World Development 114 (2019) 208–219

Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

Resettlement lies: Suggestive evidence from 29 large dam projects

Julian Kirchherr [a,*], Mats-Philip Ahrenshop [b,c], Katrina Charles [d]

[a] Faculty of Geosciences, Utrecht University, the Netherlands
[b] Department of Politics and International Relations (DPIR), University of Oxford, Oxford, United Kingdom
[c] Institute of Political Science (IfPol), University of Münster, Germany
[d] School of Geography and the Environment, University of Oxford, Oxford, United Kingdom

ARTICLE INFO

Article history:
Accepted 9 October 2018
Available online 18 October 2018

Keywords:
Resettlement
Social impacts
Hydropower
Overruns
Large dams
Project management

ABSTRACT

Reports suggest that up to 80 million people have been resettled due to the construction of large dams in the past century. Published resettlement data regarding large dam projects comes from different sources, with numbers that can be greatly dispersed. We have examined resettlement data for 29 large dam projects gathering and analysing up to 43 datum per dam project in our sample. We find that dispersion is influenced by the project cycle (with resettlement figures from the planning and design phase found to be lowest), the stakeholders releasing it (with resettlement figures released by project advocates lower than those of project opponents), the political regime (with highest data dispersion found in hybrid regimes and limited dispersion found both in democratic and autocratic regimes) and with time of completion (with dispersion increasing for dams completed since 2010). Our findings thus present some suggestive evidence for the political perspective within the project management literature which emphasizes the contested nature of truth and knowledge. Overall, our study highlights that data in the dam industry and resettlement data must be treated with caution. Furthermore, it provides the starting point for the development of a tool that helps to de-bias resettlement data provided for large dam projects.

1. Introduction

A boom in hydropower dam development is under way. At least 3,700 hydropower dams (each with a capacity of at least 1 MW) are planned or already under construction (Zarfl, Christiane, & Laura Tydecks, 2015). These are projected to increase global hydropower capacity by 73% to 1,700 GW until 2030 (Zarfl et al., 2015), with 93% of the capacity additions stemming from only 847 extremely large dams (each with a capacity of at least 100 MW) which are particularly complex to construct (Zarfl et al., 2015). The largest planned dam is the 44 GW Grand Inga Dam in the Democratic Republic of the Congo (Green, Sovacool, & Hancock, 2015). Its capacity would exceed the capacity of the world's largest completed dam, the Three Gorges Dam, by 21.5 GW (Showers, 2009; Wilmsen, & van Hulten, 2017).

The positive impacts of hydropower dams can be vast. For instance, the Grand Inga Dam is projected to provide electricity for half of the African continent (Green et al., 2015; Showers, 2009). Meanwhile, the completed Itaipu Dam on the border between Brazil and Paraguay provides 78% of Paraguay's annually used electricity and 26% of Brazil's (Folch, 2013). Overall, hydropower provides 16% of the world's global electricity and 85% of the world's renewable electricity (IEA, 2016). Additional benefits that can be provided by dams are (inter alia) flood control and irrigation (Kirchherr & Charles, 2016). For instance, dams are claimed to provide "irrigation water for feeding 800 million people" (Nombre, 2014). These direct positive social impacts of dams can induce multiple indirect positive impacts – from improved nutrition to enhanced incomes, flood protection, water storage for irrigation, additional employment opportunities and, ultimately, accelerated economic growth (Cernea, 2004; Bhatia, Cestti, Scatasta, & Malik, 2008).

Yet these positive impacts can be offset by dams' many negative impacts. The most known negative impacts may be those due to dam-induced displacement (Biswas, 2012; Scudder, 2012). WCD (2000) reported that up to 80 million people have been displaced because of dam construction in the past century; Cernea (2004) estimated at least 22.5 million people have been displaced in China by dams alone. Meanwhile, Scudder (2012) finds from a 50 dam survey, the most comprehensive survey on resettlement outcomes of large dams carried out until today, that displacement results in impoverishment for less than 70% of the communities studied. Dams' negative impacts downstream are also increasingly recog-

* Corresponding author at: Faculy of Geosciences, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, the Netherlands.

E-mail addresses: j.kirchherr@uu.nl (J. Kirchherr), mats-philip.ahrenshop@politics.ox.ac.uk (M.-P. Ahrenshop), katrina.charles@ouce.ox.ac.uk (K. Charles).

https://doi.org/10.1016/j.worlddev.2018.10.003
0305-750X/

nized (Kirchherr, Pohlner, & Charles, 2016). Worldwide, 472 million people have been estimated to be affected negatively by their downstream impacts (Richter et al., 2010); most notably, dams block fish migration routes which can reduce fish resources downstream, a threat to food security (Ziv, Baran, Nam, Rodríguez-Iturbe, & Levin, 2012; Baird, Shoemaker, & Manorom, 2015).

These various negative impacts have resulted in many massive anti-dam-protests in the past decades (Kirchherr, Charles, & Walton, 2016; Biswas, 2012). Khagram (2000, p. 83) even argues that "the most dramatic conflicts over how to pursue [...] development" have occurred over the construction of large dams. The most famous anti-dam-campaign may be the campaign against projects constructed on the Narmada River in India which started in the 1980s (Chapman, 2007; Routledge, 2003). Examples of currently contested large dam projects are the (already mentioned) Grand Inga Dam in the Democratic Republic of the Congo (Green et al., 2015; Showers, 2009), the Myitsone Dam in Myanmar (Chan, 2017; Kiik, 2016) as well as the Xayaburi Dam in Laos (Hensengerth, 2015; Yasuda, 2015).

Both the complexity of large dam projects as well as the contestation regarding their construction have resulted in significant dispersion, i.e. a lack of consistency, in cost and schedule data for large dam projects. This dispersion has been demonstrated (inter alios) by Ansar, Flyvbjerg, Budzier, and Lunn (2014), Awojobi and Jenkins (2015) and Sovacool, Gilbert, and Nugent (2014) (further discussed in the next section). The dispersion in resettlement data has not previously been reported. Smyth and Vanclay (2017) summarize the stages of a consultant's approach to defining who will require resettlement, characterizing it as an iterative approach where numbers will change through the process. However, methodologies by which the reported data are estimated from different stakeholders can often be unclear, reflecting the sensitive and contested nature of the projects, and fostering mistrust between stakeholders. The process of resettlement is complex as Vanclay (2017) highlights, describing it as a "multi-dimensional, multi-factor, multi-actor, multi-scalar and multi-level" process that affects the resettled populations in different ways depending on their "vulnerabilities, capacities, positionings and interests". Previous research, e.g. Kirchherr et al. (2016), has had to assume that resettlement data are accurate. Accurate representation of resettlement data is essential, as we detail in Section 5, for three groups: it is essential for those deciding whether to pursue or not to pursue a project, it is essential for ensuring adequate compensation for project-affected communities, and it is essential for scholars working on large dam projects to base their analyses on accurate data.

This study provides evidence via a novel data set containing 862 datum (up to 43 records per dam project) that the resettlement data regarding large dam projects is as disperse as cost and schedule overrun data. Furthermore, this study provides suggestive evidence regarding the determinants of reported resettlement data dispersion. We find that reported resettlement data is influenced by the project cycle, the stakeholders releasing it, the political regime of a country a dam is built in and the time of completion. We envisage that this study particularly contributes to the project management literature which has been frequently criticised for its lack of empirical studies (e.g. by Packendorff, 1995; Svejvig and Andersen, 2015). Furthermore, we believe that knowledge presented in this study regarding reported resettlement data dispersion can help project managers to de-bias resettlement data on large dam projects. However, we report this analysis cognizant of how contentious the relocation process is and that more accurate numbers will not necessarily equate for better outcomes for resettlees.

The remainder of this paper is organized as follows. The next section outlines our theoretical framing for this study. Section 3 describes methods adopted. We present our results in Section 4 and discuss these in Section 5. Meanwhile, our argument is summarized in Section 6 of this paper.

2. Theoretical framing

We embed our study in the project management literature which is a "relatively young field" (Bredillet, 2010, p. 5) that emerged in the mid-1980s (Svejvig and Andersen, 2015). This field is concerned with the management and organization of projects such as dam projects (Jugdev, 2008; Söderlund, 2003). We specifically relate our study to literature within this field that has investigated the dispersion of cost and schedule data for large energy/ infrastructure projects, e.g. Cantarelli, Chantal, and Bent (2013), Flyvbjerg, Holm, and Buhl (2002) and Flyvbjerg, Skamris Holm, and Buhl (2004). We note that this literature usually frames data dispersion as 'overrun' since this framing rests on the assumption that accurate project data can be identified by scholars, an assumption not confirmed by our work (further discussed below). A subset of this literature investigates large dam projects, e.g. Awojobi and Jenkins (2015), Ansar et al. (2014) and Bacon and Besant-Jones (1998). Particularly this literature has informed the various hypotheses that we present below and that we test throughout this study.

Data dispersion is the main dependent variable considered throughout this study. We define it as data lacking consistency in datum for a particular variable in a project of interest (in our case: resettlement induced by a dam project); data consistency, in turn, means obtaining identical datum for a particular variable within a project of interest when collecting data from different sources (e.g. media articles, reports from international donors etc.) and/or at different points in time (Tremblay, 2007; Cole & Frangakis, 2009; Greener, 2011) (further discussed in the next section).

We distinguish between two schools of thoughts that seek to explain data dispersion, a distinction inspired by Wachs (1987) and Flyvbjerg et al. (2002). The first school, the technical perspective, attributes dispersion mostly to planning errors and thus chance; the second school attributes it to strategic misrepresentation and it thus assumes that data dispersion was consciously introduced. The latter category has emerged as the dominant explanation in the recent literature on the topic since numerous studies, e.g. Ansar et al. (2014), Cantarelli et al. (2013) and Flyvbjerg et al. (2002) found that data dispersion was not due to randomness, but systematic error. For instance, Cantarelli et al. (2013, p. 8) write that there is "always an overestimation of traffic demand and an underestimation of costs".

These two schools of thought encompass two different epistemologies with the shift from the technical to the political one thus also including a shift in prevailing epistemological views in the project management literature. Indeed, "we are moving from an old paradigm – positivist – to a more balanced one combining positivism [and] constructivism [...] because the old one is not working anymore", as Bredillet (2010, p. 6) writes. The latter paradigm acknowledges the contested nature of truth and knowledge, that "the 'real' world out there cannot be seen independently of the [...] actors [and] their action" (Furlong and Marsh, 2010, 185).

We developed two hypotheses that can explain data volatility in reported resettlement data that (largely) build on the technical perspective (size, project cycle[1]) as well as three hypotheses that build on the political one (stakeholders, political regimes, time of completion). These are further explained below.

[1] We acknowledge that the project cycle hypothesis can also be attributed to the political perspective, as explained below.

2.1. Size

Scholars such as Flyvbjerg et al. (2004) and Awojobi and Jenkins (2015) have hypothesized, and scholars such as Giezen (2012) and Bacon and Besant-Jones (1998) have evidenced, that larger projects face greater dispersion in cost and schedule overrun data. After all, larger projects are usually more complex than smaller ones and increasing complexity, in turn, can increase the likelihood of planning errors. "The metro extension in the Rotterdam Region in the Netherlands [faced limited data dispersion] due to a strict focus on reducing complexity", Giezen (2012, p. 781) writes. Indeed, it seems plausible that estimating resettlement for larger dam projects may be more complex since it involves larger areas of land to be inundated and thus it is more prone to planning errors than planning for smaller projects. Dams can be categorized into dams with a capacity of less than 100 MW, those with a capacity between 100 MW and 1000 MW as well as mega-dams with a capacity of more than 1000 MW (Zarfl et al., 2015; Kirchherr, Charles, & Walton, 2017). Based upon the literature reviewed, we expect that the former category will feature the smallest data dispersion, while the latter category will feature the largest.

2.2. Project cycle

A second technical explanation suggested by scholars such as Vanclay (2002) and Kirchherr, Pomun, and Walton (2016) relates to changes in data dispersion during the project cycle. Kirchherr et al. (2016) found that people can move into project-affected areas after the project announcement to benefit from compensation payments which would indicate that the number of resettlees can increase throughout the project cycle. Vanclay (2002, p. 194) argues that "projects can also lead to a decline in population size where people move out because the area affected by a project becomes less desirable as a place to live" which would indicate a decrease in the number of resettlees.[2] Other various theories also arose in the development of this paper: Consultations with industry experts that helped to shape these hypotheses suggests that consultants may undertake a rapid estimation regarding resettlement figures initially (induced by strict deadlines set by project developers) and only refine this initial figure, prone to estimation error, later on in the project with this practice then also resulting in data dispersion. A reviewer highlighted the dynamic nature of agreements over compensation as the understanding of impacts (and definition of who is therefore impacted) change, for example, there was a rapid increase in the number of people requiring resettlement in the Lesotho Highlands Water Project after reservoir filling triggered an earthquake (Tilt, Braun, & He, 2009). Meanwhile, Flyvbjerg et al. (2002) introduces the political perspective to the project cycle arguing that project advocates would attempt to downplay resettlement figures in the beginning of the project to minimize opposition (further discussed below). We base our conceptualization of the dam project cycle on Kirchherr and Charles (2016) who distinguish between the planning and design, the construction and the operation phase. We expect differences in resettlement figures between the planning and design and construction phase, based upon our literature review.

2.3. Stakeholders

Scholars from the political school of thought frequently highlight that differing self-interest stakeholders can influence data. For instance, Flyvbjerg et al. (2002, p. 288) writes that "when a project goes forward, it creates work for engineers and construction firms, and many stakeholders make money". Hence, these would be incentivized to downplay resettlement data since larger resettlement figures can result in more controversy around a dam project (Tarrow, 1994; Martinez-Alier, 2002) and even a withdrawal of funding with the (originally World Bank-funded) Arun III Dam in Nepal being a case in point (Rest, 2012). Meanwhile, project opponents would be keen to inflate resettlement figures to attract additional media attention which would strengthen their anti-dam-campaign (Islam and Islam, 2016). We base our different categories of stakeholders mostly on the categorization by Nüsser (2003). We expect that governments, dam developers, consultants and international donors, typically advocates of dam projects, will report lower resettlement figures on a selected project than NGOs and the press (as long as it is not state-controlled[3]) which are project opponents. We also argue that academics belong to the category of project opponents and thus report higher resettlement figures than project advocates on average since these were recently found to view large dams' impacts mostly negatively (Kirchherr et al. 2016).

2.4. Political regime

Relevant studies such as Ansar et al. (2014) have investigated whether the political regime also impacts data dispersion. Research from political science suggests differences in data dispersion for democratic, autocratic and hybrid countries. The checks and balances in democratic countries result in increased accountability and transparency which, in turn, implies that "democratic leaders can never be entirely free from a commitment to truth-telling" (Shapiro, 2003, 200) (echoed inter alios by Landemore, 2014; Hollyer, Peter Rosendorff, & Vreeland, 2015). Meanwhile, those releasing information in autocratic regimes can restrict that it is scrutinized which also results in limited data dispersion (Morlino 2008; Epstein, Bates, Goldstone, & Kristensen, 2006). The mix of deliberation and oppression in hybrid regimes has resulted in scholars arguing that these are "more [data disperse] than either pure democracies or autocracies" (Epstein et al., 2006, p. 555). We define a country's political regime using the proxy of Freedom House (2017) definitions. We adopt those labelled 'free' as a proxy for 'democratic', those labelled as 'partly free' as a proxy for 'hybrid' and those labelled as 'not free' as a proxy for 'autocratic', mirroring Merkel (2010). Freedom House's (2017) classification is based on political rights and civil liberties which is considered to provides an appropriate reflection of the mechanisms that affect data dispersion. Lastly, we note that the political regime view is also the reason why Flyvbjerg et al. (2002, p. 285) conclude that "geography matters" regarding data dispersion; after all, the type of prevailing regime differs significantly from continent to continent. We expect the greatest data dispersion in reported resettlement figures in those continents with the greatest share of hybrid countries; furthermore, we expect least data dispersion in those continents with the greatest share of democratic and autocratic regimes.

2.5. Time of completion

We hypothesised that data dispersion would not be static over time. There are two mechanisms that we particularly highlight. Firstly, communications have improved and, with it, wider access to data has increased the number of stakeholders involved in major projects. The struggle against large dam projects is frequently no longer restricted to the country of those to be displaced by it.

[2] Motivations outlined here for moving in and out are strategic. Nevertheless, changes in resettlement data in such cases (undertaken by consultants) would be updates, not deliberate falsehoods. Hence, we attribute this explanation to the technical perspective instead of the political one.

[3] State-controlled press was coded as 'government' for this study.

Rather, relevant international NGOs such as International Rivers and the World Wildlife Fund (WWF) join forces with project-affected communities around the world and can thus bring anti-dam-campaigns in any country to global scale (Eichert, 2014; Ives, 2017). Hence, we hypothesize, albeit this is not suggested in the relevant literature as far as we are aware, that the increase in actors and communication may contribute to the level of data dispersion, as more project opponents report on resettlement data. Work from social movement studies suggests that an ambiguous (instead of a homogeneous) global attitude increases data dispersion since stakeholders will contribute varying data to the discourse in a contested project (Porta & Diani, 2015; Chowdhury, 2013; McCormick, 2006).

Secondly, the changes in attitudes of major funders to large dam projects may have shifted the reporting of resettlement data. Biswas (2012) suggests that the contentiousness regarding large dam projects peaked in 2000 with the publication of the World Commission on Dams (WCD, 2000), which, in turn, fostered consensus not to pursue large dam projects (McCully, 2001). However, with time this contentiousness regarding large dam projects has decreased (approximately since 2010) with many large dam projects now being constructed again (Zarfl et al., 2015). We thus hypothesise shifts in data dispersion depending on when the dam was completed, which is categorized into three time periods: before 2000, 2000–2010 and after 2010.

We close this section by noting that our different explanatory variables can be divided into structural and proximate variables, a division common in the social sciences (Ide, 2015). This distinction is relevant for our respective data dispersion calculation approach (further discussed below). Structural conditions are relatively stable for a single dam project of interest and we thus examine data dispersion between projects for these, while proximate variables vary significantly within projects and we thus examine data dispersion within projects for these (Schneider and Wagemann, 2006; Ide, 2015). We label *size, political regime* and *time of completion* as structural conditions since usually only a single datum can be found for these explanatory variables per dam project (e.g. only a single MW size per dam project). We label *project cycle* and *stakeholders* as proximate variables since different data can be found for these (different data per part of the project cycle, different values per stakeholder) in a single dam project.

3. Methods

Our data collection process to gather resettlement values comprised two steps. First, we built a long list of dam projects. To test our hypotheses outlined in the previous section, we needed dam projects with multiple resettlement data. It is acknowledged that this creates some bias in the sample, as outlined in Section 4.1. To identify dams with multiple resettlement data, we needed to identify dam projects that much has been written about previously. We included all dams examined by Scudder (2012) in his 50 dam survey, the 12 dams examined in Kirchherr et al. (2016), those discussed in the scholarly literature according to Kirchherr et al. (2016) as well as all dams examined in a forthcoming paper on cost and schedule overruns of large dam projects that one of the authors of this paper is involved in. Our long list based upon this approach includes 121 dam projects (cf. Supplementary Material). Two recognized scholarly authorities reviewed this long list, confirming that our list contains the major dam projects of the past decades. While we do not claim that our long list is exhaustive, we are fairly confident that it is largely representative regarding recent large dam projects, given the experts' inputs.

Using the long list, we searched for resettlement data for each of the dams with the intention to gather every reported resettlement value available online for each of the dams. This included a wide range of types of reports from academic and technical to media and advocacy. With this range it was not possible to identify the methodology for each estimate or trace the source of the information, so each source was used as an individual datum. This method does bias the results, so that they reflect the reporting of resettlement numbers, not which ones are being calculated. We acknowledge that the reliance on online sources may also bias our sample towards more recent dam projects. Key words used in the searches were the name of the dam in combination with the word 'resettle*', alternatively 'displac*','reloc*', 'disloc*'. We used Google, Google News Archive Search, Google Scholar, Thomson Reuters' Web of Science, the University of Oxford Search Oxford Libraries Online (SOLO) and Elsevier's Scopus for our searches. Furthermore, resettlement data sources were provided by another research team also working on dams. A search for resettlement data for a single dam project could take up to four hours. Overall, it took us eight months to gather the data presented in this study. We found most resettlement data (n = 43) for the Three Gorges Dam. In total, we gathered 862 resettlement values and we found at least 10 resettlement values for 29 dam projects. The full list is presented in Section 4. Only dam projects were included in our analyses for which at least 10 resettlement values could be obtained which reflects that both the C_V as well as the QCD require a minimum number of values for calculation (Benesch, 2013; Bonett, 2006). We acknowledge that additional resettlement data may be available in non-online formats and other languages, but such formats and non-English sources were not accessible for the authors of this study.

We also gathered from the identified sources the publication data, start and end data of dam construction, the size of the dam, political regime data regarding the country of the dam using the proxies from Freedom House (2017) and the type of source based on the stakeholder differentiation presented in the previous section since that was also needed data to test our hypotheses. We note that not all information was available for dams on the long list (further discussed below). Regarding the type of source, resettlement values found on dam developers' websites were coded as 'dam developer', those released by NGOs as 'NGOs' etc.; we further drew on a qualitative data set, described in Kirchherr et al. (2017), that includes more than 100 expert interviews with various relevant dam industry stakeholders (e.g. dam developers, consultants, NGOs) for explanations regarding our results presented in the next section.

We note that we only included resettlement values in our database if it was evident that the respective author referred to actual resettlement which we define as a "co-opted or coerced process by which local people surrender land for a project (such as dam) and are relocated elsewhere" (Vanclay, 2002, 195). This also includes people who lose a rented house due to the project, but not those who lose farming land, but not their housing due to a project (Smyth and Vanclay, 2017; Vanclay, 2002). Lastly, we note that resettlement does not include downstream communities that are negatively affected by a dam project, but do not surrender land for it (Smyth and Vanclay, 2017; Vanclay, 2002).

Admittedly, sources identified did not always indicate the number of people resettled because of a dam project. Instead, the number of families displaced or the number of communities displaced were reported for 20% of the resettlement values used in the eventual sample. We converted this data to the number of people resettled in all instances to ensure comparability of our data. The conversion approach taken in many of the instances was straight-forward since authors would frequently indicate the average number of people in a family, for instance, in the source at question. If this was not the case, we would adopt suggested conversion rates from other dam projects (first from those in the same country, if not available, those in neighbouring countries and if these were also not available, those from the same continent).

Analysis employed descriptive statistics. This study was designed to be exploratory, and while it investigates the potential sources of dispersion in the data for these large and controversial dams, it is not considered appropriate to apply parametric statistics to this database at this stage. As highlighted below, the sample size is limited, and there these results are not significant, but rather used to highlight the dispersion in data.

The sample was compared to the remainder of the long list to identify biases using Cramer's V to compare the nominal (continent, stakeholder) and ordinal (time category) categories.

We analyze structural variables (*size, political regime* and *time of completion*) with the coefficient of variation (C_V) and the quartile coefficient of dispersion (QCD). The C_V which is the ratio of the standard deviation to the mean "can be found virtually in all statistics texts" (Bonett 2006, 2953). Due to its reliance on the mean, the C_V is susceptible to outliers. The QCD which is the ratio of interquartile range to the median is viewed as a more robust alternative (Tam, Valera, Tan, & Koh, 2016) since it is unaffected by scores below the first and above the third quartile (Benesch, 2013; Francis, 2008). As the C_V remains the most commonly used descriptive statistic for data dispersion, we report both the C_V and QCD in Section 4 of this paper. For all structural variables, we first calculated the C_V and QCD across all data for a single dam project in our sample. We then calculated the mean and median across these results (further explained in Section 4.2 of this paper).

For proximate variables (*political cycle* and *stakeholders*), we first calculated the median resettlement figure for a respective category of a single dam project (e.g. NGOs) and then divided this by the median of all resettlement values of a specific dam before subtracting one. Second, we calculated the mean across the results from this exercise from all dam projects (also noted underneath the relevant tables in Section 4). We report the standard deviation regarding the mean for the calculation results of both our structural and proximate variables to indicate the amount of dispersion regarding the aggregated results presented, whereas we note that none of the results presented in the next section are statistically significant due to our limited sample size. However, the sample size could not be increased further despite significant efforts undertaken (further discussed below). Evidence presented throughout this paper is thus best considered as suggestive.

We ran all analyses presented in the next section both with only unconverted population data as well as with converted and unconverted data. We found no substantive difference in results in Sections 4.2 and 4.3 at large (cf. appendix). Hence, we deemed our conversion approach to be robust. We also feared that projects not yet completed may be different to projects completed which could, in turn, skew results. We thus also ran all analyses reported in Sections 4.2 and 4.3 only with projects completed (cf. Supplementary Material), but no substantive difference was found either. Results reported in the next section are those based on both unconverted and converted data as well as data from projects completed and not completed; after all, we are keen to base our findings on a database as large as possible to enhance external validity. The entire sample database is provided in the Supplementary materials of this paper to enable readers to run further analyses with it in order to accelerate cumulative knowledge development on this topic.

4. Results

4.1. Sample Overview

Table 1 presents a summary of the sample. A boxplot (including explanation) for every dam project in our sample is included in the Supplementary material. The boxplot for the Three Gorges Dam, the largest project in our sample, is depicted in Fig. 1 to visualize data dispersion in this dam project. Our long list and the sample are largely representative regarding the distribution of large dams at the beginning of this century, as outlined by WCD (2000) (Fig. 2). Only Africa is overrepresented, particularly in our sample (by 25 percentage points in comparison to the WCD sample) which is driven by an overrepresentation of academic sources in our sample (38% of our figures are from academic sources) with scholars reported to over-study African dam projects (Kirchherr et al., 2016). Meanwhile, the underrepresentation of Europe in the long list and our sample is driven by our reliance on online sources in combination with the fact that no large dams have been built in Europe recently (WCD, 2000; Biswas, 2012).

The overrepresentation of academic sources in our sample may also explain why our sample is skewed to extremely large dams since these are studied most by scholars (Kirchherr et al., 2016). The median size of a dam in our sample is 1148 MW, whereas the median size of a dam on the long list (excluding the sample) is only 250 MW (further information on this depicted in Table 2); megawatts, a measure of hydropower capacity, are used here as a proxy for size. We further note that the increased attention extremely large dams receive from scholars and also from stakeholders such as the press (Biswas, 2012) increases the number of records available per dam which also explains why the (in our view) most notorious large dam projects such as Three Gorges Dam (Wilmsen and van Hulten, 2017), Sardar Sarovar Dam (Khagram, 2004) and Nam Theun 2 (Shoemaker, Baird, and Manorom, 2014) are all included in our sample.

More than 75% of the dam projects in our sample (67% of long list projects) were built in either hybrid regimes or autocratic ones (Table 3). This observation resonates with recent social movement literature (e.g. Xie and van der Heijden, 2010; Evren, 2015; Kirchherr et al., 2016) which identifies anti-dam-movements as a major barrier to dam construction, while noting that countries with limited political opportunity structures (POS) which are both hybrid and autocratic ones, in turn, prevent the emergence of anti-dam-movements.

Twenty records on average were found for the dam projects in our sample in democratic regimes, 18 for those in hybrid and 23 for those in autocratic regimes. On average, 21 records were found per dam project. Lastly, we note that the number of records available did not change much over time despite expansion of the internet and data access during the period: 19 records were collected on average for projects completed before 2000, 23 for projects completed between 2000 and 2010 and 20 for projects completed after 2010.

There was no significant difference ($p > 0.05$) between the sample and the remainder of the long list in terms of end date of construction or continent. There was a significant difference ($p < 0.05$) in the sources of information available, with a greater representation of the media and NGOs in our sample than in the remainder of the long list, potentially suggesting that more contested dams are over-represented in the sample; there was similar representation by academics, governments and developers. We now turn to the presentation of our findings regarding the initially outlined hypotheses.

4.2. Technical perspective

4.2.1. Size

The hypothesis regarding size cannot be confirmed with the data at question. Indeed, the largest dam projects face smaller data dispersion than the smaller ones (Table 4).[4] Here we expand on the

[4] A category of projects with a size of <100 MW is not included since none of the projects in our sample featured <100 MW.

Table 1
Overview of sample of 29 dam projects.

Dam	Country	Completion date	Height (m)	# data	Median resettlement	C_V	QCD
Itaparica Dam	Brazil	1988	105	20	40,000	0.14	0.13
Irapé Dam	Brazil	2006	208	12	5,000	0.14	0.01
China Three Gorges Dam	China	2009	181	43	1,300,000	0.55	0.17
Xiaolangdi I & II Dam	China	2000	154	40	181,250	0.36	0.12
Jinghong Dam	China	2013	108	10	2,849	0.46	1.02
Lingjintan Dam	China	2001	38.5	11	4,060	0.21	0.05
Shuikou Dam	China	1996	101	16	67,100	0.07	0.02
El Quimbo Dam	Colombia	2011	151	19	1,912	1.35	0.33
Gilgel Gibe III	Ethiopia	2015	243	10	1,220	1.07	2.55
Bui Dam	Ghana	2013	108	32	1,216	0.29	0.00
Chixoy Dam	Guatemala	1983	90	21	3,500	0.83	5.26
Hirakud Dam	India	1957	61	20	100,067	0.24	0.26
Sardar Sarovar Dam	India	2006	163	30	200,000	0.42	0.57
Saguling Dam	Indonesia	1987	99	18	11,185	0.92	2.04
Kedung Ombo Dam	Indonesia	1991	60	15	24,498	0.16	0.15
Kossou Dam	Ivory Coast	1972	58	22	75,000	0.05	0.00
Nam Theun 2	Laos	2010	39	25	6,200	0.06	0.03
Xayaburi Dam	Laos	2019	33	21	2,100	0.17	0.04
Bakun Dam	Malaysia	2011	205	26	10,000	0.10	0.00
Manantali Dam	Mali	1988	65	19	10,000	0.10	0.10
Cahora Bassa Dam	Mozambique	1974	171	10	42,000	0.25	0.40
Myitsone Dam	Myanmar	2018	140	29	10,000	0.85	0.99
Kainji Dam	Nigeria	1968	65	20	44,000	0.08	0.14
Ghazi Barotha Dam	Pakistan	2004	69	16	899	0.18	0.13
San Roque Dam	Philippines	2004	200	10	4,160	0.22	0.19
Upper Kotmale Dam	Sri Lanka	2013	35	11	2,835	1.18	0.19
Bujagali Dam	Uganda	2012	30	15	820	1.98	0.49
Son La Dam	Vietnam	2012	138	29	92,000	0.08	0.10
Kariba Dam	Zambia	1959	128	28	57,000	0.13	0.00

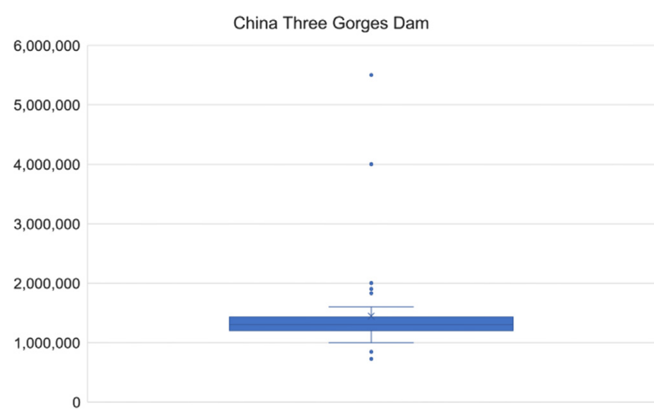

Fig. 1. Collected resettlement figures regarding the Three Gorges Dam as a boxplot. The QCD statistic uses the interquartile range, divided by the median, therefore ensuring that outliers don't influence the result.

interpretation of the results in Table 4 to aid readers unfamiliar with these methods. For the smaller dams (≤1000 MW), the median of the coefficient of dispersion (QCD) for the 13 dams is 0.24, which indicates that the interquartile range is 0.24 of the median for this group. This is considered the most robust statistic for comparing groups. The median of QCD of projects with a size of larger dams (>1000 MW) is lower, indicating less dispersion around the mean. Hence, there was less variation in the data found for the >1000 MW category. The lower standard deviation for projects with >1000 MW indicates that the scores in this category were more similar to each other for this category. The coefficient of variation (C_V) represents the standard deviation divided by the mean which is heavily influenced by outliers.

We tested the robustness of these findings by varying the categories showcased in Table 4 (cf. Supplementary Material). Yet none of these additional analyses indicated the expected relationship between data dispersion and project size, although they did highlight that the two dams with the greatest capacity also had the greatest dispersion. Experts we consulted attributed our result in Table 4 to noise arguing that the difficulty of estimating the needed resettlement and thus the probability of planning error would not depend on the project's size, but rather on a variety of topographical conditions which vary significantly from project to project.

We believe the difference in dispersion is related to the different political regimes (further outlined in 4.3): 46% of dams with a capacity of ≤1000 MW are in hybrid regimes, compared to only 20% of dams with a capacity of >1000 MW, while we find the greatest dispersion for dams in hybrid regimes.

4.2.2. Project cycle

Our data provides evidence for the hypothesis, posited both by the technical and political perspective, that resettlement values during the planning and design phase tend to be lower than those later on in the project cycle. Indeed, resettlement values from the planning and design phase are 28.3% lower on average than the overall median resettlement value per dam project, according to our analysis (Table 5). For instance, the Chinese government estimated during the planning and design phase that 725,500 people would need to be resettled for the Three Gorges Dam, as reported by Heming and Rees (2000), an estimate that was corrected to 1,500,000 in its operation phase (Watts, 2010). Our data does not indicate if this is to be explained by people moving into the resettlement area upon project announcement to benefit from compensation, by estimation error due to rapid assessment, by project proponents consciously downplaying resettlement figures to push the project through, or by changes in design or impacts over time affecting total resettlement figures. The expert interview data consulted suggested all of these explanations. There is no suggestion that estimates post-completion, in the operation phase, are more accurate; estimates in this phase are dominated by academics (47%) and NGOs (21%) who generally reported higher figures (see Section 4.3), whereas government and developers, who report lower figures, were better represented in the construction phase.

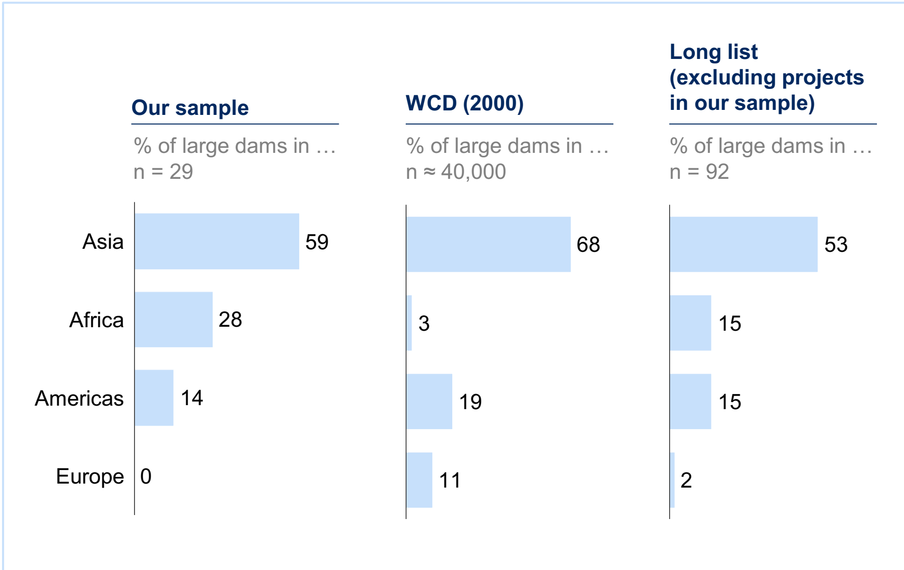

Fig. 2. Geographical distributions of dams in our sample, the World Commission on Dams report, and our long list.

Table 2
Hydropower capacity distribution from sample and long list (excluding sample) perspective.

	Sample (n = 29)	Longlist (excluding sample) (n = 83)
<100 MW	0%	22%
100–1000 MW	46%	60%
>1000 MW	54%	18%

Capacity of the following dams could not be retrieved: Kedung Ombo Dam, Alborz Dam, Kaeng Suea Ten Dam, Metolong Dam, Ramial Dam, Mohale Dam, Alemán Dam, Pimburetewa Dam.

4.3. Political perspective

4.3.1. Stakeholders

Our data largely provides evidence for the hypothesis that differing self-interests of stakeholders influences resettlement values, whereas, admittedly, standard deviations particularly for project opponents are quite high. For instance, one academic source, Nakayama and Fujikura (2003), reports a resettlement figure of 3,000 families for the Saguling Dam, while another, Jain and Singh (2003), reports 65,000 people. The methodologies for collecting these data are unclear, as is almost unanimously the case, so we rely on the reported figures. The government resettlement values are, on average, the most below the median resettlement value per dam project (−22.2%), while NGO resettlement values on average exceed the median resettlement value most (28.4%). For instance, the Indian government estimated in 2014 that 154,227 people would be resettled because of the Sardar Sarovar Dam (NCA, 2014), while NGOs put this figure at 250,000 in the same year (Patterson, 2014).

The only result in Table 6 that defies our hypothesis regarding stakeholders is the one regarding consultants. While we had initially assumed that these would report values below the median to cater to their clients, the project advocates, we found them to report values 9.5% above the median on average. In consultation on development of this paper, experts have suggested that consultants may fear being sued by project opponents if it becomes evident upon resettlement completion that resettlement figures were downplayed initially in social impact assessments (SIAs) authored by the respective consultants. However, our findings regarding

Table 4
Project size: Greater data dispersion associated with smaller dams.

Project size (MW)	n		C_V	QCD
≤1000	13	Mean	0.70	0.58
		Median	0.19	0.24
		Standard deviation	1.47	0.61
>1000	15	Mean	0.42	0.33
		Median	0.13	0.18
		Standard deviation	0.68	0.30

Size of one of the dams in the sample (Kedung Ombo Dam) could not be retrieved.

Table 5
Project cycle: Released resettlement data is lowest in the planning and design phase.

	Planning and design	Construction	Operation
n	12	23	27
Mean	−28.3	4.1	−3.7
Standard deviation	33.80	27.00	19.77

Reported mean is relative to the median resettlement value per cycle; standard deviation is standard deviation around the reported mean. Two cases were not completed by 2017 and did not have any data post completion; a full sensitivity analysis without these dams is included in the Supplementary Material.

Table 3
Regime types represented in sample and long list, by continent. Columns (continents) sum to 100%.

	Asia (S)	Asia (L)	Africa (S)	Africa (L)	Americas (S)	Americas (L)	Europe (L)	Total (S)	Total (L)
n	10	15	8	10	3	9	2	21	36
Democratic	20%	13%	25%	30%	33%	78%	0%	23%	33%
Hybrid	30%	47%	50%	40%	67%	22%	50%	35%	39%
Autocratic	50%	40%	25%	30%	0%	0%	50%	42%	28%

(S) refers to dams in the sample, (L) refers to dams which are on the longlist, but not in the sample.

Table 6
Stakeholders: Resettlement data released by project advocates was usually lower than those of project opponents.

	Project advocates				Project opponents		
	Govt.	Developers	Donors	Consultants	Academia	Press	NGOs
n	13	16	21	7	29	21	27
Mean	−22.2	−21.5	−1.1	9.5	12.7	23.5	28.4
Standard deviation	40.33	34.38	30.01	30.22	101.82	115.93	101.41

Reported mean is relative to the median resettlement value per stakeholder; standard deviation is standard deviation around the reported mean.

consultants may also be due to coincidence – we only have data from consultants for 7 of our 29 projects studied.

4.3.2. Political regime

Furthermore, our data provides evidence for the hypothesis that the type of political regime influences dispersion in reported data. Indeed, the relevant data dispersion statistics are lower for democratic and autocratic regimes than for hybrid ones (Table 7). We suggest that differences in C_V and QCD from a continent perspective can also be explained by the prevailing political regime: The relevant descriptive statistics are highest in Latin America, the continent with the greatest share of hybrid regimes in our sample (67%), and lowest in Asia, the continent with the lowest share of hybrid regimes in our sample (30%) (Table 3; Table 8). For instance, identified resettlement figures for Guatemala's Chixoy Dam range from 450 (Johnston, 2004) to 22,000 (AP, 2014), while those for Laos' Nam Theun 2 Dam only range from 4,500 (Guttal, 2000) to 6,738 (Souksavath and Nakayama, 2013).

We also examined median resettlement values from both a stakeholder and regime perspective (Table 9). We acknowledge that there can be few observations per category (e.g. 'Hybrid' in combination with 'Govt.'). Yet it is necessary to break these down as the role of different stakeholders changes between regimes. For instance, we find that press median resettlement values are above median resettlement values on average in both democratic and hybrid regimes (by 14.5% and 72.7%), but not in autocratic ones (−7.2% below the median resettlement value). This can be explained by the fact that autocratic regimes which act as project advocates control the media (even if it is not officially state-controlled) and thus resettlement data published in the media in these regimes.

Table 7
Political regimes: Highest reported data dispersion found in regimes, limited dispersion found both in democratic and autocratic regimes.

Type of regime	n		C_V	QCD
Democratic	7	Mean	0.22	0.18
		Median	0.22	0.13
		Standard deviation	0.11	0.20
Hybrid	10	Mean	0.70	0.90
		Median	0.54	0.26
		Standard deviation	0.66	1.64
Autocratic	12	Mean	0.34	0.44
		Median	0.20	0.11
		Standard deviation	0.33	0.76

Table 8
Geographical distributions: Highest data dispersion in Americas, lowest in Asia.

	n		C_V	QCD
Asia	17	Mean	0.37	0.36
		Median	0.22	0.15
		Standard deviation	0.33	0.53
Africa	8	Mean	0.49	0.46
		Median	0.19	0.12
		Standard deviation	0.69	0.87
Americas	4	Mean	0.62	1.43
		Median	0.49	0.23
		Standard deviation	0.59	2.56

4.3.3. Time of completion

The data demonstrates that dispersion of reported data is highest in projects completed post 2010 with all relevant measures in Table 10 showcasing a dispersion increase for this period (e.g.

Table 10
Time period of completion: Greatest data dispersion (median QCD) is evident for the most recent period.

Project completion	n		C_V	QCD
Before 2000	11	Mean	0.27	0.77
		Median	0.14	0.14
		Standard deviation	0.31	1.60
2000–2010	8	Mean	0.27	0.16
		Median	0.22	0.13
		Standard deviation	0.16	0.18
After 2010	10	Mean	0.75	0.57
		Median	0.66	0.26
		Standard deviation	0.64	0.79

Table 9
Political regimes from a stakeholder perspective: Press reporting lower resettlement values in autocratic regimes than in democratic and hybrid ones. Mean and standard deviation relative to median resettlement value per dam.

		Project advocates				Project opponents		
		Govt.	Developers	Consultants	Donors	NGOs	Press	Academia
Democratic	n	3	4	1	5	7	4	7
	Mean	2.8	−14.5	0.1	−14.0	6.9	14.5	−1.3
	Standard deviation	6.19	15.95	N/A	29.70	16.27	29.29	7.91
Hybrid	n	2	6	1	7	9	7	10
	Mean	−61.1	−21.8	−16.0	−6.1	59.9	72.7	48.8
	Standard deviation	36.77	25.53	N/A	22.41	168.68	192.66	170.14
Autocratic	n	8	6	5	9	11	10	12
	Mean	−21.8	−25.9	16.5	9.8	16.2	−7.2	−9.1
	Standard deviation	43.40	51.77	33.52	34.25	49.39	43.20	29.03

Reported mean is relative to the median resettlement value per stakeholder; standard deviation is standard deviation around the reported mean.

almost a doubling of the median QCD for the completion period 2000 to 2010 to the period after 2010; the median QCD is likely the most robust descriptive statistic in Table 9). For instance, the Myitsone Dam whose completion is still pending features resettlement data ranging from 3,000 people (Min and Win, 2016) to 40,000 (Phanida, 2016). This data may be indicative of a shift in reporting of resettlement data on large dams in recent years. The data remains inconclusive regarding the level of data dispersion for projects completed before 2000. The comparison of mean QCDs for the periods 'Before 2000' and '2000–2010' indicates that data dispersion declined. Overall, the data from Table 10 suggests that data dispersion is a recent instead of a recurring issue.

5. Discussion

Accurate resettlement data is essential for at least three reasons. First, it is essential for those deciding whether to pursue or not to pursue a project. Pickrell (1989) studied cost and ridership estimates for large urban rail transit projects and found that both tend to be inflated by project advocates. Accordingly, Pickrell (1989, p. 8) argues that "it is certainly possible that decision-makers acting on more accurate forecasts of costs and future ridership for the projects reviewed here would have selected [other] projects" if accurate estimates had been presented to them.[5] Similarly, those deciding on large dam projects may also choose alternative energy sources if confronted upfront with the actual resettlement values. After all, resettlement is a major cost item in large dam projects with greater resettlement thus potentially threatening a project's economic viability. For instance, one-third of total costs of the Three Gorges Dam were spent on resettlement (Wilmsen and van Hulten, 2017).

Second, accurate resettlement data is essential for project-affected communities. Vanclay (2017, p. 4) notes that "no matter how effectively [...] resettlement is undertaken, there is always some amount of harm and hurt" (echoed by Scudder, 2005; Tilt et al., 2009). Yet this harm and hurt can be amplified if initial resettlement values are reported too low, as indicated by our data. After all, these initial values are the basis for the allocation of resettlement funding, found to be key to avoid impoverishment as a result of resettlement (Fujikura & Nakayama, 2013). Yet this funding is unlikely to be increased throughout the project once it emerges that eventual resettlement is larger than the initial estimates. The slashing of social safeguard budgets to minimize cost overruns – which are estimated at 99% for large dams projects on average (Ansar et al., 2014) – is said to be a common practice in the dam industry (Haas & Skinner, 2015; Kirchherr, Matthews, Charles, & Walton, 2017).

Third, accurate resettlement data is essential for scholars working on large dam projects. For instance, consider that Kirchherr et al. (2016) recently examined the size of resettlement as a causal condition that may induce significant anti-dam-protests. The scholars do not indicate any dispersion regarding their resettlement data, although we found significant dispersion for multiple dam projects in their sample (e.g. Three Gorges Dam, Myitsone Dam, Nam Theun 2 Dam). Kirchherr et al. (2016) conclude that the size of resettlement is not part of either of the two causal recipes identified, through a fuzzy set Qualitative Comparative Analysis (fsQCA), that induce significant anti-dam-protests. However, a replication of their analyses using the minimum and maximum resettlement data identified in this study indicates that the results of their analyses can change, depending on the data used. If minimum resettlement data from our study is used for their analyses, the absence of resettlement is found to contribute to massive anti-dam protests (see Supplementary Material), whereas the results remain unchanged compared to their original analysis if maximum resettlement data is used. Hence, scholars who consider resettlement data in their work are urged to carefully examine the resettlement data used. Indeed, significant dispersion may not only be found for resettlement data for large dam projects since determinants of dispersion found in this study (e.g. differing stakeholder interests, political regime) hold across various infrastructure projects.

These three reasons explain why accurate resettlement data is urgently needed. While resettlement "of a large mass of people [...] is a formidably intricate [...] process" (Cernea, 2004, 10), a fixed number of people is eventually resettled (assuming that a dam project induces resettlement). We have collected 21 resettlement values on average per dam project in our sample; we have even collected 43 resettlement values for the Three Gorges Dam. However, we do not know which of these values are accurate. Rather, our study only showcases the range of reported resettlement values with the intention to highlight the contested nature of truth and thus knowledge in the global dam industry. It indicates that most published resettlement values are false, particularly for large, contested dams such as were represented in this sample. Caution is thus warranted regarding any resettlement figure published.

We have presented the technical and political perspective regarding data dispersion in Section 2 of this paper. The political perspective argues that data dispersion is consciously introduced. Overall, our findings confirm this perspective with resettlement values seemingly determined by the project cycle (which can be viewed as a hypothesis of the political perspective), stakeholders, the political regime and the time of completion. Most values found in our dataset may thus be framed as resettlement lies with a lie defined as "a false statement made with the intent to deceive others" (Carson, 2006, 286). Overall, our study suggest that resettlement lies are omnipresent in the global dam industry.

A data transparency initiative is needed to overcome these resettlement lies. Those releasing resettlement figures must exactly explain how they have developed these. Sadly, the diverging self-interests of the different stakeholders outlined in this study make it unlikely that this recommendation will be implemented anytime soon. Even if it was implemented, only few experts would be able to judge if the calculation approach was sound. Furthermore: If these experts are paid to certify selected resettlement figures, the problem arises that these may cater to the needs of those paying them. Indeed, ensuring accurate resettlement figures is likely to be an intricate process, particularly in countries with limited checks and balances, as outlined in Section 4.3.

Nevertheless, we are hopeful that this study may also serve as the starting point towards more accurate resettlement figures. We envisage that it may help to develop a tool to de-bias resettlement figures. The starting point for this tool would be true resettlement values that could then be compared to the ones already identified. While Kirchherr et al. (2016) argue that many academics do not act as independent voices regarding large dams, but rather as activists, we still suggest to task academics with obtaining resettlement values. First, this suggestion reflects our normative belief that "scientists are armed with the truth" (Perkins, 2017). Second, it reflects that academics remain the ones of all stakeholders involved in a dam project with the least stake in it (Nüsser, 2003; Nüsser and Baghel, 2017) and thus those with the highest probability to generate true resettlement figures. Hence, we propose a research project that starts out by attempting to identify true resettlement values across many dam projects with

[5] Scholars have also found that water demand estimates created as part of water infrastructure projects are frequently higher than eventual demand with these estimates influencing whether a project is pursued or not (Pohlner, 2016; Walker, 2013; van Niekerk and Du Plessis, 2013).

the intention to ultimately develop a tool to de-bias resettlement values. These would be de-biased based upon the planning cycle, the stakeholders releasing them, the relevant political regime and point in time as well as additional variables that determine resettlement values that are possibly identified throughout the research project.

6. Conclusion

The idea for this paper developed when one of the authors of it worked on a (forthcoming) study that investigates cost and schedule overruns for large dam projects. While researching cost and schedule data, the author realized not only that these records varied widely for a specific dam project, but that resettlement data also seemed to be extremely dispersed. A subsequent literature review suggested that resettlement data dispersion was apparently not yet acknowledged in the scholarly discourse. Rather, academics took specific resettlement values for granted in the relevant studies, e.g. Kirchherr et al. (2016), although even academic studies published in the same year could feature very different resettlement values for the same dam project.

This study presents the very first analysis of reported resettlement data dispersion regarding large dam projects, as far as we are aware. Our sample was limited to dams with at least 10 reports of resettlement figures, which biased it toward the more contested and more studied projects, and drew on reported data uncritically to represent the different narratives rather than the (opaque) methodologies. We find suggestive evidence for the political perspective within the project management literature which claims that data dispersion is consciously introduced. Indeed, we find that resettlement data from the planning and design phase of a project are 28.3% lower on average than its median resettlement values which can be explained by project proponents downplaying resettlement figures in the early stages of the project to minimize resistance. Furthermore, we find that the resettlement figures reported by project advocates (government, dam developers, international donors) are on average lower than the median resettlement figure per dam project, while those of project opponents (NGOs, press, academia) are on average higher. Dispersion in reported data is least in projects pursued in countries with democratic and autocratic regimes which may indicate the presence of a functioning system of checks and balances and hence accountability and transparency and additionally, or perhaps conversely, a tight grip on institutions and society found in autocracies limit data dispersion. Furthermore, we find an uptick in dispersion in reported data since 2010.

Overall, the dispersion in reported data found in our dataset highlights the contested nature of truth and knowledge in the dam industry. We suspect that our dataset does not contain falsehoods, but lies, given that our findings correspond mostly with the predictions from the political perspective within the project management literature. "Accurate" figures may be difficult to measure but an agreed, transparent methodology to determine resettlement figures and other negative impacts of large dam construction, which is reviewed following construction, would help generate trust between stakeholders. Further, considering the low resettlement figures in the planning and design phase, ensuring adequate resources are available for additional resettlement is essential. Resettlement lies particularly deserve the attention of practitioners given that inaccurate resettlement figures can result in decision-makers choosing projects that are ultimately not economically viable as well as insufficiently funded resettlement programs which can lead to impoverishment of displaced communities. Indeed, resettlement lies can offset dams' vast positive impacts.

Admittedly, the major limitation of this study is its limited sample size. We spent eight months attempting to collect all resettlement data on large dam projects that is currently available online. We identified a total of 862 resettlement records and we found at least 10 resettlement records for 29 dam projects. While this database constitutes the very first medium-N-resettlement database on large dam projects that has ever been created, as far as we are aware, much more resettlement data needs to be identified (it is assumed this data exists in other languages and other formats than were identified by our methods). After all, our sample size was too limited to yield any results that are statistically significant. Our entire database is presented in the Supplementary materials to help accelerate cumulative knowledge development on this timely topic.

Future research on this topic may attempt to expand the database we have created for this study with the end goal being the creation of a tool to de-bias resettlement data for large dam projects, as described in Section 5 of this paper. Specifically this may result in creation of context appropriate weighting-factors for resettlement figures for large dams that donors and planners can apply to better budget for resettlement costs. Furthermore, scholars working on energy/infrastructure projects beyond dams are urged to replicate this study for their respective focus projects, as also outlined in Section 5. After all, the influences of data dispersion identified in this study (e.g. self-interest of stakeholders, political regime) are not specific to dams, but will likely hold across a variety of different energy/infrastructure projects.

Conflict of interest statement

No conflict of interest is reported by the authors.

Acknowledgements

We thank Frank Vanclay and Bryan Tilt for content and framing advice regarding this piece of work. Furthermore, we thank Bernd Schlipphak for methodological advice regarding this article. In addition, we thank Michael Ioannides for sharing selected resettlement data points with us. Lastly, we thank the editors and anonymous reviewers for their contributions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.worlddev.2018.10.003.

References

Ansar, A., Flyvbjerg, B., Budzier, A., & Lunn, D. (2014). Should we build more large dams? The actual costs of hydropower megaproject development. *Energy Policy, 69*, 43–56. https://doi.org/10.1016/j.enpol.2013.10.069.

AP. (2014). *Guatemala's President Apologizes to 33 Communities Displaced for Dam, Outlines Compensation*. Fox News. http://www.foxnews.com/world/2014/11/08/guatemala-president-apologizes-to-33-communities-displaced-for-dam-outlines.html.

Awojobi, O., & Jenkins, G. P. (2015). Were the hydro dams financed by the World Bank from 1976 to 2005 worthwhile? *Energy Policy, 86*, 222–232. https://doi.org/10.1016/j.enpol.2015.06.040.

Bacon, R. W., & Besant-Jones, J. E. (1998). Estimating construction costs and schedules: Experience with power generation projects in developing countries. *Energy Policy, 26*(4), 317–333.

Baird, I. G., Shoemaker, B. P., & Manorom, K. (2015). The people and their river, the World Bank and its dam: Revisiting the Xe Bang Fai River in Laos. *Development and Change, 46*(5), 1080–1105. https://doi.org/10.1111/dech.12186.

Benesch, T. (2013). *Schlüsselkonzepte Zur Statistik: Die Wichtigsten Methoden, Verteilungen, Tests Anschaulich Erklärt*. Berlin, Germany: Springer Spektrum.

Bhatia, R., Cestti, R., Scatasta, M., & Malik, R. P. S. (2008). *Indirect economic impacts of dams – Case studies from India, Egypt and Brazil*. The World Bank.

Biswas, A. K. (2012). Impacts of large dams: Issues, opportunities and constraints. *Impacts of Large Dams: A Global Assessment*. Berlin, Germany: Springer.

Bonett, D. G. (2006). Confidence interval for a coefficient of quartile variation. *Computational Statistics & Data Analysis, 50*(11), 2953–2957. https://doi.org/10.1016/j.csda.2005.05.007.

Bredillet, C. N. (2010). Blowing hot and cold on project management. *Project Management Journal, 41*(3), 4–20. https://doi.org/10.1002/pmj.20179. Wiley Subscription Services, Inc., A Wiley Company.

Cantarelli, C.C., Bent, F., Molin, E.J.E., van Wee, B. (2013). Cost Overruns in Large-Scale Transportation Infrastructure Projects: Explanations and Their Theoretical Embeddedness, July. http://arxiv.org/abs/1307.2176.

Carson, T. L. (2006). The definition of lying. *Nous, 40*(2), 284–306. https://doi.org/10.1111/j.0029-4624.2006.00610.x. Blackwell Publishing Ltd/Inc..

Cernea, M. (2004). Social impacts and social risks in hydropower programs: Preemptive planning and counter-risk measures. http://www.un.org/esa/sustdev/sdissues/energy/op/hydro_cernea_social impacts_backgroundpaper.pdf.

Chan, D. S. W. (2017). Asymmetric bargaining between Myanmar and China in the Myitsone dam controversy: Social opposition akin to David's stone against goliath. *The Pacific Review*, 1–18. https://doi.org/10.1080/09512748.2017.1293714. Routledge.

Chapman, J. (2007). India's Narmada dams controversy. *Journal of International Communication, 13*(1), 71–85. https://doi.org/10.1080/13216597.2007. 9674708. Taylor & Francis.

Chowdhury, A. R. (2013). 'Repertoires of contention' in movements against hydropower projects in India. *Social Movement Studies, 13*(3), 399–405. https://doi.org/10.1080/14742837.2013.830564. Routledge.

Cole, S. R., & Frangakis, C. E. (2009). The consistency statement in causal inference: A definition or an assumption? *Epidemiology, 20*(1), 3–5.

Eichert, S. (2014). NGOs as strategic actors in the promotion of sustainable development. *Evolution of dam policies: Evidence from the big hydropower states*. Berlin, Germany: Springer.

Epstein, D. L., Robert, B., Jack, G., Ida, K., & O'Halloran, S. (2006). Democratic transitions. *American Journal of Political Science, 50*(3), 551–569. https://doi.org/10.1111/j.1540-5907.2006.00201.x. Blackwell Publishing Inc.

Evren, E. (2015). The rise and decline of an anti-dam campaign: Yusufeli dam project and the temporal politics of development. *Water History, 6*(4), 405–419. https://doi.org/10.1007/s12685-014-0120-8.

Flyvbjerg, B., Holm, M. S., & Buhl, S. (2002). Underestimating costs in public works projects: *Error or Lie?. Journal of the American Planning Association, 68279–95*(3). https://doi.org/10.1080/01944360208976273. Taylor & Francis Group.

Flyvbjerg, B., Skamris, M. K., Holm & Buhl, S. (2004). What causes cost overrun in transport infrastructure projects? *Transport Reviews, 24*(1), 3–18. https://doi.org/10.1080/0144164032000080494a. Taylor and Francis Ltd.

Folch, C. (2013). Surveillance and state violence in stroessner's Paraguay: Itaipú Hydroelectric Dam, Archive of Terror. *American Anthropologist*. https://doi.org/10.1111/j.1548-1433.2012.01534.x.

Francis, A. (2008). *Business mathematics and statistics*. London, United Kingdom: Cengage Learning.

Freedom House (2017). About 'Freedom in the World.' https://freedomhouse.org/report-types/freedom-world.

Fujikura, R., & Nakayama, M. (2013). The long-term impacts of resettlement programmes resulting from dam construction projects in Indonesia, Japan, Laos, Sri Lanka and Turkey: A comparison of land-for-land and cash compensation schemes. *International Journal of Water Resources Development, 29*(1), 4–13. https://doi.org/10.1080/07900627.2012.741032. Taylor & Francis Group.

Furlong, P., & Marsh, D. (2010). A skin not a sweater: Ontology and epistemology in political science. *Theory and Methods in Political Science*. Basingstoke, United Kingdom: Palgrave Macmillan.

Giezen, M. (2012). Keeping it simple? A case study into the advantages and disadvantages of reducing complexity in mega project planning. *International Journal of Project Management, 30*(7), 781–790. https://doi.org/10.1016/j.ijproman.2012.01.010.

Green, N., Sovacool, B. K., & Hancock, K. (2015). Grand designs: Assessing the African energy security implications of the grand Inga dam. *African Studies Review, 58*(1), 133–158.

Greener, I. (2011). *Designing social research: A guide for the bewildered*. London, United Kingdom: SAGE Publications.

Guttal, S. 2000. Public consultation and participation in the Nam Theun 2 hydroelectric project in the Lao PDR. https://focusweb.org/publications/researchpapers/Research Papers/public consultation.html.

Haas, L.J. & Skinner, J. (2015). Managing the environmental and social risks of hydropower: Private and public roles. http://pubs.iied.org/17292IIED/.

Heming, L., & Rees, P. (2000). Population displacement in the three gorges reservoir area of the Yangtze River, Central China: Relocation policies and migrant views. *International Journal of Population Geography, 6*(6), 439–462. https://doi.org/10.1002/1099-1220(200011/12)6:6<439::AID-IJPG198>3.0.CO;2-L. John Wiley & Sons Ltd..

Hensengerth, O. (2015). Where is the power? Transnational networks, authority and the dispute over the Xayaburi Dam on the lower Mekong Mainstream. *Water International, 40*(5–6), 911–928. https://doi.org/10.1080/02508060.2015.1088334. Routledge.

Hollyer, J. R., Peter Rosendorff, B., & Vreeland, J. R. (2015). *Democracy and transparency*. New York, USA: Cambridge University Press. 10.1017/S0022381611000880.

Ide, T. (2015). Why do conflicts over scarce renewable resources turn violent? A qualitative comparative analysis. *Global Environmental Change, 33*, 61–70. https://doi.org/10.1016/j.gloenvcha.2015.04.008.

IEA (2016). About hydropower. https://www.iea.org/topics/renewables/subtopics/hydropower/.

Islam, M. S., & Islam, M. N. (2016). 'Environmentalism of the poor': The Tipaimukh Dam, ecological disasters and environmental resistance beyond borders. *Bandung: Journal of the Global South, 3*(1), 27. https://doi.org/10.1186/s40728-016-0030-5. Springer International Publishing.

Ives, M. (2017). *A Chinese-backed dam project leaves Myanmar in a bind*. New York Times https://www.nytimes.com/2017/03/31/world/asia/myanmar-china-myitsone-dam-project.html?_r=0.

Jain, S. K., & Singh, V. P. (2003). *Water resources systems planning and management*. Amsterdam, the Netherlands: Elsevier.

Johnston, B. (2004). 'Reparations for Dam-Displaced Communities?' Report from the Chixoy Dam Legacy Issues Meeting, July 25, 2003 Pacux, Guatemala. *Capitalism Nature Socialism, 15*(1), 113–119. https://doi.org/10.1080/1045575032000189046. Taylor & Francis Group.

Jugdev, K. (2008). Good theory: Developing a foundation for project management. *International Journal of Product Development, 6*(2), 177–189.

Khagram, S. (2000). Toward democratic governance for sustainable development: Transnational civil society organizing around big dams. *Third force : The rise of transnational civil society*. Carnegie Endowment for International Peace.

Khagram, S. (2004). *Dams and development: Transnational struggles for water and power*. Ithaca, United States: Cornell University Press.

Kiik, L. (2016). Nationalism and anti-ethno-politics: Why 'Chinese Development' failed at Myanmar's Myitsone Dam. *Eurasian Geography and Economics, 57*(3), 374–402. https://doi.org/10.1080/15387216.2016.1198265. Routledge.

Kirchherr, J., & Charles, K. J. (2016). The social impacts of dams: A new framework for scholarly analysis. *Environmental Impact Assessment Review*. https://doi.org/10.1016/j.eiar.2016.02.005.

Kirchherr, J., Charles, K. J., & Walton, M. J. (2016). Multi-causal pathways of public opposition to dam projects in Asia: A fuzzy set qualitative comparative analysis (fsQCA). *Global Environmental Change, 41*, 33–45. https://doi.org/10.1016/j.gloenvcha.2016.08.001.

Kirchherr, J., Charles, K. J., & Walton, M. J. (2017). The interplay of activists and dam developers: The Case of Myanmar's mega-dams. *International Journal of Water Resources Development, 33*(1), 111–131. https://doi.org/10.1080/07900627.2016.1179176.

Kirchherr, J., Matthews, N., Charles, K. J., & Walton, M. J. (2017). 'Learning It the Hard Way': Social Safeguards Norms in Chinese-Led Dam Projects in Myanmar, Laos and Cambodia. *Energy Policy, 102*, 529–539. https://doi.org/10.1016/j.enpol.2016.12.058.

Kirchherr, J., Pohlner, H., & Charles, K. J. (2016). Cleaning up the big muddy: A meta-synthesis of the research on the social impact of dams. *Environmental Impact Assessment Review*. https://doi.org/10.1016/j.eiar.2016.02.007.

Kirchherr, J., Pomun, T., & Walton, M. J. (2016). Mapping the social impacts of 'damocles projects': The case of Thailand's (as yet Unbuilt) Kaeng Suea Ten Dam. *Journal of International Development*. https://doi.org/10.1002/jid.3246.

Landemore, H. (2014). Truth and democracy. *Contemporary Political Theory, 13*(2), e7–e11. https://doi.org/10.1057/cpt.2013.22. Palgrave Macmillan UK.

Martinez-Alier, J. (2002). *The environmentalism of the poor: A study of ecological conflicts and valuation*. Cheltenham, United Kingdom: Edward Elgar Publishing.

McCormick, S. (2006). The Brazilian anti-dam movement: Knowledge contestation as communicative action. *Organization & Environment, 19*(3), 321–346. https://doi.org/10.1177/1086026606292494.

McCully, P. (2001). The use of a trilateral network: An activist's perspective on the formation of the world commission on dams. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.920.5479&rep=rep1&type=pdf.

Merkel, W. (2010). Are dictatorships returning? Revisiting the 'democratic Rollback' hypothesis. *Contemporary Politics, 16*(1), 17–31. https://doi.org/10.1080/13569771003593839. Routledge.

Min, K. M., & Win, Z. M. (2016). *Chinese delegation visits myanmar locals displaced by controversial dam project*. Mekong Eye. https://www.mekongeye.com/2016/06/07/chinese-delegation-visits-myanmar-locals-displaced-by-controversial-dam-project/.

Morlino, L. (2008). Hybrid regimes or regimes in transition? http://fride.org/download/WP70-Hybrid_regimes_ENG_sep08.pdf.

Nakayama, M., & Fujikura, R. (2003). Comparative study on resettlement schemes applied for two dam construction projects In Indonesia. http://www.iwra.org/congress/resource/MADRID2003_MIKIYASU_NAKAYAMA_EN.pdf.

NCA (2014). Thirty fifth annual report 2014-2015. http://nca.gov.in/forms_pdf/annual-report/annual-report-2014-15.pdf.

van Niekerk, P. H., & Du Plessis, J. A. (2013). Hydrologic-economic appraisal of life-cycle costs of inter-basin water transfer projects. *Water SA, 39*(4), 539–548. https://doi.org/10.4314/wsa.v39i4.13. Water Research Commission (WRC).

Nombre, A. (2014). Yes, we need to build more large dams for water storage and energy for sustainable development! http://www.icold-cigb.org/share/article/0/icold-president-answers-oxford-misleading-study.

Nüsser, M., & Baghel, R. (2017). The emergence of technological hydroscapes in the anthropocene: Socio-hydrology and development paradigms of large dams. In *Handbook on geographies of technology* (pp. 287–301). Cheltenham, United Kingdom: Edward Elgar Publishing.

Nüsser, M. (2003). Political ecology of large dams: A critical review. https://www.researchgate.net/profile/Marcus_Nuesser/publication/273380663_Political_ecology_of_large_dams_A_critical_review/links/54ff6fc70cf2eaf210ba5e21.pdf.

Packendorff, J. (1995). Inquiring into the temporary organization: New directions for project management research. *Scandinavian Journal of Management, 11*(4), 319–333. https://doi.org/10.1016/0956-5221(95)00018-Q.

Patterson, B. (2014). *Modi approves increased height for Sardar Sarovar Dam*. The Council of Canadians. https://canadians.org/blog/modi-approves-increased-height-sardar-sarovar-dam.

Perkins, A. (2017). *Scientists are armed with the truth. But it won't win them the culture war*. The Guardian https://www.theguardian.com/commentisfree/2017/apr/23/scientists-truth-culture-war-donald-trump-pope.

Phanida (2016). *Residents in injanyang call for permanent cancellation of myitsone project*. Mizzima. http://www.mizzima.com/news-domestic/residents-injanyang-call-permanent-cancellation-myitsone-project.

Pickrell, D. H. (1989). Urban Rail Transit Projects: Forecast versus Actual Ridership and Costs. https://ntl.bts.gov/lib/56000/56600/56645/urbanrailtransit00unse_0.pdf.

Pohlner, H. (2016). Institutional change and the political economy of water megaprojects: China's south-north water transfer. *Global Environmental Change, 38*, 205–216. https://doi.org/10.1016/j.gloenvcha.2016.03.015.

Porta, D. D., & Diani, M. (2015). *The Oxford handbook of social movements*. Oxford, United Kingdom: Oxford University Press.

Rest, M. (2012). Generating power: Debates on development around the Nepalese Arun-3 hydropower project. *Contemporary South Asia, 20*(1), 105–117. https://doi.org/10.1080/09584935.2011.646071. Routledge.

Richter, B. D., Postel, S., Revenga, C., Scudder, T., Lehner, B., Churchill, A., & Chow, M. (2010). Lost in development's shadow: The downstream human consequences of dams. *Water Alternatives, 3*(2), 14–42.

Routledge, P. (2003). Voices of the dammed: Discursive resistance amidst erasure in the Narmada Valley, India. *Political Geography, 22*(3), 243–270. https://doi.org/10.1016/S0962-6298(02)00095-1.

Schneider, C. Q., & Wagemann, C. (2006). Reducing complexity in qualitative comparative analysis (QCA): Remote and proximate factors and the consolidation of democracy. *European Journal of Political Research, 45*(5), 751–786. https://doi.org/10.1111/j.1475-6765.2006.00635.x.

Scudder, T. (2005). *The future of large dams: Dealing with social, environmental, institutional and political costs*. London, United Kingdom: Earthscan.

Scudder, T. (2012). Resettlement outcomes of large dams. In C. Tortajada, D. Altinbilek, & A. K. Biswas (Eds.), *Impacts of large dams: A global assessment*. Berlin, Germany: Springer.

Shapiro, I. (2003). *The moral foundations of politics*. New Haven, United States: Yale University Press.

Shoemaker, B., Baird, I. G., & Manorom, K. (2014). Nam Theun 2: The World Bank's narrative of success falls apart. *International Rivers* https://www.internationalrivers.org/resources/8479.

Showers, K. B. (2009). Congo river's grand Inga hydroelectricity scheme: Linking environmental history, policy and impact. *Water History, 1*(1), 31–58. https://doi.org/10.1007/s12685-009-0001-8. Springer Netherlands.

Smyth, E., & Vanclay, F. (2017). The social framework for projects: A conceptual but practical model to assist in assessing, planning and managing the social impacts of projects. *Impact Assessment and Project Appraisal, 35*(1), 65–80. https://doi.org/10.1080/14615517.2016.1271539. Taylor & Francis.

Söderlund, J. (2003). On the development of project management research: Schools of thought and critique. *International Project Management Journal, 8*(1), 20–31.

Souksavath, B., & Nakayama, M. (2013). Reconstruction of the livelihood of resettlers from the Nam Theun 2 hydropower project in Laos. *International Journal of Water Resources Development, 29*(1), 71–86. https://doi.org/10.1080/07900627.2012.738792. Taylor & Francis Group.

Sovacool, B. K., Gilbert, A., & Nugent, D. (2014). An international comparative assessment of construction cost overruns for electricity infrastructure. *Energy Research & Social Science, 3*, 152–160.

Svejvig, P., & Andersen, P. (2015). Rethinking project management: A structured literature review with a critical look at the brave new world. *International Journal of Project Management, 33*(2), 278–290. https://doi.org/10.1016/j.ijproman.2014.06.004.

Tam, L. T., Valera, A. C., Tan, H.-P., & Koh, C. (2016). Online detection of behavioral change using unobtrusive eldercare monitoring system. In *Proceedings of the 11th International Conference on Queueing Theory and Network Applications – QTNA '16* (pp. 1–8). New York, New York, USA: ACM Press. https://doi.org/10.1145/3016032.3016053.

Tarrow, S. (1994). *Power in movement: social movements and contentious politics*. Cambridge, United Kingdom: Cambridge University Press.

Tilt, B., Braun, Y., & He, D. (2009). Social impacts of large dam projects: A comparison of international case studies and implications for best practice. *Journal of Environmental Management, 90*(Suppl 3), S249–S257. https://doi.org/10.1016/j.jenvman.2008.07.030.

Tremblay, M. C. (2007). *Uncertainty in the information supply chain: Integrating multiple health care data sources*. University of South Florida.

Vanclay, F. (2002). Conceptualising social impacts. *Environmental Impact Assessment Review, 22*(3), 183–211. https://doi.org/10.1016/S0195-9255(01)00105-6.

Vanclay, F. (2017). Project-induced displacement and resettlement: From impoverishment risks to an opportunity for development? *Impact Assessment and Project Appraisal, 35*(1), 3–21. https://doi.org/10.1080/14615517.2017.1278671. Taylor & Francis.

Wachs, M. (1987). Forecasts in urban transportation planning: uses, methods, and dilemmas. *Climatic Change, 11*(1–2), 61–80. https://doi.org/10.1007/BF00138795. Kluwer Academic Publishers.

Walker, G. (2013). A critical examination of models and projections of demand in water utility resource planning in England and Wales. *International Journal of Water Resources Development, 29*(3), 352–372. https://doi.org/10.1080/07900627.2012.721679. Taylor & Francis Group.

Watts, J. (2010). *Three Gorges Dam may force relocation of a further 300,000 people*. The Guardian.

WCD (2000). Dams and development – A new framework for decision-making. http://www.unep.org/dams/WCD/report/WCD_DAMS report.pdf.

Wilmsen, B., & van Hulten, A. (2017). Following resettled people over time: The value of longitudinal data collection for understanding the livelihood impacts of the three gorges dam, China. *Impact Assessment and Project Appraisal, 35*(1), 94–105. https://doi.org/10.1080/14615517.2016.1271542. Taylor & Francis.

Xie, L., & Van Der Heijden, H.-A. (2010). Environmental movements and political opportunities: The case of China. *Social Movement Studies, 9*(1), 51–68. https://doi.org/10.1080/14742830903442527. Taylor & Francis Group.

Yasuda, Y. (2015). *Rules, Norms and NGO Advocacy Strategies: Hydropower Development on the Mekong River*. London, United Kingdom: Earthscan.

Zarfl, C., Lumsdon, A. E., Berlekamp, J., Tydecks, L., & Tockner, K. (2015). A global boom in hydropower dam construction. *Aquatic Sciences, 77*(1), 161–170. https://doi.org/10.1007/s00027-014-0377-0. Springer Basel.

Ziv, G., Baran, E., Nam, S., Rodríguez-Iturbe, I., & Levin, S. A. (2012). Trading-off Fish Biodiversity, Food Security, and Hydropower in the Mekong River Basin. *Proceedings of the National Academy of Sciences of the United States of America, 109*(15), 5609–5614. https://doi.org/10.1073/pnas.1201423109.