPUR is generally less than 7% (support = 19.8%, confidence = 61. 3%), which is illustrated in rules 1 to 2 (Table 1). Based on rules 3 and 4, we conclude that countries with a higher SHPUR generally have a lower poverty rate. When the utilization rate is greater than 27%, the poverty rate is usually less than 4.9% (support = 33.8%, confidence = 78.0%). As the SHPUR increases from 2 to 3, the confidence of this rule increases from 66.7% to 89.7%.

The association rules between SHP development and the abundance of water are shown in rules 5 and 6. The fifth rule indicates that WAPC is less than 5.4 thousand m^3 if the SHPUR ranges from 8% to 26% (support = 12.5%, confidence = 58.6%). Overall, there are fewer significant correlation rules between WAPC and SHP utilization, indicating that the relationship between poverty and SHP development is closer ($p < 0.01$) (Fig. 3d).

3.2.2 HEPR and poverty

HEPR (Fig. 2e) reveals the electricity-generating capacity of each reservoir. Table 1 shows the results of the association rules among poverty, HEPR, and WAPC with a minimum support of 5% and a minimum confidence of 65%.

The HEPR is less than 605.3 MW/a if the WAPC is less than 5.4 thousand m^3 (support = 30.8%, confidence = 65.5%). When the power generation per reservoir is between 1796.8 and 3993.0 MW/a, the poverty rate is generally low (support = 11.1%, confidence = 76.5%). According to rule 3, the conclusion is that if the poverty rate ranges from 16.1% to 38.6%, HEPR is less than 605.3 MW/a. Basically, this indicator is associated with both water availability and poverty, but the association is not very strong (Fig. 3e).

3.3 Poverty and Water Use

The United Arab Emirates, Libya, and Saudi Arabia are the top 3 countries with high ratios of water utilization (Fig. 2f). Table 1 illustrates the results of the association rules among the poverty rate, the ratio of water utilization, and WAPC with a minimum support of 5% and a minimum confidence of 70%.

Rules 1 to 3 in Table 1 state that in countries with a high ratio of water utilization, water resources are generally scarce. When the ratio of water utilization is higher than 7%, the WAPC in the country is generally less than 5.4 thousand m^3 (support = 36.9%, confidence = 90.8%). Rules 4 and 5 illustrate the relationship between the ratio of water utilization and poverty. In 77.8% of the countries with a ratio of water utilization between 0.22 and 1.0, the poverty rate is less than 4.9%, while in 81.5% of the countries with a poverty rate higher than 38.7%, the ratio of water utilization is lower than 6%. When people can use more than 5.5 thousand m^3 per year, the water utilization ratio is likely to be lower than 6% (support = 44.4%, confidence = 85.5%). In brief, the ratio of water utilization is affected by two factors, namely, WAPC and poverty, and the effect of water availability may be stronger if it involves more association rules. The ratio of water utilization relates to poverty but not through a linear relationship (Fig. 3f), considering a lower R_p of −0.01, a higher R_S of −0.31 and R_k of −0.21 ($p < 0.01$).

3.4 Poverty and Water Development

The accuracy of the CART tree in predicting the poverty class based on water development factors is 79.1%, and the pruned decision tree is plotted in Fig. 3g. In total, eight rules are set to classify the poverty class with a maximum depth of five and a minimum depth of two. If the reservoir density is less than $0.15/10^4$ km^2 and the ratio of water utilization and reservoir storage ratio are low, the country may be moderately or somewhat poor. In the decision tree, the moderately poor, somewhat poor and poor countries appear on the left side, while the non-poor countries are always located on the right side of the node, indicating that non-poor countries have a higher development of water resources.

In the random forest, the number of trees (*ntree*) was set to 600 because the accuracy converges when the number of trees is larger than 400, and the increase in this parameter ensures the robustness of the variable importance metrics (Genuer et al. 2008). The accuracy of OOB in the random forest was 67.4%. As shown in Fig. 3h, the ratio of water utilization, WAPC and reservoir density contributed the most to predicting the poverty class and are related to poverty alleviation according to the mean decrease in accuracy and Gini. Among the non-poor countries, the SHPUR was the most important feature, while the ratio of water utilization was paramount for the moderately poor countries, followed by the reservoir density. The reservoir storage ratio and HEPR were paramount for somewhat poor countries and poor countries, respectively.

4 Discussion and Conclusion

4.1 General Links between Water Development and Poverty

This study explored the links between water development and poverty rates using data mining methods in a worldwide context. By focusing on water development, this study added to the perspective of "water poverty" (Sun et al. 2018), which assesses accessibility to water. Using a quantitative association and correlation analysis, we found that the reservoir density, RNRL, SHPUR, and ratio of water utilization were highly linked to a country's poverty rate likely because hydropower can contribute to solving problems associated with providing rural electrification, improving living standards and production conditions, promoting rural economic development, and reducing emissions (Zhou et al. 2009). Nevertheless, the SHPUR in both Southern Asia and Southeastern Asia is only 17%. In contrast, SHP has reached 61% and 48% in North America and Europe (Fig. 4a), respectively, which are affluent regions. Great potential remains to be developed in South America, Africa and Southern and Southeastern

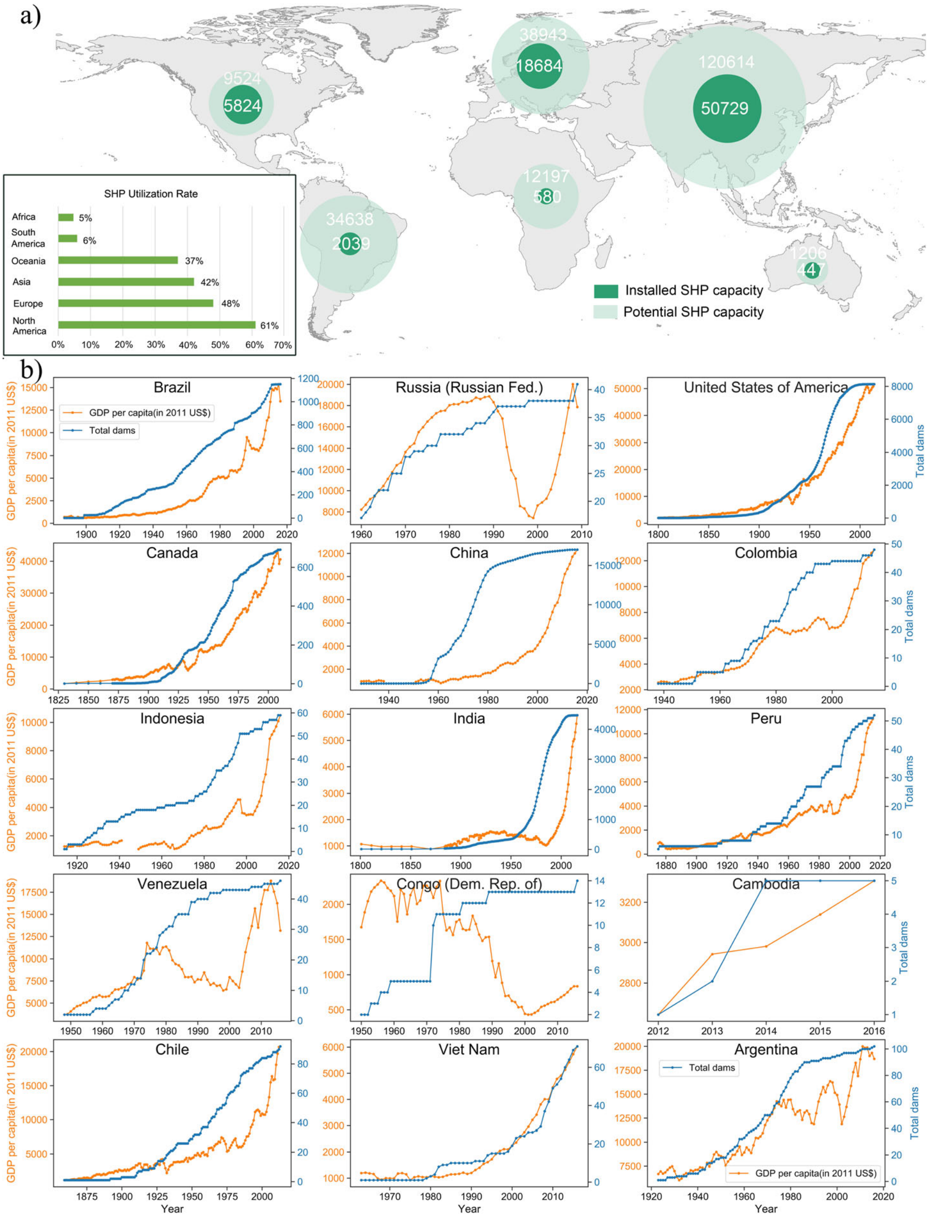


Fig. 4 SHP utilization per continent (**a**) and the relationships between GDP per capita (in 2010 US$, source: Maddison Project Database (Bolt et al. 2018)) and the construction of reservoirs in the top 16 countries with the most total renewable water resources

Asia as Brazil reached only 22%, Kenya reached only 1.7%, Angola reached only 1.5% and Cambodia reached only 0.3% in 2016 (United Nations Industrial Development Organization and International Center on Small Hydro Power); these countries also have high poverty rates.

Although the ratio of water utilization is affected by WAPC, it is low in extremely poor regions. Water use includes agricultural, industrial and domestic water consumption. Irrigation accounts for approximately 70% of the total water withdrawal (Hejazi et al. 2014a). Some scholars found that investment in irrigation can contribute to poverty alleviation in China and Africa (Hanjra and Ferede 2009; Huang et al. 2006). However, the percentage of the irrigation area is low in Central South America and Africa (Siebert et al. 2015). Therefore, more investments in agriculture irrigation are needed to improve the ratio of water utilization, such as in the Central African Republic (using 0.1% of its total renewable water), Bolivia (0.4%), Mozambique (0.7%) and Brazil (0.9%).

The linkage between water development and poverty discovered by the association rule mining differs from that obtained by traditional causality rules and is suitable for all countries. However, the confidence of the association rules rarely reaches 100%. The limited confidence suggests that the association rules are sub-optimal for individual countries but are able to show the global prevailing pattern. Data mining methods aim to find a pattern hidden in the data and provide alternative comprehensive results. For example, one should be careful in interpreting the linkage between the SHPUR and poverty such that a high SHPUR leads to a low poverty rate or that a low poverty rate results in a low poverty rate because we did not fully understand the linkage. We only found a linkage and deliver some conjectures. Hence, more work could be considered to explore the mechanism responsible for the association between water development and poverty reduction.

In this paper, the level of poverty was estimated by a consumption-based monetary threshold of \$1.9 day^{-1}, which is defined by the World Bank as the international poverty line. However, this consumption-based indicator does not reflect living conditions from all aspects (Steele et al. 2017). Using the same poverty criteria ensured comparability but neglected the differences in the living standard across the countries (Alkire and Santos 2014), which may slightly bias the poverty level results; however, after discretization with a cluster analysis, this monetary threshold bias may have a limited effect.

4.2 Strong Links between Reservoir Construction and Poverty

Reservoirs represent the most important infrastructure for hydropower and water utilization in most countries (Yüksel 2010). We found strong links between poverty and the reservoir density. Because reservoirs are among the main types of water infrastructure in countries with sufficient surface water, they play a crucial role in the control and management of water resources (Gao et al. 2012). However, there are serious concerns about dam performance in alleviating poverty. The construction of dams changes aquatic ecosystems and leads to the loss of the habitats of some species (Wilcove and Wikelski 2008). Additionally, it has been estimated that 40–80 million indigenous people have been replaced during the building of dams globally (World Commission on Dams 2000). However, dam construction never stops in prosperous countries until the best sites for dams have been developed (Moran et al. 2018). Figure 4b shows the relationship between the GDP and construction of reservoirs in the top 16 countries with the most total renewable water resources, except for Bangladesh, which only has one reservoir recoded in the GRanD database, and reveals that an increase in GDP is accompanied by an increase in the number of reservoirs.

Therefore, we need to balance the trade-off between economic growth and environmental conservation. The construction of reservoirs is necessary in countries with significant water resources to improve the irrigation percentage, ratio of water utilization and hydropower utilization. Accordingly, additional effort from governments is needed to minimize the effects on the environment and ecosystems and improve the condition of the environment. For example, the Three Gorges Dam has created some challenges for the environment and ecological systems, but great efforts have been exerted to make it an environmentally friendly dam, such as controlling the water quality, regulating sediment, and conserving biodiversity (Fu et al. 2010). Reservoir construction is not feasible in regions with less runoff and a high evaporative index, such as the countries in Northern Africa and Western Asia and island countries, such as the Maldives.

5 Conclusion

In this paper, a quantitative linkage is discovered between water development and poverty by association rules. These association rules show that the reservoir density in countries with a high poverty rate is generally low because reservoirs can comprehensively enhance water development in hydroelectricity utilization and water use in countries with sufficient surface runoff to fill reservoirs and turn turbines. Moreover, countries with a higher SHPUR are generally more prosperous. If the utilization rate exceeds 27%, the poverty rate is usually less than 4.9%. Based on the feature importance metrics derived from the random forests, the ratio of water utilization, WAPC and reservoir density contributed the most to predicting the poverty class and are related to poverty alleviation. Therefore, water development needs more attention because the development of countries with a high poverty rate is highly constrained by water development. Additionally, the development of water is not the only one aspect that constrains poor countries, and there are many other shortcomings, such as education and infrastructure.

Acknowledgements We thank the World bank, FAO, World Energy Council and International Center on Small Hydro Power for providing the data used in this research. This study was financially supported by the National Key Research and Development Program (2016YFA0600304) and the National Natural Science Foundation of China (41561144013).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflicts of interest.

References

Agarwal R, Srikant R (1994) Fast algorithms for mining association rules. In: Proc. of the 20th VLDB Conference, pp 487–499

Agrawal R, Imieliński T, Swami A (1993) Mining association rules between sets of items in large databases. ACM SIGMOD Rec 2:207–216

Alcamo J, Döll P, Henrichs T, Kaspar F, Lehner B, Rösch T, Siebert S (2003) Development and testing of the WaterGAP 2 global model of water use and availability. Hydrol Sci J 48:317–337. https://doi.org/10.1623/hysj.48.3.317.45290

Alkire S, Santos ME (2014) Measuring acute poverty in the developing world: robustness and scope of the multidimensional poverty index. World Dev 59:251–274. https://doi.org/10.1016/j.worlddev.2014.01.026

Bolt J, Inklaar R, de Jong H, van Zanden JL (2018) Maddison project database, version 2018 rebasing 'Maddison': new income comparisons and the shape of long-run economic development
Breiman L (2001) Random forests. Mach Learn 45(1):5–32
Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and regression trees. CRC press
Chen J, Shi H, Sivakumar B, Peart MR (2016) Population, water, food, energy and dams. Renew Sust Energ Rev 56:18–28. https://doi.org/10.1016/j.rser.2015.11.043
CIA The world Factbook. https://www.cia.gov/library/publications/the-world-factbook/fields/2046.html. Accessed 20/06 2018
Diao X, Hazell P, Thurlow J (2010) The role of agriculture in African development. World Dev 38:1375–1383. https://doi.org/10.1016/j.worlddev.2009.06.011
Fu B-J et al (2010) Three gorges project: efforts and challenges for the environment. Prog Phys Geogr 34:741–754
Gao H, Birkett C, Lettenmaier DP (2012) Global monitoring of large reservoir storage from satellite remote sensing. Water Resour Res 48. https://doi.org/10.1029/2012wr012063
Genuer R, Poggi J-M, Tuleau C (2008) Random Forests: some methodological insights. arXiv preprint arXiv:08113619
Han J, Pei J, Kamber M (2011) Data mining: concepts and techniques. Elsevier
Hanjra MA, Ferede T, Gutta DG (2009) Reducing poverty in sub-Saharan Africa through investments in water and other priorities. Agric Water Manag 96:1062–1070. https://doi.org/10.1016/j.agwat.2009.03.001
Hejazi M et al (2014a) Long-term global water projections using six socioeconomic scenarios in an integrated assessment modeling framework. Technol Forecast Soc Chang 81:205–226
Hejazi MI et al (2014b) Integrated assessment of global water scarcity over the 21st century under multiple climate change mitigation policies. Hydrol Earth Syst Sci 18:2859–2883. https://doi.org/10.5194/hess-18-2859-2014
Huang Q, Rozelle S, Lohmar B, Huang J, Wang J (2006) Irrigation, agricultural performance and poverty reduction in China. Food Policy 31:30–52
Hussain I, Wijerathna D, Arif SS, Murtiningrum, Mawarni A, Suparmi (2006) Irrigation, productivity and poverty linkages in irrigation systems in Java, Indonesia. Water Resour Manag 20:313–336. https://doi.org/10.1007/s11269-006-0079-z
Jalilov S-M, Amer SA, Ward FA (2013) Water, food, and energy security: an elusive search for balance in Central Asia. Water Resour Manag 27:3959–3979. https://doi.org/10.1007/s11269-013-0390-4
Jenks GF (1967) The data model concept in statistical mapping. Int Yearb Cartogr 7:186–190
Kemp-Benedict E et al (2011) Connections between poverty, water and agriculture: evidence from 10 river basins. Water Int 36:125–140
Lehner B et al (2011a) High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. Front Ecol Environ 9:494–502
Lehner B et al (2011b) Global reservoir and dam database, version 1 (GRanDv1): dams, revision 01. NASA Socioeconomic Data and Applications Center (SEDAC), Palisades
Li X, Vernon CR, Hejazi MI, Link RP, Huang Z, Liu L, Feng L (2018) Tethys – a python package for spatial and temporal downscaling of global water withdrawals. J Open Res Softw 6. https://doi.org/10.5334/jors.197
Manorom K, Baird IG, Shoemaker B (2017) The World Bank, hydropower-based poverty alleviation and indigenous peoples: On-the-ground realities in the Xe Bang Fai river basin of Laos. In: Forum for Development Studies. vol 2. Taylor & Francis, pp 275–300
Mehta L (2014) Water and human development. World Dev 59:59–69. https://doi.org/10.1016/j.worlddev.2013.12.018
Mennis J, Liu JW (2005) Mining association rules in spatio-temporal data: an analysis of urban socioeconomic and land cover change. Trans GIS 9:5–17
Moran EF, Lopez MC, Moore N, Muller N, Hyndman DW (2018) Sustainable hydropower in the 21st century. Proc Natl Acad Sci U S A 115:11891–11898. https://doi.org/10.1073/pnas.1809426115
Page J, Shimeles A (2015) Aid, employment and poverty reduction in Africa. Afr Dev Rev 27:17–30
Peña-Barragán JM, Ngugi MK, Plant RE, Six J (2011) Object-based crop identification using multiple vegetation indices, textural features and crop phenology. Remote Sens Environ 115:1301–1316
Pingali PL (2012) Green revolution: impacts, limits, and the path ahead. Proc Natl Acad Sci U S A 109:12302–12308. https://doi.org/10.1073/pnas.0912953109
Roser M, Ortiz-Ospina E (2017) Global extreme poverty. https://ourworldindata.org/extreme-poverty#the-evolution-of-poverty-by-world-regions
Siebert S, Kummu M, Porkka M, Döll P, Ramankutty N, Scanlon BR (2015) A global data set of the extent of irrigated land from 1900 to 2005. Hydrol Earth Syst Sci 19:1521–1545. https://doi.org/10.5194/hess-19-1521-2015

Steele JE et al (2017) Mapping poverty using mobile phone and satellite data. J R Soc Interface 14. https://doi.org/10.1098/rsif.2016.0690
Sun C, Wu Y, Zou W, Zhao L, Liu W (2018) A rural water poverty analysis in China using the DPSIR-PLS model. Water Resour Manag 32:1933–1951. https://doi.org/10.1007/s11269-017-1819-y
Teegavarapu RS (2009) Estimation of missing precipitation records integrating surface interpolation techniques and spatio-temporal association rules. J Hydroinf 11:133–146
United Nations Industrial Development Organization, International Center on Small Hydro Power, The World Small Hydropower Development Report 2016. https://www.unido.org/sites/default/files/2016-11/WSHPDR_Executive_Summary_2016_0.pdf
Watmough GR, Atkinson PM, Saikia A, Hutton CW (2016) Understanding the evidence base for poverty–environment relationships using remotely sensed satellite data: an example from Assam, India. World Dev 78:188–203
Wilcove DS, Wikelski M (2008) Going, going, gone: is animal migration disappearing. PLoS Biol 6:e188
World Bank. World Bank Open Data. https://data.worldbank.org/. Accessed 1/23 2019
World Commission on Dams (2000) Dams and development: a new framework for decision-making: the report of the world commission on dams. Earthscan
Yüksel I (2010) Hydropower for sustainable water and energy development. Renew Sust Energ Rev 14(1):462–469
Zhou S, Zhang X, Liu J (2009) The trend of small hydropower development in China. Renew Energy 34:1078–1083. https://doi.org/10.1016/j.renene.2008.07.003