3. Research Design and Methodology

This chapter presents the research design and methodology of this study. In the first section, I introduce the comparative case study design which combines a cross-case analysis of three case studies with a within-case analysis by focusing on Action Situations, with the overarching aim to uncover causalities (Section 3.1). In this context, I also discuss the selection of case studies, which is guided by the theoretical framework of this study, as well as the selection of Action Situations for the within-case analysis. In the second section of this chapter, I justify my methods for data collection and data analysis, namely Process Tracing and Qualitative Content Analysis, and discuss different types of assessment of variables (Section 3.2).

3.1 Comparative case study design

The empirical objective of my study, in a nutshell, is to understand how and why environmental objectives of the Water Framework Directive (WFD) have not been achieved in Spain despite strong public efforts. I thus aim to understand and explain governance processes, their determinants, as well as outcomes. To do so, a comparative case study is deemed particularly suitable. A single-case study is defined as an in-depth examination of a "spatially delimited phenomenon […] observed at a single point in time or over some period of time", with the intention to "shed light on a larger class of cases" (Gerring 2006: 19–20); whereas in a comparative case study, several single-case studies are comparatively analysed, which allows, inter alia, to detect similarities, differences, or patterns across cases. The main reason why I employ a comparative case study is that single as well as comparative case studies enable researchers to answer "how" or "why" questions (Yin 2018). Comparative case studies hence allow to explain certain phenomena by identifying causal relationships through the method of comparison (Yin 2018; Lauth, Pickel, and Pickel 2015). Furthermore, to meaningfully uncover causalities, the broader context in which causal mechanisms unfold need to be taken into account (see also next paragraph), which makes case studies particularly advantageous. More specifically, I undertake a *cross-case comparison* of three River Basin Districts (RBDs), and combine this with a *within-*

case analysis to reveal causal mechanisms unfolding within each case in the different Action Situations (George and Bennett 2005).

Since uncovering causal mechanisms is key to case studies as well as to this work, the understanding of causality underpinning this study needs to be explained. Causal mechanisms are defined in this study as unobservable physical, social, or psychological processes through which, in specific contexts, outcomes are generated (George and Bennett 2005). This definition adopts the view of contingent causal relations, meaning that causal mechanisms operate under scope conditions and are context dependent; which is why the effects of causal mechanisms also depend on interaction with other mechanisms (George and Bennett 2005). Similarly, Falleti and Lynch (2009: 1144) argue that causal explanations in social science can be identified "if and only if" the "interaction between causal mechanisms and the context in which they operate" is considered, since causal mechanisms operate differently in different contexts and under different conditions. The importance of contingency is also in line with much of the research on social-ecological systems, which understands social-ecological systems as highly context dependent; and where causality is seen as non-linear and dynamic (Preiser et al. 2021). The study's approach to identify causal pathways through which particular configurations of variables under certain conditions lead to specific outcomes thus corresponds with George and Bennett's (2005) "typological theory"; as well as with the Social-Ecological Systems (SES) framework, which is about "typologically decomposing" resource and governance systems and relating different system subtypes to outcomes (E. Ostrom and Cox 2010: 10).

However, identifying causalities in social science research, and in my study, is not without challenges. First caveats concern the fundamental challenge of isolating one causal mechanism from another, and identifying the specific circumstances under which causal mechanisms become activated (George and Bennett 2005); or, as Steinberg (2007: 183–4) states, to "say something meaningful about isolated components […] in a world that is in fact highly connected". Indeed, fully uncovering causalities requires undertaking a perfectly controlled experiment where the researcher changes one variable to observe the effect on the outcome – an endeavour which is obviously not possible in social science research. Despite these constraints, the research of this study is designed to nonetheless approximate causalities. Indeed, small-N analysis (Steinberg 2007), comparative case studies (George and Bennett 2005), and process tracing (Blatter and Haverland 2014; Trampusch and Palier 2016) are all methods that allow, albeit to a limited extent, to capture causalities.

A second challenge of drawing (causal) inference in comparative case studies concerns the extent to which generalizations are possible. According to Gerring (2006: 79–80), case studies always "partake of two worlds: they are particularizing *and* generalizing". Thus, while in-depth understanding of the single cases is of high importance – especially because case studies are often chosen to understand a

particular empirical puzzle where existing knowledge is limited – they also allow to "generalize across a larger set of cases of the same general type" (Gerring 2006: 65). Yin (2018) thereby highlights the importance to distinguish between statistical generalizations and generalizations from case study research. The former is about drawing inferences from a population of cases, based on data collected from a sample of that population. In contrast, generalizations in case study research are analytical, i.e., they are valid for theoretical propositions rather than populations (Yin 2018). Notwithstanding, all forms of generalizations in social science have their limitations, since they are, as George and Bennett (2005: 130–131) argue, "necessarily contingent and time-bound, or conditioned by ideas and institutions that hold only for finite periods", and are therefore "increasingly narrow". Thus, once again, it is important to be specific about the different contextual conditions under which configurations of variables are at work. Therefore, in the following I explain my rationales for case study selection, as well as similarities and differences of the three cases.

3.1.1 Selection of case studies and cross-case comparison

To undertake case study research, "*the* key question" concerns the definition of criteria for case study selection, as well as the case study selection itself (Herron and Quinn 2016: 459, italics in original). Indeed, the case study selection procedure is highly important because to meaningfully compare cases, they also need to have comparable characteristics. Furthermore, generalizations that can be drawn from case studies ultimately depend on how they have been selected – thus, whether findings of selected case studies are also relevant for other cases depends on how they relate to each other. Although there is no "general theory of purposive sampling", as argued by Agrawal (2001: 1662), it is clear that "selected cases should represent variation on theoretically significant causal factors". Thus, to select cases for cross-case comparison, I undertake a theory-guided purposive sampling. The selection is hence based on particular variables of the theoretical framework of this study (see Chapter 2), combined with a thorough understanding of the empirics of the cases, thereby aiming to ensure that selected cases are also of empirical relevance in the context of the topic under investigation. By doing so, I can ensure external validity, referring to the generalizability of empirical findings beyond the single case study (Yin 2018).

A wide range of methods exists for the selection of cases (for an overview, see Gerring and Cojocaru 2016). In this study, I undertake a combination of John Stuart Mill's method of agreement *and* method of difference, which Mill frames as Joint Method of Agreement and Difference (Seawright and Gerring 2008). I thus combine the Most Different System Design with the Most Similar Systems Design. In the Most Similar Systems Design, relying on the method of difference, researchers

compare very similar cases that show differences in the outcome variable (George and Bennett 2005; Lauth, Pickel, and Pickel 2015) – which, as I will discuss below, is represented by the Jucar and the Guadalquivir in my study design. In contrast, in the Most Different Systems Design, relying on the method of agreement, researchers compare very different cases that nonetheless share the same outcome (Lauth, Pickel, and Pickel 2015) – which is reflected by the Jucar and the Mediterranean Basins (see below, Table 4).[1] Gerring (2006) calls this case selection technique the method of "diverse cases", which has also been applied in empirical research on water governance in Europe (Kochskämper, Challies, et al. 2017). However, since it has not been discussed much in literature on qualitative research methods, a generally recognized name does not exist yet (Gerring 2006); but the method resembles the "Method of Agreement and Difference" of Stuart Mill; or the "maximum variation" sampling of Patton (2015).

The main reason why I use the diverse cases selection technique is that the method allows me to identify various causal pathways that may lead to an outcome, based on the assumption of equifinality (Gerring and Cojocaru 2016; Gerring 2006). Equifinality refers to the fact that different causal mechanisms can lead to similar outcomes (George and Bennett 2005). This is because a full range of values on both, independent as well as dependent variables, can be covered through this method, facilitating to achieve a "maximum variance along relevant dimensions" (Seawright and Gerring 2008: 300). The method is thus in line with what George and Bennett (2005) understand as "typology theory". Further, a particular strength of this method is that it "probably has stronger claims to representativeness than any other small-*N* sample" (Seawright and Gerring 2008: 301). However, the abovementioned limitations of drawing generalizations in case study research similarly apply to this method.

Rationales for the selection of the Guadalquivir, Jucar, and the Mediterranean Basins of Andalusia

In the following, I explain the different steps of case study selection, guided by the study's theoretical framework while at the same time ensuring empirical relevance; which ultimately result in the selection of the Guadalquivir, Jucar, and the Mediterranean River Basins of Andalusia (hereafter: Mediterranean Basins) in Southern and South-west of Spain (see Figure 2). First, I decided to select different cases within

1 I am aware that these cases only reflect the Most Different Systems Design if I assume that the population of all possible cases includes only Spanish RBDs. Looking only at Spain, the Jucar and Mediterranean Basins indeed do show significant differences in the independent variable. However, if I enlarged the population of all cases to all European RBDs, for example, these two Spanish RBDs would need to be framed as being very similar. Compared to other European RBDs, contextual conditions would then be constant.

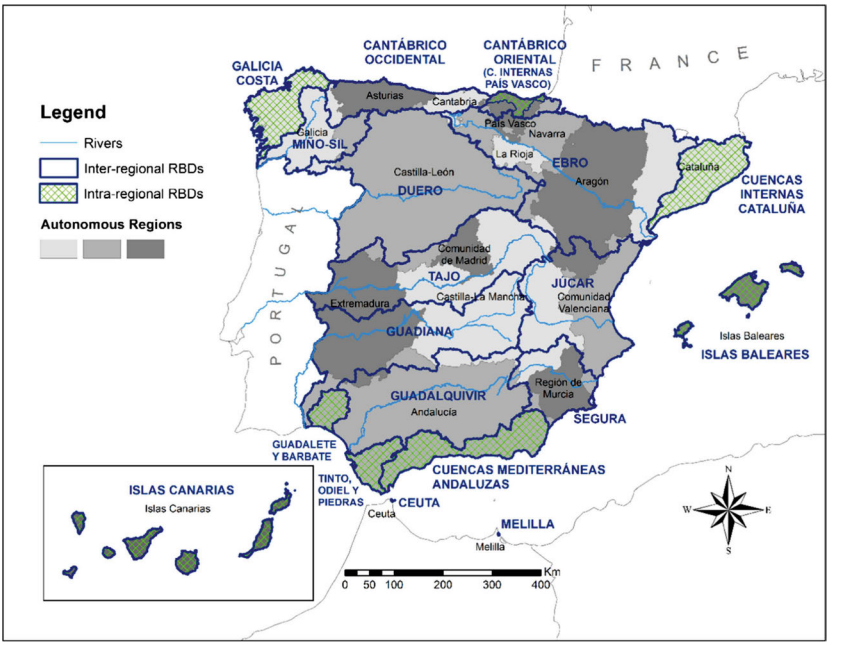
one country to keep the broader context in which cases are embedded constant. As mentioned above, an ideal setting to identify causalities is an experimental design where the external environment is strictly controlled (George and Bennett 2005). Since this is hardly possible in social sciences, the focus on one country nevertheless allows to create a relatively stable external environment and minimize confounding variables. Internal validity, referring to the correctness of the causal inference drawn by a researcher, can thereby be increased. Reasons to focus on Spain are of empirical nature: First, although the WFD implementation has been widely studied (Boeuf and Fritsch 2016), issues of water quality (see e.g., Boezeman, Wiering, and Crabbé 2020) received much higher attention than of water quantity (Acreman et al. 2010). This occurs despite the fact that over-abstraction of water is the second most common pressure on water bodies in Member States (European Commission 2012). Furthermore, the European Commission (2012: 6) highlighted already a decade ago the need to "put water quantity management on a much more solid foundation". Thus, research on governance processes to reduce over-abstraction certainly is of high empirical importance. Second, in the context of increasing irrigation efficiency, Spain is a highly relevant country, having the fifth largest sprinkler and micro irrigated area worldwide, and the second largest among the countries of the Global North, after the United States.[2]

To select cases within Spain, I aim for a variation on specific independent and dependent variables that are part of the theoretical framework; thereby following the above-mentioned method of diverse cases. Concerning the independent variable, I chose cases based on their variance along the *governance structure of the RBD*. This variable distinguishes between intra- and inter-regional river basins (see Chapter 2). While inter-regional basins are governed by regional authorities, intra-regional basins are governed by the national state through *Confederaciones Hidrográficas*. Furthermore, the legal framework differs in the two types of river basins: while the National Water Law is fully applicable in inter-regional basins, it only sets the broader legal context in intra-regional basins. Intra-regional basins can thus specify or go beyond the National Water Law through an own regional water law. Despite these differences, all RBDs are, obviously, embedded in a multi-level governance system where the EU law, and most importantly the WFD, applies. The legal status of the WFD implies, as with any other EU directive, that the EU sets specific goals which all Member States must achieve in a given period. At the same time, though, Member States have considerable leeway on *how* to achieve them. Therefore, even though all Spanish RBDs need to fulfil the same aim, we can expect to observe differences in the governance processes for WFD implementation between inter- and intra-regional RBDs. This selection criterion also means that transboundary RBDs are excluded as they have a different governance structure. The number of potential cases,

2 https://www.icid.org/sprinklerandmircro.pdf (accessed 30.06.2021)

i.e., the population of cases represented by all Spanish RBDs, can thereby be reduced from 25 to 18 cases, namely, 4 inter-regional RBDs and 14 intra-regional RBDs (see Table 14, Appendix 1 for all pre-selected RBDs).

Figure 2: Map of Spanish River Basin Districts



Source: De Stefano, Hernandez-Mora (2018)

In relation to the dependent variable, I selected cases based on their variance along the variable *development of agricultural water use,* which is also part of the theoretical framework. I chose this variable because the study's main empirical foci are processes that reduce agricultural water use; thereby representing a key explanandum. To ensure that the reduction of agricultural water consumption is also of empirical relevance in the respective RBDs, I pre-selected those that have a high share of agricultural water use. RBDs where agriculture accounts for less than 50% of total water use are therefore excluded. Followingly, six RBDs remain, namely the Guadalquivir, Jucar, and Segura as inter-regional RBDs; and the Mediterranean Basins, Guadalete-Barbate, and Tinto-Odiel-Piedras as intra-regional RBDs (see Table 14, Appendix 1). As a next step, I assessed the actual *development of agricultural water use*. I therefore analysed data from 2009 and 2016/17 included in the respective

River Basin Management Plans (RBMPs) of the first, second, and partially third planning cycle, depending on data availability in the different RBDs (see Table 15, Appendix 1). For data triangulation, and since these numbers refer to estimations of water use instead of actual water use (European Commission 2015b), I undertook scoping interviews and reviewed secondary literature (see also section 3.2). Based on these different data sources, I selected the Guadalquivir and Jucar as inter-regional RBDs, and the Mediterranean Basins as intra-regional RBD. In the following, I explain the empirical reasons for the selection of the respective cases, which are also summarized in Table 4.

The Guadalquivir was selected as a first case, representing a RBD where agricultural water use increased after the implementation of irrigation efficiency measures by 8,7%, from 2.569 hm^3 in 2009 to 2.792 hm^3 in 2016/17 (own calculations based on CHG 2013; 2020a). Furthermore, the Guadalquivir is often mentioned as an important example where a rebound effect (see Chapter 1) occurred (WWF/Adena 2015; Corominas and Cuevas 2017), and where the empirical relevance of irrigation efficiency measures is particularly high. This is because Andalusia, where almost the entire RBD is located, is the region where the largest areas were affected by irrigation efficiency measures, representing 40% of the so-called modernized area in Spain (Berbel and Gutiérrez-Martín 2017a).

The Jucar was selected as second inter-regional river basin, aiming to increase the variance on the variable *development of agricultural water use* – in line with the rationale of case selection procedure explained above. Indeed, the Jucar is the only inter-regional RBD where agricultural water use (slightly) decreased in the analysed time period, namely by 1.8% from 2009 (1.412 hm^3/year) to 2016/17 (1.386 hm^3/year) (own calculations based on CHJ 2014a; 2019a). Furthermore, the Jucar was mentioned by several interview partners from scoping and stakeholder interviews (Interview 21/2018, 22/2018, 14/2019, 15/2019) and in several empirical studies (Sanchis-Ibor et al. 2016) as an important case in Spain in terms of having prevented the rebound effect.

For the third case, I selected an intra-regional RBD, thereby increasing variance on the independent variable; as well as a case that also shows a decrease in agricultural water consumption, aiming to have a further case that contrasts the Guadalquivir. Having these criteria in mind, I selected the Mediterranean Basins, since between 2009 and 2015, its agricultural water use slightly decreased by 0.8 %, from 824 hm^3/year to 817 hm^3/year (own calculations based on Junta de Andalucía 2014a; 2019a). Even though also the RBD Tinto-Odiel-Piedras meets this criterion (see Table 15, Appendix 1), experts indicated that the political importance of increasing irrigation efficiency and reducing agricultural freshwater consumption is much higher in the Mediterranean Basins (Interview 2/2018).[3] Furthermore, agricultural

3 For the list of interviews, see Appendix 2

water use in the Tinto-Odiel-Piedras corresponds to less than half of what is used by agriculture in the Mediterranean Basins, which suggests a higher empirical relevance of the latter. The Mediterranean Basins of Andalusia include several river basins (see Chapter 4), but represents one River Basin District for the WFD implementation, which is why it composes a single case.

Table 4: Case study selection and its criteria

		Variance along the environmental outcome: Change in agricultural water use (2009–2016/17)	
		(Slight) decrease	Increase
Governance Structure of the RBD	Inter-regional RBD	**Jucar**	**Guadalquivir**
	Intra-regional RBD	**Mediterranean River Basins of Andalusia**	

Similarities and differences between case studies

To be able to meaningfully compare findings derived from case studies, it is important to know whether cases are actually comparable with each other. Furthermore, as discussed above, the possibility to generalize findings to other cases hinges on how cases relate to each other. It is therefore important to have a sound understanding of parallels and variations of case studies, in terms of variables that are considered of theoretical significance for my research. While selection criteria have been discussed for each case, in the following, I briefly present key similarities and differences of further independent variables included in the theoretical framework. More specifically, I focus on *contextual conditions* and *characteristics of heterogenous actors* which are both part of the theoretical framework (see Chapter 2 for definition of variables). *Social problem characteristics* as well as *overarching rules* are not discussed here since they apply to the level of Action Situation and therefore go beyond the scope of this chapter. The implications of these differences as well as similarities for drawing (causal) inference and deriving generalizations will be considered in the Discussion (Chapter 7). All variables will be analysed more in-depth in the empirical chapters (see Chapter 4, 5, and 6).

First, regarding *contextual conditions* of the case studies, it is to mention the second-tier variable *geographic and hydrological characteristics of the river basin district,* which are quite different among and within the three cases. Indeed, case studies show major differences concerning the size of the respective RBD, number of river basins governed within the RBD, main ecosystems, landscapes, or administrative boundaries. However, there are also important differences within each case:

all three RBDs have mountainous as well as flat areas, which considerably shape agricultural production systems; both the Guadalquivir and Jucar have protected wetlands where agriculture is restricted, as well as large-scale areas of intensive farming; and climatic conditions vary within the different RBDs, also affecting agricultural production. Concerning the second-tier variable *socio-economic role of irrigated agriculture*, cases are relatively similar. Indeed, agricultural production in all three cases depends on irrigation which plays an important role for employment as well as the social and political context in rural areas. Third, relative numbers of *water supply and demand* are alike in the three cases, with all cases having a high share of agricultural water demand, and total water demand approximating or even equalling water supply. Yet, cases differ in their absolute numbers of water demand and supply – mainly due to the different sizes of the RBDs –, as well as in their division between surface, groundwater, and non-conventional water resources. In the Guadalquivir and the Jucar, main water resources for irrigation are surface water, while groundwater and non-conventional water resources dominate in the Mediterranean Basins.

Second, *characteristics of heterogenous actors* are relatively similar in the three cases. More specifically, we can observe that in all three cases, *financial and human resources* of environmental actors are considerably lower than those of Water User Associations (WUAs). Further, financial resources also vary among WUAs, depending mostly on whether they are traditional WUAs using rainwater harvesting techniques, or financially better endowed WUAs that use regulated surface water distributed by the state. State actors in the three cases all report that they lack financial means and that they were significantly affected by the Euro crisis in 2008/09 and its consequences. However, regional actors and most importantly the Regional Ministry of Andalusia seem to have, in relative terms, lesser financial and human resources than its national counterparts, i.e., the River Basin Authorities of the Guadalquivir and the Jucar. Further, similar *narratives on water management* are used by actors in the three cases, even though the relative importance of the respective narratives vary. Actor groups in all three cases seem to agree on the problem of limited availabilities of water resources, but they identify different reasons as well as solutions to these problems, ranging from increasing water supply to improving governance or restricting water demand.

Having discussed the selection of case studies, I know turn to selection of action situations for the within case-analysis.

3.1.2 Selection of Action Situations for within-case analysis

Decision-making processes are studied in this book through Ostrom's (2005) Institutional Analysis and Development (IAD) Framework and the Network of Adjacent Action Situations (NAAS), developed by McGinnis (2011) (see Chapter 2). The

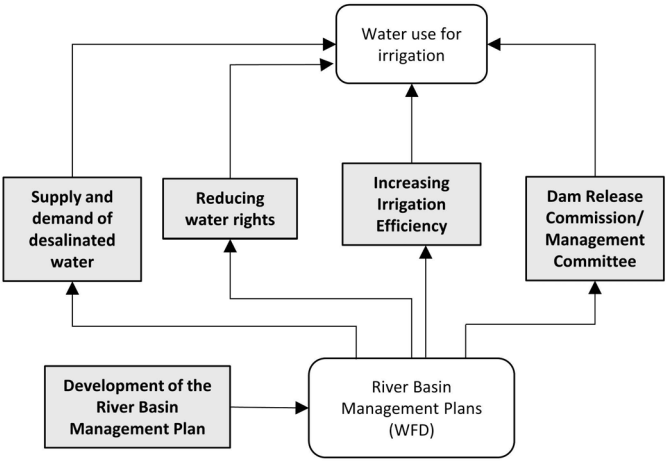
unit of analysis are Action Situations where participants interact with each other (E. Ostrom 2005). The in-depth analysis of different decision-making processes represents a *within-case analysis*. As the name suggests, a within-case analysis allows researchers to observe causal processes within a case (Goertz and Mahoney 2013). Gerring (2006: 204) even argues that it is unlikely that "one has satisfactorily explained an outcome until one has explored *within-case* evidence". In this book, I thus combine cross-case comparison of the three case studies with a within-case analysis (George and Bennett 2005; Gerring 2006) through the focus on Action Situations. Rohlfing (2012) calls such a research design an integrative comparative case study. A common method to undertake within-case analyses is process tracing (Goertz and Mahoney 2013; Collier 2011), which I will introduce below.

I selected four Action Situations that occurred in all three case studies, and one additional Action Situation that is only of relevance in the Mediterranean Basins, namely Demand and Supply of Desalinated Water. All Action Situations are embedded in the overarching processes of WFD implementation of the first and second planning cycle (see Figure 3). There are multiple ways used in the literature to delineate Action Situations, ranging from boundary drawing along governance functions (McGinnis 2011) to selecting Action Situations according to their type of social interaction (for an overview, see Oberlack et al. 2018). Thus, delineating Action Situations is left to the discretion of the researcher. In this work, I draw on the Management and Transition Framework of Pahl-Wostl et al. (2010), and delineate Action Situations broadly based on the policy cycle, albeit only regarding the phases of planning and implementation; as well as based on the type of actors participating in the different decision-making processes. The focus on the policy cycle seems suitable since the governance process stipulated by the EU for WFD implementation undergoes phases as delineated in the policy cycle (Newig and Koontz 2014). However, I acknowledge that focusing on the WFD implementation risks overlooking other and more informal processes which nevertheless may influence farmers' decision-making regarding their water consumption. Furthermore, policy processes are usually more complex than their representation in a policy cycle (Wegrich and Jann 2006). Indeed, instead of undergoing a sequence of different steps, policies may be adapted while being implemented, e.g., due to lack of finances or changed political priorities. However, since I analyse policy stages through the analytical lens of an Action Situation, I explicitly consider institutions, as well as actors' interests and incentives, which allows me to better capture the complexity of the policy process.

The selection and delineation of Action Situations was based on scoping interviews (see below). Although selected Action Situations occur in all three case studies, their relative importance varies; and formal coordination processes sometimes differ, e.g., between intra- and inter-regional RBDs. Nevertheless, for analytical purposes and to facilitate cross-case comparison, these partly varying decision-making

processes are subsumed under the same Action Situations in each case. In the following, I explain the different Action Situations and justify their selection.

Figure 3: Network of Action Situations



Source: Own illustration. The Action Situation Supply of Desalinated Water will only be analysed in the Mediterranean Basins due to little empirical relevance in the other two cases.

First, the Action Situation *Development of River Basin Management Plans* concerns the planning phase for the WFD implementation, ranging from compiling measures to participatory processes and the final RBMP approval. RBMPs need to be developed every six years, outlining all measures which will be taken to meet the WFD objectives. RBMPs are thus the cornerstone of the WFD implementation, which is why they are included as an Action Situation. Second, the Action Situation *Dam Release Commissions* (denominated *Management Committee* in the Mediterranean Basins) is about decision-making processes regarding water allocation to different groups of water users. Members of Dam Release Commissions decide on the reservoirs' filling level during the wet season and upon the schedule and volume of water storage releases during the dry season. Thereby, decisions by the Dam Release Commission may immediately affect the amount of agricultural water use. Third, the Action Situation *Increasing Irrigation Efficiency* analyses what is commonly called "modernization of irrigation" in Spain, namely the implementation of new irrigation techniques such as drip irrigation as well as the replacement of irrigation canals and ditches with pipes (see Chapter 1). These measures aim to increase irrigation efficiency and

are of high empirical importance in all three case studies. Furthermore, the Action Situation *Supply and Demand of Desalinated Water* addresses the implementation of desalination plants of seawater and brackish water, also aiming to reduce freshwater consumption in agriculture. This Action Situation only concerns the Mediterranean Basins since there are no desalination plants of empirical relevance in the other two RBDs. The last Action Situation *Reducing Water Rights* is about reducing water rights after the increase of irrigation efficiency in order to avoid a rebound effect. Furthermore, it also includes changing the type of water right in the context of desalination, i.e., replacing the right to withdraw surface water or groundwater with the right to use desalinated water. Both measures are inherently linked with the technical measures of increasing irrigation efficiency and desalinating water, which is why they are included as an Action Situation in this study as well.

Selected Action Situations can be seen as different phases of the policy cycle, as mentioned above. The *Development of River Basin Management Plans* relates to the planning phase of the policy cycle, while the other three Action Situations concern policy implementation. However, there is a main difference between a rather classical policy cycle and the policy cycle for WFD implementation. Traditionally, policy implementation is understood as bureaucrats carrying out decisions taken by political actors (Newig and Koontz 2014). In contrast, in the context of WFD implementation, those actors who are in charge of the planning phase (i.e., *Confederaciones Hidrográficas*) are also responsible for implementing the respective plans, as well as evaluating their implementation. Newig and Koontz (2014) term this the "EU's mandated participation planning approach", where the formulation of plans is mandated to subnational actors.

Having outlined the comparative case study design, including selection of case studies and of Action Situations, I describe the process of data collection and analysis in the next section.

3.2 Collection and analysis of data

This study follows a mixed methods approach even though the major focus lies on qualitative data. Mixed methods combine and integrate qualitative and quantitative data aiming to compare various perspectives drawn from the different types of data (Creswell 2014). Especially in research on social-ecological systems, mixed methods are considered useful in order to "acquire more support for a potential explanation of a complex phenomenon" (de Vos et al. 2021: 52). The study's main focus on qualitative data is due to my interest in *understanding* social and political processes and their outcomes (Lauth, Pickel, and Pickel 2015). Indeed, qualitative research allows to uncover decision-making processes in-depth by integrating different perspectives and multiple realities of persons involved in these processes (Creswell 2014),

and by analysing a broad range of variables. In addition, I use quantitative data to complement and triangulate findings from qualitative data, especially concerning the environmental outcome.

In the following, I explain the processes of data collection and data analysis in the case studies, as well as the variables' assessment. With this section, I aim to ensure reliability of the study, i.e., to provide the possibility to repeat the study and arrive at similar results (Yin 2018).

3.2.1 Data collection in case studies

Empirical data was collected through scoping and stakeholder interviews (N=53) and document analysis (Yin 2018), thereby aiming to increase the validity of the measurement. Data was collected until a certain degree of saturation was achieved, meaning that collection of new data would most probably not have revealed new insights.

Scoping interviews

I conducted scoping interviews (N=6) with scholars and external experts in October 2017 and June 2018. Scoping or key informant interviews are often used at the beginning of an empirical study to generate contextual and background information from people who hold useful knowledge for the study (Shackleton et al. 2021). Aims of the first two to three scoping interviews thus were to identify and gain an overarching insight of the main empirical field of my study, as well as to detect empirical research gaps. Subsequent scoping interviews were used to select and discuss cases and relevant Action Situations.

I selected interview partners for scoping interviews based on pre-established contacts, as well as through snowball sampling. Scoping interviews were open-ended, although guided by some general questions. They were not recorded, but detailed notes were taken during and after the respective interviews.

Stakeholder interviews

Stakeholder interviews are the main means for data gathering in this study. Through stakeholder interviews, data is collected from people who are themselves part of the case study. The main reasoning behind stakeholder interviews is that they allow to reconstruct and explain social and political processes (Gläser and Laudel 2010), i.e., to generate descriptive as well as explanatory knowledge (Shackleton et al. 2021) – thereby corresponding to the overarching rationale of this study.

More specifically, I conducted semi-structured in-depth interviews (N=47) in June 2018, October/November 2018, June/July 2019, and October 2019 for all three cases. All but one of the in-depth interviews were recorded. Interviewees were guaranteed anonymity. I excluded two of the 47 interviews at the stage of the analysis since the interviewees' expertise did not match with the Action Situations under in-

vestigation. Number of interviews are divided between the case studies as follows: 16 on the Guadalquivir, 14 on the Jucar, 14 on the Mediterranean River Basins, and three on the national level (see Table 16, Appendix 2). Interviews were conducted in Spanish and their recordings were fully transcribed by a student research assistant, also in Spanish. While I have very good Spanish language skills, the fact that I am not a native speaker may have affected the conduction of interviews, e.g., the accuracy of the questions asked. Further, there is a risk that in those cases where interviews were not recorded (scoping interviews and one in-depth interview), information may have been lost. Yet, I argue that due to the relatively high number of interviews conducted, transcriptions carried out by a native speaker, and the use of data triangulation with documents, the overall data quality of this study is not affected.

Interview partners were selected aiming to achieve a balanced representation from the water and agricultural sector operating at different levels, i.e., the local, regional and the national level. This includes national and regional public administrations, WUAs, agricultural organizations, or environmental NGOs. The identification of interview partners consisted in several steps. I first analysed RBMPs of the second planning cycle, namely participant lists of the participatory processes and written statements (*alegaciones)* submitted by actors to the RBMP. This was complemented by snowball sampling in the scoping as well as stakeholder interviews. Interview partners within the identified organisations were chosen based on their experience with the WFD implementation in the respective case study, with a particular focus on the management of agricultural water use. In many cases, these persons were in a leading position of the respective organization and were male.

As mentioned above, interviews were semi-structured, and therefore steered by an interview guideline. Semi-structured interviews are suitable when the research aim is to reconstruct social processes. The use of a guideline then ensures that all topics relevant to understand the particular process are covered, while at the same time questions can be adapted to different interview situations and emerging issues (Gläser and Laudel 2010). I tailored interview guidelines to the case study and the respective type of actor, i.e., public, private, or civil society actor. Thereby, I tried to ensure that questions related to the empirical context of the respective interviewee. Guidelines covered independent as well as dependent variables, and were developed deductively.

Documents and grey literature

Lastly, I collected policy documents and grey literature to better capture the complexity of water governance systems under investigation. Indeed, this allows to triangulate interview data as well as to integrate quantitative data to the study, thereby undertaking the mixed-methods approach. In this context, I identified policy documents and grey literature based on formal documents for WFD im-

plementation, snowball sampling as well as through stakeholder and scoping interviews. Most importantly, these documents include the RBMPs of the first, second, and third planning cycle, including the different accompanying and/or related documents such as draft RBMPs, Scheme of Important Issues, annexes, etc. Data in these documents are of qualitative as well as quantitative nature. It was mainly used to measure output performance, i.e., *political output performance* and *environmental outcome performance*, but also for some of the independent variables, such as *contextual conditions* and *overarching rules*. Further, grey literature includes inter alia press releases, public statements, or reports, mostly from the European Commission, national and regional authorities, as well as stakeholder groups which were published in the period of analysis (i.e., 2009–2019).

3.2.2 Data analysis

Process tracing

To identify causal relationships in the three case studies, I conduct process tracing and analyse primary and secondary data through Qualitative Content Analysis. I use process tracing since this method enables researchers to identify intervening causal processes between the independent and dependent variables; which is why it is particularly suitable for within-case analysis (George and Bennett 2005; Collier 2011), as undertaken in this study through the focus on Action Situations (see also above). The method has received increasing attention in political science in the last decades, which is why various definitions and forms of process tracing are used in the literature (for an overview, see Trampusch and Palier 2016). Here, it is defined as an "analytic tool for drawing descriptive and causal inferences from diagnostic pieces of evidence" (Collier 2011: 824). Furthermore, Gerring (2006: 173) argues employing "multiple types of evidence […] for the verification of a single inference" to do process tracing, mainly based on qualitative, but also on quantitative data. The mixed-method approach applied in this study is therefore also suitable for process-tracing. In a next step, noncomparable observations drawn from different types of data need to be "ordered, categorized, 'narrativized'" (Gerring 2006: 180). This helps the researcher to uncover the timing and sequence of events or situations (Collier 2011). Breaking down the overarching process of WFD implementation into several interdependent Action Situations is thus considered helpful in this regard.

Process tracing complements the study's research design as well as its theoretical framework, since it is based on similar underlying assumptions than those of comparative case studies as well as of Ostrom's (2005) IAD Framework. Indeed, process tracing (Blatter and Haverland 2014) as well as case studies (Yin 2018) are particularly suitable to answer "why" questions, i.e., to explain outcomes. Furthermore, as Blatter and Haverland (2014) explain, process tracing is based on configurational thinking, and by focusing on contexts and intervening variables to

understand causalities, it takes contingency into account. Thus, causal paths that are identified through process tracing consist of multiple independent variables, feedback loops as well as contextual evidence. This is in line with the theoretical framework of this study, where variables are understood as being configural, interacting and mutually influencing each other (E. Ostrom 2005).

Qualitative Content Analysis

The data which is used to conduct process tracing is analysed through Qualitative Content Analysis. Qualitative Content Analysis is a research method that allows me to identify and categorize patterns in texts, and to make inferences which are replicable (Patton 2015; Krippendorff 1989). Furthermore, it is a rule-guided approach, which allows for tracing the process of data analysis also at a later stage. It combines strengths of quantitative content analysis with a more qualitatively oriented procedure for text interpretation (Mayring 2015). To carry out Qualitative Content Analysis, several methodological and analytical steps are required, from the development of codes to coding of data, and writing the analysis (Kuckartz 2019), as explained in the following.

Elaborating a coding scheme is a key part of Qualitative Content Analysis. I developed codes in an iterative way, thereby combining a "concept-driven and data-driven development of codes" (Kuckartz 2019: 185). I thus first developed codes deductively based on the theoretical framework; and added further codes during the process of coding itself, i.e., based on the interview material. Coding and developing the code book thus underwent several cycles of respective adjustments. This iterative approach on the one hand allowed me to fully consider the theoretical framework; and on the other, it was possible to incorporate all aspects that are relevant to answer the research questions, but which were not expected or unpredicted when designing the coding scheme in the first place. This approach is considered appropriate to be able to make theoretical contributions at a later stage and is in line with the iterative development of the theoretical framework. This means that changes and additions to the code book were then also considered in the theoretical framework. The code book includes all variables, except variables categorized under *social problem characteristic*. This is because I only added them at a later stage, after interview material had been coded. The analysis of these variables is thus based on a thorough understanding on and interpretation of the different Action Situations, instead of an analysis of the interview material.

In the coding process itself, I selected text segments and assigned the respective categories by using the software program atlas.ti. During this process, I additionally paraphrased every coded text segment. The purpose was to further condense the interview material, as well as to translate the content of the different text segments from Spanish into English. In a third step, these paraphrased texts severed as a basis

to write descriptive summaries of every case study, which were then used to write the empirical chapters.

According to Kuckartz (2019: 196), Qualitative Content Analysis "tries to reach a consensus – as far as this is possible – on the subjective meaning of statements". I do not assume that other researchers would code this study's interview material identically as I have done it; however, this method as well as defining codes in the code book aims to make the lens through which I analysed the data explicit. Furthermore, I discussed coded material of the Guadalquivir case with three colleagues; we therefore held several online meetings during a period of approximately two months. In this process, codes were refined as well as coded segments were refined and adapted. Therefore, some degree of intercoder reliability could be ensured. However, this subjectivity in the analysis of the data is neither due to the particular method of Qualitative Content Analysis, nor to the research design of the case study. As Gerring (2006: 69–70) points out: "All data requires interpretation, and in this respect all techniques of evidence gathering are *interpretive*. Rarely, if ever, does the evidence speak for itself. […] Social science is, of necessity, an interpretive act."

3.2.3 Assessment of variables

The final step in condensing information in this study consists of determining the value of each variable. I thereby make use of nominal as well as ordinal scales, but also qualitatively describe some variables, depending on the respective type of variable. Reasons to use nominal and ordinal scales, which I describe below, are to reduce complexity of the collected data, as well as to make the assessment more transparent. Furthermore, assigning values – such as high, moderate or low – to a variable enables to undertake a more structured comparison of the three case studies, ultimately helping to identify causal mechanisms. Similar to what I point out regarding Qualitative Content Analysis, also the method of assigning values is a subjective and interpretative act. However, by doing this exercise explicitly rather than implicitly, the procedure is made more comprehensible, thereby increasing reliability. However, the reduction of complexity which I aim to achieve through this method necessarily implies a certain loss of information. In the cross-case comparison (see Chapter 7) it is thus important to not only compare the values of each variable, but also to consider the underlying justification. In the following, I explain the three different ways to assess variables in this study. The more specific form of assessment of each variable, including the operationalization of the different scales, is displayed in Table 5. I developed the different categories for categorial and ordinal variables inductively, i.e., after having gained an in-depth understanding of the different values of every variable in all three cases (George and Bennett 2005).

The largest group of variables in this study are ordinal variables. These are variables where we can assign discrete categories that can be ranked from lowest to high-

est, but where the distance between the different ranks is without meaning (Cox 2015). I use three-point ordinal scales, defined separately for each variable (see Table 5). Examples of ordinal variables in this study are *development of agricultural water use,* where respective scores are *reduced, constant* and *increased agricultural water use;* or the variable *human and financial resources of actors*, where respective scores are *high, moderate,* and *low*. I use ordinal scales for those variables where it is possible to apply a ranking, and where also the underlying research interest is in line with this ranking exercise. To give an example, if the *amount* of financial resources of actors is of interest to me – rather than the type or source of resources – I would use an ordinal scale. To get to the respective rank, I will first qualitatively describe each variable, and then base the assignment of categories on these descriptions.

One of the main difficulties in ranking variables certainly relates to having clear benchmarks. I did not define graded statements for each variable as part of the scoring scheme, indicating how to arrive at a particular score (see e.g., Dombrowsky et al. 2022). This decision is because in the stage of defining scores and respective statements it is not possible to foresee the full complexity as well as all the nuanced differences that will arise between case studies. However, also without defining graded statements, choosing a certain score is not arbitrary. In contrast, it is based on a weighing process that considers how often certain statements were raised by different interviewees; and, more importantly, it is the result of comparing values of variables across Action Situations as well as across cases. Arriving at a final score – indicating, for example, whether behaviour of actors is highly, moderately or not/marginally coordinated – therefore is an iterative process, where results are compared, and scores weighed and re-weighed. Since each score is preceded by a qualitative description, I ensure that it is comprehensible and understandable how the respective scores are arrived at.

The second group of variables is nominal variables. These are variables whose values are also classified into discrete qualitative categories, but unlike ordinal variables, it is not possible to rank them in a meaningful way. An example is gender, where the categories male, female, or non-binary stand side by side without hierarchical meaning. In this study, this group is much smaller than those of ordinal variables, and includes, for example, the variable *narratives on water management* with the categories *supply-side management*, *demand-side management*, *knowledge and governance*, and *deep ecology.* Very importantly, modes of coordination also represents an ordinal variable, where each pure form of coordination, i.e., *cooperation,* different forms *competition,* and *hierarchy,* as well as *information exchange, gaps* and *conflict* represent discrete categories on a nominal scale. The underlying reason to use a nominal scale is that I am interested in the *type* to which I can assign a particular interaction; and not in whether an interaction is more cooperative than another, for instance. Also for this group of variables, the comparison across Action Situations and across case studies is crucial to arrive at a category. Indeed, for instance, to find out

whether a behaviour can be classified as *hierarchical* depends on if hierarchy dominates in contrast to other Action Situations. This is because traits of hierarchy are likely to be found in any type of interaction within in a multi-level governance process. However, to avoid having to classify any pattern of interaction as a hybrid of all pure forms, it seems reasonable to compare results across Action Situations and across case studies.

Lastly, there are two variables which cannot be grouped under nominal or ordinal variables. This concerns *geographic and hydrological characteristics of the river basin district,* which I describe in a qualitative way. This reason is that it is not possible to use any kind of standardized measurement approach, or to structure the variable in a reasonable way (Cox 2015). In contrast, I focus on those characteristics that were considered important by interviewees to understand water governance and their outcomes in the respective case studies. Lastly, the variable *water supply and demand* is described based on quantitative information since I am interested in absolute numbers of different types of water resources available in the case studies.

Table 5: Overview of variables and their assessment scheme

	Variable	Definition	Assessment
Contextual conditions	Geographic and hydrological characteristics of the river basin district	Location, administrative and hydrological boundaries of the river basins; geography; main ecosystems.	Qualitative description
	Socio-economic role of irrigated agriculture	Relative importance of irrigated agriculture and the agri-food industry compared to other economic sectors, for economy and society.	Ordinal scale: high importance; medium importance; low importance
	Water supply and demand	Type and amount of water resources available for consumption.	Quantitative information
Characteristics of heterogeneous actors	Financial and human resources of actors	Endowments of public, private, and civil society actors in relation to the case study focus.	Ordinal scale: high; moderate; low
	Narratives on water management	Causal and explanatory beliefs of actors regarding status and reasons of existing water management problems.	Nominal scale: supply-side management; demand-side management; knowledge and governance; deep ecology
Overarching rules	Governance structure of the river basin district	Distinction between intra- and inter-regional RBD.	Nominal scale: intra-regional RBD; inter-regional RBD
	De jure autonomy	Extent of formal rights and competencies of governmental and non-governmental actors as stated by laws and regulations with respect to the case study focus.	Ordinal scale: high; moderate; low
	Formal rules for coordination	Formal institutions creating the structure for actors to interact with each other.	Nominal scale: cooperation, competition, hierarchy, hybrids

Social problem characteristics	Uncertainty	Complete lack of information, or insufficient information.	Ordinal scale: high uncertainty; moderate uncertainty; low uncertainty
	Asset specificity	Investments for a specific good or service which cannot be easily transferred to alternative uses.	Ordinal scale: high specificity; moderate specificity; low specificity
	Frequency	Number of times specific activities occur within a particular time period.	Ordinal scale: high frequency; moderate frequency; low frequency
	Spatial and jurisdictional scales	Dimension to study a particular phenomenon.	Nominal scale: Jurisdictional levels: national, regional, local Hydrological levels: basin, sub-basin
	Excludability	Possibility to exclude additional actors from using or suffering from a produced good or service at reasonable costs.	Ordinal scale: high excludability, moderate excludability, non-excludability
Modes of coordination	Hierarchy	Process of forced alignment of activities by a superior actor vis-à-vis an inferior actor.	Category on a nominal scale: Hierarchy
	Competition	Process of alignment of activities based on prices, (economic) incentives, or ideas.	Categories on a nominal scale: Idea-based competition, price-based competition, incentive-based competition
	Cooperation	Process of voluntary alignment of activities of actors to achieve a shared aim.	Category on a nominal scale: Cooperation
	Hybrid	Process of alignment of activities based on a combination of pure forms of coordination (hierarchy, competition, or cooperation).	Combinations of different categories of interaction (hierarchy, competition, cooperation)

Additional categories of interaction	Information exchange	Minimum form of coordination: One-way or two-way exchange of information among actors.	Category on a nominal scale: Information exchange
	Gaps of interaction	Situation where actors intentionally or unintentionally do not coordinate with each other (no information exchange, no alignment of behaviour).	Category on a nominal scale: Gap of interaction
	Conflict	Disagreements or disputes of actors that are not solved through any of the three pure forms of coordination.	Category on a nominal scale: Conflict
Process performance	Coordinated behaviour *(second-tier variable)*	Extent to which different types of interaction (*cooperation*, *competition*, *hierarchy*, and *hybrids*) lead to ordered patterns.	Ordinal scale: highly coordinated; moderately coordinated; not/marginally coordinated
	Information exchanged *(evaluative criterion)*	Extent to which information among actors within a process is exchanged; as well as to which information about the process and its output are available to outsiders of the process.	Ordinal scale: information exchanged; moderately exchanged information; information not/marginally exchanged
	Competing interests considered *(evaluative criterion)*	Extent to which contradictory interests which exist in society in relation to the case study focus are taken into account.	Ordinal scale: competing interests considered; competing interests partly considered; competing interests not/marginally considered
	Incentives aligned *(evaluative criterion)*	Extent to which an incentive structure is established which makes it rational for actors to behave in the expected way.	Ordinal scale: incentives aligned; incentives partly aligned; incentives not/lowly aligned

Output performance	Effectiveness of RBMP *(level of Action Situation)*	Extent to which the RBMP is likely to achieve the political goal of reducing agricultural water consumption.	Ordinal scale: effective; moderately effective; not/marginally effective
	Distribution of surface water adapted *(level of Action Situation)*	Extent to which surface water distribution has been adapted in the Dam Release Commission, compared to what would be required in order to meet ecological flow requirements.	Ordinal scale: distribution adapted; distribution partly adapted; distribution not/marginally adapted
	Status of implementation of measures *(level of Action Situation)*	Status of implementation of measures (reduction of water rights; irrigation efficiency measures; use of desalinated water), compared to what has been prescribed in the RBMP.	Ordinal scale: measures implemented; measures partly implemented; measures not/marginally implemented
	RBMP implemented *(level of overarching governance process)*	Status of implementation of measures included in the RBMP which relate to the management of agricultural water consumption.	Ordinal scale: RBMP implemented; RBMP partly implemented; RBMP not/marginally implemented
Environmental outcome performance	Development of agricultural water use	Change in consumptive, as well as total agricultural water use (consumptive and non-consumptive) from 2009 to 2021.	Ordinal scale: reduced agricultural water use; constant agricultural water use; increased agricultural water use
	Development of irrigated area	Change in irrigated surface area from 2009 to 2021.	Ordinal scale: reduced irrigated area; constant irrigated area; increased irrigated area
	Development of status of water bodies	Change in the water status from 2009 to 2021 according to the WFD assessment.	Ordinal scale: status improved; constant status; status deteriorated