

# Estimating the Performance of International Regulatory Regimes

by Tobias Siegfried  
and Thomas Bernauer  
ETH Zurich



# **Estimating the Performance of International Regulatory**

## **Regimes:**

*Methodology and Empirical Application to International Water Management in the  
Naryn / Syr Darya Basin*

Tobias Siegfried and Thomas Bernauer

ETH Zurich

## **Authors:**

Dr. Tobias Siegfried

Earth Institute, Columbia University in the City of New York, 500 W. 120<sup>th</sup> St. Room 842A,  
New York, NY 10027, USA, Phone: +1-212-854-1695, [ts2293@columbia.edu](mailto:ts2293@columbia.edu)

Prof. Thomas Bernauer

Center for Comparative and International Studies (CIS), Seilergraben 49, SEI - G.9, CH-8092  
Zurich, Switzerland, Phone: +41-44-632-6466 (office) or +41-44-632-6771 (direct),  
Fax: +41-44-632-1289, [bernauer@ir.gess.ethz.ch](mailto:bernauer@ir.gess.ethz.ch)

## **Correspondence:**

Prof. Thomas Bernauer

## **Abstract**

We develop a methodology for estimating the performance of international regulatory regimes, building on work by Underdal, Sprinz, Helm, and Hovi. Our performance metric ( $PER^*$ ) relies on assessments, over time, of actual performance, counterfactual performance, and optimal performance. To demonstrate the empirical relevance of this methodology we examine international water management in the Naryn / Syr Darya basin, a major international river system in Central Asia. The emphasis is on the Toktogul reservoir, the main reservoir in the Naryn / Syr Darya basin, and its downstream effects. The biggest policy challenge in this case has been to design and implement international trade-offs among water releases for upstream hydropower-production in winter and water releases for downstream irrigation in summer. We find that the international regime in place since 1998 is characterized by low average performance and high variability. We compare these results with results from a compliance-oriented assessment approach to highlight the analytical problems inherent in the latter.

**Keywords:** International regimes, performance, effectiveness, water management, Syr Darya, Aral Sea, Toktogul dam

## **Acknowledgements**

We are grateful to Andrey Yakovlev, Head of the Department of Operational Hydrology of the Uzbek Hydrometeorological Service, for providing hydrological data that would have been impossible to obtain otherwise. We also thank the Global Runoff Data Center at the Federal Institute of Hydrology (BfG) in Koblenz, Germany, for providing the pre-1991 flow data.

Finally, we are grateful to Christoph Schär, Jean Fried, Detlef Sprinz, Gabi Ruoff, Vally Koubi, Thomas Hinkebein, Nils Petter Gleditsch and Peter Bauer for valuable comments.

# 1. Introduction

Most quantitative political science research on the determinants of international cooperation operates with simple notions of the outcome to be explained – most commonly, the existence of or membership in international agreements, treaties, alliances, or regimes (e.g. (Bernauer 1995; Brochmann and Gleditsch 2006; Neumayer 2002a; Neumayer 2002b)). Substantive assessment of the contents of cooperative arrangements and their performance in terms of solving problems that motivate their establishment is usually left to qualitative case study research. Recent work on the effectiveness of international environmental cooperation suggests that a more sophisticated quantitative approach is feasible (Helm and Sprinz 2000; Underdal 1992; Young 2001). Such an approach could help in systematically and substantively measuring and comparing success or failure in international cooperation over time and across cases. Hence it could provide a more solid foundation for explaining variation in success or failure in this regard, and for assessing whether and when good news about compliance is also good news about cooperation (Downs, Rocke, and Barsoom 1996). Moreover, it would be of practical relevance for policy evaluation.

The existing literature offers only very limited concepts for estimating the performance of international regulatory regimes. Most assessments of performance rely on non-causal criteria. The most common approach is to describe the development of a particular problem targeted by a regulatory regime (e.g. pollution, protectionism) over time and to assess compliance with international obligations in this respect. This is usually done without systematic analysis of

whether and how changes in the outcome and in compliance levels have, *ceteris paribus*, been affected by international cooperation. Economic studies define performance chiefly in terms of efficiency (in a cost-benefit sense) rather than effectiveness in the sense of behavioral effects and problem solving. Policy performance in the local or national context (e.g. (Benneer and Coglianesi 2005)) is sometimes assessed through quasi-experimental research designs and statistical analysis of differences among “treatment” and “non-treatment” groups (e.g., (Greenstone 2004)). Such studies require a wealth of data that often does not exist in the international realm. Moreover, the statistical approach to performance measurement is usually not based on a clear notion of what outcomes would be desirable; neither does it address the counterfactual element inherent in the concept of effectiveness.

In this paper we develop a methodology for estimating the performance (or effectiveness) of international regulatory regimes, building on work by (Helm and Sprinz 2000; Hovi, Sprinz, and Arlid 2003; Sprinz and Helm 2000; Underdal 1992). Our policy performance metric (*PER*<sup>\*</sup>) is a function of the outcome that should ideally be reached (optimum), the performance of a given policy at the time of measurement (actual performance), and the outcome that would have occurred in the absence of this policy (counterfactual performance). The advantages of this measurement concept are: first, it makes explicit reference to optimal performance and thus problem solving; second, it focuses explicitly on the causal relationship between international policies and outcomes; third, it can be used to assess international policy performance at specific points in time in contexts marked by very little data, but also to assess performance dynamics over time in contexts where more data exist; fourth, cooperative efforts can be disaggregated

with reference to particular objectives, policy performance can then be measured for these objectives and aggregated or not.

To demonstrate the empirical relevance of this methodology we examine international water management in the Naryn / Syr Darya basin, a major international river system in Central Asia. The emphasis is on the Toktogul reservoir, the main reservoir in the Naryn / Syr Darya basin, and its downstream effects. The biggest policy challenge in this case has been to design and implement international trade-offs among water releases for upstream hydropower-production in winter and water releases for downstream irrigation in summer. We observe that the international regime in place since 1998 is generally characterized by low average performance and high variability. We compare these results with results from a compliance-oriented assessment approach to highlight the analytical problems inherent in the latter.

The remainder of the paper is organized as follows. In Section 2 we introduce the basic measurement concept, as proposed in previous research, and discuss the problems with this concept. In Section 3 we develop a new concept that solves the problems discussed in the preceding section. In Section 4 we apply the new concept to the Naryn / Syr Darya case. Section 5 concludes.

## **2. Basic Measurement Concept**

The international regimes performance metric as proposed by (Sprinz and Helm 2000; Underdal 1992) is defined as

$$PER_i = \frac{AP - CP}{OP - CP} \quad (1)$$

where  $AP$ : actual performance,  $CP$ : counterfactual performance,  $OP$ : optimal performance<sup>1</sup>. This approach to measuring the performance (i.e., effectiveness) of international regulatory regimes is referred to by the authors as the ‘Oslo-Potsdam Solution’. The subscript  $i$  denotes the  $i$ th criteria with regard to which  $PER$  is estimated. In international water management, for example, such criteria may relate to hydropower production, irrigation water provision, and water quality<sup>2</sup>.

$PER$  can be estimated in relation to any public demand addressed by a public policy. In effect, this equation captures the extent to which a given problem has actually been solved ( $AP - CP$ ) relative to the problem solving potential ( $OP - CP$ ). The first difference alone would only tell us that the relevant policy or regime has had some effect. Only by adding the second difference (and  $OP$  in particular) do we gain information on the extent to which the problem has been solved. Moreover, adding the second difference facilitates comparisons across policies within and across policy-domains, and over time: provided we distinguish between maximizing ( $CP \geq AP \geq OP$ ) and minimizing ( $CP \leq AP \leq OP$ ) cases it sets a lower and upper bound and (with some exceptions) standardizes  $PER$  values between 0 and 1. The limiting behavior of  $PER$  is described in Appendix A.

---

<sup>1</sup> The names of the variables we use differ from the original.

<sup>2</sup> For clarity purposes, the subscript is dropped in the following discussion.



There are several problems with such a performance definition. The first problem stems from the fact that the basic measurement concept is not symmetric around  $OP$  (see also Figure 2 and Appendix A). In other words,  $PER$  is a strictly increasing or decreasing function of  $AP$ , depending on the sign of the difference ( $OP - CP$ ). A simple example demonstrates why this is of relevance. Imagine, for example, that  $PER$  is assessed with regard to public demand coverage. Let us assume that  $OP$  corresponds to freshwater demand of a particular economic sector. If the actual performance of the international water management regime is suboptimal, i.e.  $AP < OP$ , we obtain  $PER < 1$ . However, if too much water is allocated to a particular sector and hence wasted, i.e.  $AP > OP$ , we obtain  $PER > 1$ <sup>3</sup>. This result suggests that wasting resources in allocating ‘too much’ is preferable over the allocation of ‘too little’. Both conditions are clearly undesirable from the point of view of economic efficiency. Similar arguments could be made in regard to policy performance in other areas where policies may over-supply public (or collective) goods.  $PER$  thus fails to provide meaningful results in such situations and its application necessitates an arbitrary scaling of observed values to an ordinal scale (e.g. (Rieckermann et al. 2006)). The latter approach introduces additional uncertainty because of the ad-hoc assignment and scaling of  $AP$ ,  $CP$  and  $OP$  values.

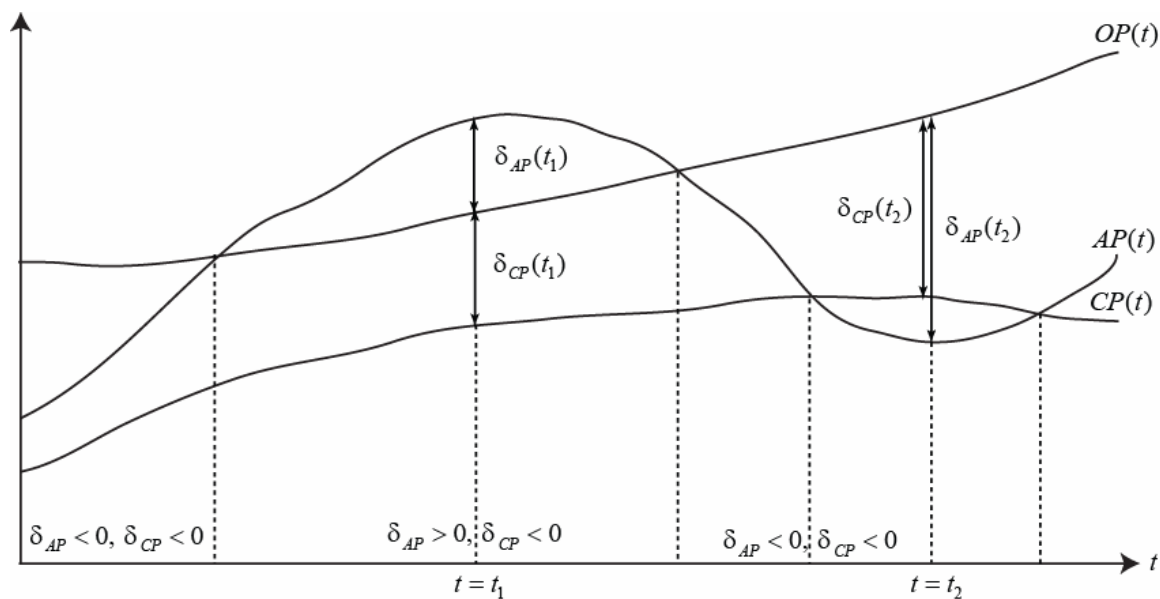
The second problem is that the basic measurement concept may lead to ad hoc integral assessments over time and to wrong conclusions, as shown in Figures 1 and 2. The estimation of  $PER$  at time  $t = t_1$  leads to the value  $b$ , as highlighted in Figure 2. If  $PER$  is assessed at time  $t = t_2$ , performance  $c$  is obtained, which clearly differs from the performance value  $b$ .

---

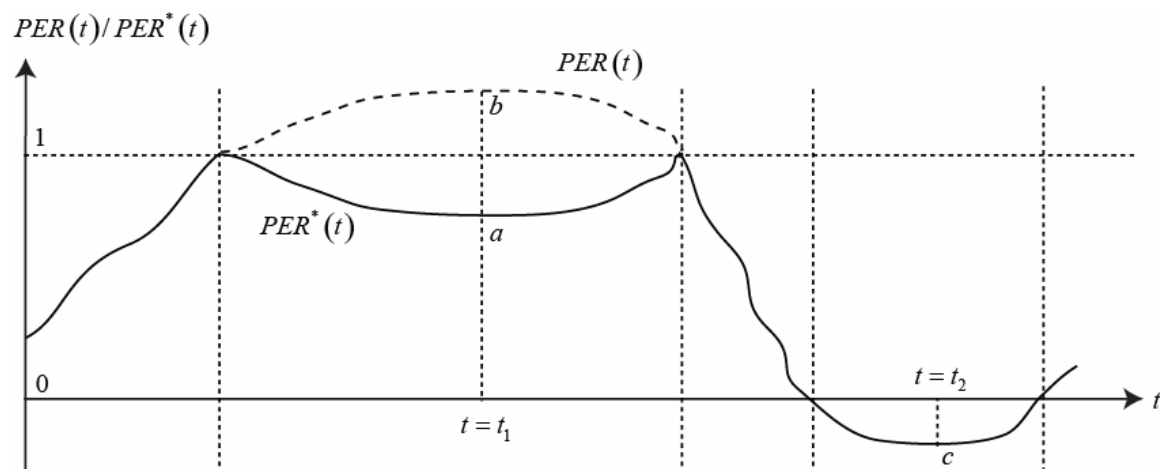
<sup>3</sup> See section 4.2.

Policy performance usually varies in time since public management efforts include time-varying state and demand variables. Imagine, for example, that one tries to assess post-impoundment impacts of a large international dam project over a period of 50 years. Assume, furthermore, that the catchment initially benefits from the hydropower production resulting from the dam project. The negative downstream effects on soil and deltaic systems, however, accumulate in time and gradually appear only after some decades. If performance is viewed as a concept related to demand coverage, initial hydropower demand may have been fully met ( $PER = 1$ ). But subsequent demand coverage in respect to downstream environmental services experiences a dramatic decline. Any assessment of *PER* at a certain time thus provides only a partial picture of performance.

Various scientific disciplines have come up with time dependent measurement concepts. For example, the water engineering literature has defined several performance criteria (e.g. reliability, resilience and vulnerability) that account for time dynamics (Kjeldsen and Rosbjerg 2004). Similarly, climate science has defined several concepts for assessing computational model forecasting quality (Nurmi 2003). The importance of accounting for time dependence is also emphasized by Young (2001) who argues that a static mode of reasoning leads to ad hoc assessments and introduces arbitrariness. We view the lively debate that followed Young's critique as an expression of the need for more research in this field (Hovi, Sprinz, and Underdal 2003; Hovi, Sprinz, and Arlid 2003; Sprinz 2005; Young 2003).



**Figure 1:** Stylized development of  $AP(t)$ ,  $CP(t)$  and  $OP(t)$  over time.  $\delta_{AP}$  and  $\delta_{CP}$  as defined in Equation (3) are shown at different times  $t_1$  and  $t_2$ .



**Figure 2:** Stylized development of  $PER$  and  $PER^*$  as a time dependent function of the stochastic processes as depicted in Figure 1. Note that  $PER(t) > 1$  during a certain time interval, which would lead us to assume falsely that during such wasteful allocation the performance of the respective regulatory regime is highest.

### 3. Upgraded Policy Performance Concept

#### 3.1 Definition

To account for the problems discussed in Section 2, we propose the definition of a new performance measure as given by

$$PER^*(t) = 1 - \frac{|AP(t) - OP(t)|}{|CP(t) - OP(t)|} \quad (2)$$

where  $PER^*(t)$  is a measure of policy performance at a certain time  $t$ .  $PER^*(t)$  measures performance relative to optimal performance  $OP$  at a specific observation time  $t$ . If we use the notation  $\delta_{AP}(t) = |AP(t) - OP(t)|$  and  $\delta_{CP}(t) = |CP(t) - OP(t)|$ , then Equation (2) becomes

$$PER^*(t) = 1 - \frac{\delta_{AP}(t)}{\delta_{CP}(t)} \quad (3)$$

by the definition of the absolute value and its properties. If  $CP(t) < AP(t) < OP(t)$  or  $CP(t) > AP(t) > OP(t)$ , it is easy to see that the two performance measures as defined by Equations (1) and (2) are equal, i.e.  $PER^*(t) = PER$ . Note that  $PER^*(t)$  is symmetric around  $OP(t)$  and that, according to Equation (3),  $PER^*(t)$  is defined as long as  $\delta_{CP}(t) \neq 0$ .

### **3.2 Accounting for Temporal Development and Variation**

Successive observations in times series data are usually not independent of each other. Effectively, each observation for the measured variable is a bivariate observation with time as the second variable. Variation in time can for example be caused by seasonal variation, trends and irregular fluctuations, or a combination of the above. Most series are stochastic in that future values are only partly determined by past time-series values. Simple examples include stochastic rainfall, recharge and run-off processes (for an example, see Figure 4) as well as future per capita and sectoral demand developments.

In our context, we regard the time series  $AP(t)$ ,  $CP(t)$  and  $OP(t)$  (as well as the derived  $\delta_{AP}(t)$  and  $\delta_{CP}(t)$ ) as finite realizations of underlying stochastic processes. In the subsequent analysis, we restrict our focus to stationary processes<sup>4</sup>. The goal is to provide a general and straightforward approach to the characterization of policy performance over a certain period of time by making use of basic concepts and definitions of probability theory and statistics. This approach assumes neither knowledge of the underlying probability distribution functions, nor of the stochastic processes that eventually produce  $AP(t)$ ,  $CP(t)$  and  $OP(t)$ . We submit that the expected value as well as the variance of  $PER^*(t)$  are two descriptions that permit a useful characterization of regime performance over time.

---

<sup>4</sup> A process is stationary if the properties of the underlying model do not change. Precipitation patterns need not be particular realizations of stationary processes since, for example, climate change can affect the underlying model. However, the time horizon for performance assessment is short compared to such model changes and is therefore neglected.

As shown in Appendix B, we can approximate the expected value of  $PER^*(t)$  by

$$\langle PER^* \rangle = 1 - \frac{\mu_{\delta_{AP}}}{\mu_{\delta_{CP}}} + \frac{1}{\mu_{\delta_{CP}}^2} \text{Cov}(\delta_{AP}, \delta_{CP}) \quad (4)$$

where  $\text{Cov}(\delta_{AP}, \delta_{CP})$  denotes the covariance and  $\mu_{\delta_{AP}}$  as well as  $\mu_{\delta_{CP}}$  the mean of the time series  $\delta_{AP}(t)$  and  $\delta_{CP}(t)$ . Taking the covariance into account is relevant in many cases. Imagine for example pre- and post-impoundment run-off in a river. Depending on the management of the constructed dam, pre- and post- flow regimes are still correlated to variable degrees<sup>5</sup>. The magnitude of such covariance depends on the variances  $\sigma_{\delta_{AP}}^2$  and  $\sigma_{\delta_{CP}}^2$  of  $\delta_{AP}(t)$  and  $\delta_{CP}(t)$ . If the latter are entirely uncorrelated, then  $\text{Cov}(\delta_{AP}, \delta_{CP}) = 0$ .

As shown in Appendix C, the variance of  $PER^*(t)$  is approximated by

$$\sigma_{PER^*}^2 = \frac{4\sigma_{\delta_{AP}}^2}{\mu_{\delta_{CP}}^2} - \frac{\mu_{\delta_{AP}}^2 \sigma_{\delta_{CP}}^2}{\mu_{\delta_{CP}}^4} - \frac{\text{Cov}(\delta_{AP}, \delta_{CP})^2}{\mu_{\delta_{CP}}^4} - \frac{2 \text{Cov}(\delta_{AP}, \delta_{CP}) \mu_{\delta_{AP}}}{\mu_{\delta_{CP}}^3} \quad (5)$$

In Equations (4) and (5),  $\mu_{\delta_{AP}}$ ,  $\mu_{\delta_{CP}}$ ,  $\sigma_{\delta_{AP}}^2$ ,  $\sigma_{\delta_{CP}}^2$  and  $\text{Cov}(\delta_{AP}, \delta_{CP})$  have to be empirically estimated from available data (see for example (Loucks, Stedinger, and Haith 1981) for a detailed explanation of the standard estimation procedure).

---

<sup>5</sup> See Section 4 for a real world example of pre- and post-impoundment flow correlation.

(Young 2001) states that procedures involving counterfactual analysis to assess international regime effectiveness have rarely been applied in a transparent and systematic fashion. According to him, they have relied too much on subjective judgments in scoring individual cases based on simplistic categories. We submit that the upgraded measurement concept presented above addresses the most important shortcomings of the approach proposed by (Sprinz and Helm 2000). In the remainder of this paper, we demonstrate the empirical relevance of the concept with a case study on international water management.

## **4. Application to International Water Management**

We begin with a description of the case to be studied: the Naryn / Syr Darya river basin in Central Asia, and the Toktogul reservoir in particular. We then estimate international regulatory regime performance in this case.

### ***4.1 Naryn / Syr Darya Basin and Toktogul Reservoir***

The Naryn / Syr Darya river system is part of the Aral Sea basin; the other main river of this basin is the Amu Darya. The size of the Aral Sea basin is approx. 1.55 million km<sup>2</sup>, its population around 40 million. The economies of the riparian countries (Kazakhstan, Kyrgyzstan, Uzbekistan, Tajikistan, Turkmenistan) are heavily dependent on irrigated agriculture (with shares of 40 – 50 % of GDP in 1960 – 1990, and around 20 – 30 % thereafter). Farming employs ca.

60 % of the rural population and 25 – 60 % of the total labor force (World Bank 1996). Most water for irrigation is abstracted from the two Daryas. While upstream parts of the basin are mostly mountainous and humid, the mid- and downstream areas are arid (low frequency and irregular, high intensity precipitation with large daily and seasonal temperature differences). Over the past 40 years, excessive water withdrawals have led to a drastic shrinkage of the Aral Sea; the latter receives the bulk of its water from the two Daryas. The Aral Sea has thus been reduced to around 25 % of its original volume and has received worldwide attention as an ecological disaster zone (Dukhovny and Sokolov 2005).

The Syr Darya river originates as the Naryn river in the mountains of Kyrgyzstan (see Figure 3). It then flows through Uzbekistan and Tajikistan and ends in the Aral Sea in Kazakhstan (total length around 2'800 km). In total, approximately 20 million people inhabit this river catchment, which covers an area of ca. 250'000 km<sup>2</sup>. The river is mainly fed by snowmelt and water from glaciers. The natural run-off pattern, with annual flow ranges of 23.5 – 51 km<sup>3</sup> (around 40 km<sup>3</sup> in the past few years) is characterized by a spring / summer flood. It usually starts in April and peaks in June. Nowadays, around 93% of the Syr Darya's mean annual flow is regulated by storage reservoirs. Approximately 75% of the run-off stems from Kyrgyzstan (Dukhovny and Sokolov 2005). Water abstraction from the Syr Darya basin is mainly for irrigated farming. Of the approx. 3.4 million ha of irrigated farm land around 1.7 million ha is irrigated with water taken directly from the river. Figure 4 shows the time series of the Naryn / Syr Darya river flow over the last 72 years as measured at Uch Kurgan gauge station, Uzbekistan (see Figure 3 for the location of the latter).



The run-off of the Naryn / Syr Darya, as measured at Uch Kurgan gauge station, i.e. at the foot of the Naryn / Syr Darya cascade shortly after the river enters Uzbekistan from Kyrgyzstan, varies strongly over time. It is marked by four distinct periods as shown in Figure 4. During the phase of largely natural run-off (1933–1974), mean flow was  $388 \text{ m}^3/\text{s}$ , with a high variability in summer (see Figure 7 for mean monthly flows as well as Table 4 in Annex D for data on flow variance). In this period, the pronounced differences in flows are entirely determined by seasonal and climatic variability.



**Figure 3:** This map shows the part of the Naryn and Syr Darya catchment that is of most interest in this paper. The Uch Kurgan gauge station is located in the center of the map. The Toktogul reservoir is located at the top of the Naryn / Syr Darya cascade in Kyrgyzstan.

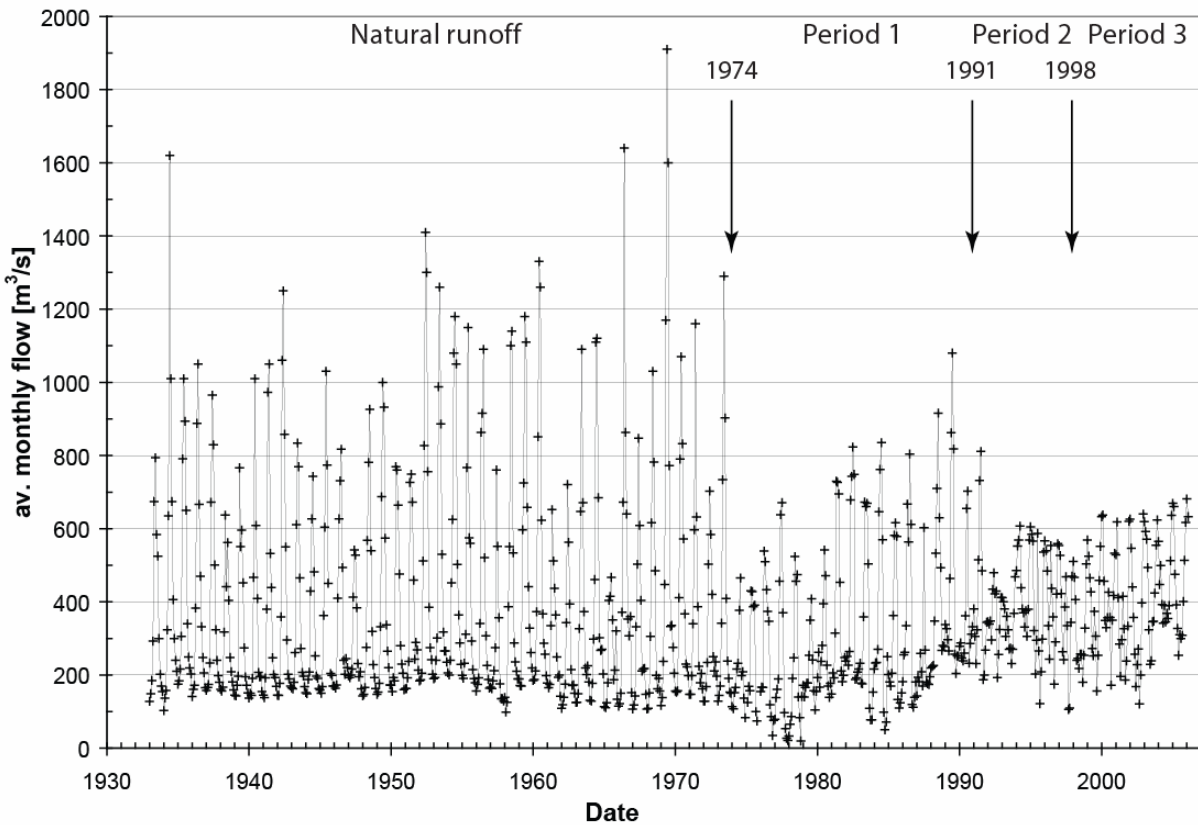
A substantial change in flow patterns occurred with the commissioning of the Toktogul dam in 1974<sup>6</sup>. This event marks the beginning of the first river management period (1974 – 1990) in our analysis. The latter was characterized by centralized management of the Toktogul reservoir and the river basin as a whole. The Toktogul dam is by far the largest storage facility in the Aral Sea basin. It has 14 km<sup>3</sup> effective capacity, 8.7 km<sup>3</sup> firm yield and a full capacity of ca. 19.5 km<sup>3</sup> (Figure 6 shows storage volumes in 1974 – 2006). The reservoir area is around 280 km<sup>2</sup>, its length around 65 km.<sup>7</sup> Hydropower capacity of the Toktogul power plant is 1'200 MW, i.e. the second biggest in the Aral Sea basin (Antipova et al. 2002). After the commissioning of the dam, a general attenuation of peak downstream flows was observed (see Figure 4). Furthermore, an overall decline of monthly flow variability occurred. This decline was most pronounced in the summer months.

During this first period, the management system was oriented primarily towards adequate water provision for irrigated agriculture (above all, cotton production) in Uzbekistan and Kazakhstan. The timing of winter and summer flow releases did not change substantially compared to the natural runoff pattern. This is indicated by seasonal ratios  $r$  of inflow vs. outflow that oscillate around  $r = 1$  (see inflow/outflow ratios in Figure 5 in 1980 – 1990).

---

<sup>6</sup> Some smaller reservoirs downstream of Uch Kurgan, notably the Kairakkum and Chardara reservoirs, had been put in place earlier.

<sup>7</sup> In the whole Naryn / Syr Darya basin, total usable reservoir capacity is around 27 km<sup>3</sup>.



**Figure 4:** Mean monthly flow of Naryn / Syr Darya river at Uch Kurgan gauge from January 1933 to February 2006. The four flow patterns – pre-Toktogul (1933 – 1974), USSR Naryn–Syr Darya management (Period 1: 1974 – 1990), post-USSR operation (Period 2: 1991 – 1997), and new Inter-State Commission for Water Coordination (ICWC) regime (Period 3: 1998 – today) are clearly visible in the time-series. Data Sources: Global Runoff Data Center (GRDC) and Andrey Yakovlev, Uzbek Hydrometeorological Service.

In the early 1980s, two basin water organizations (BWO) were added to this system; the one for the Naryn / Syr Darya was set up in Tashkent, Uzbekistan. Their mandate was to operate and maintain all head water structures with a discharge of more than 10 m<sup>3</sup>/s. This management

system and its infrastructure was fully funded from the federal budget of the USSR. In consultation with the governments of the five republics and based on forecasts by the Central Asia Hydromet Service, the ministry of water resources (Minvodgoz) in Moscow defined annually (based on a multi-year master plan for each river system) how much water was to be released for irrigation during the growing season (April to September) to each water management region.

The BWOs were responsible for implementing the water allocations and maintaining the infrastructure. They also had the authority to increase or reduce allocations to each republic by up to 10%. The electricity produced at Toktogul during that period went into the Central Asian Energy Pool (CAEP) and was thus shared among the riparian republics. In exchange, the neighboring republics supplied coal, oil, and natural gas to Kyrgyzstan in winter to cover increased Kyrgyz energy demand during the colder months. The fossil fuel was used primarily in thermal power plants in Bishkek and Osh. (Cai, McKinney, and Lasdon 2002).

The collapse of the Soviet Union in 1991 led to the breakdown of centralized water resources management and water-energy trade-off arrangements, causing serious disputes between the newly independent states over water allocation issues. With these events, the second river management period commenced. Coal, oil, natural gas, and electricity supplies to Kyrgyzstan declined dramatically between 1991 and 1997, and so did the thermal and electric power output of Kyrgyz thermal power plants (TPP).<sup>8</sup> Consumers turned to electricity, which increased winter demand by more than 100%. Purchases of energy from abroad were (and still are) difficult

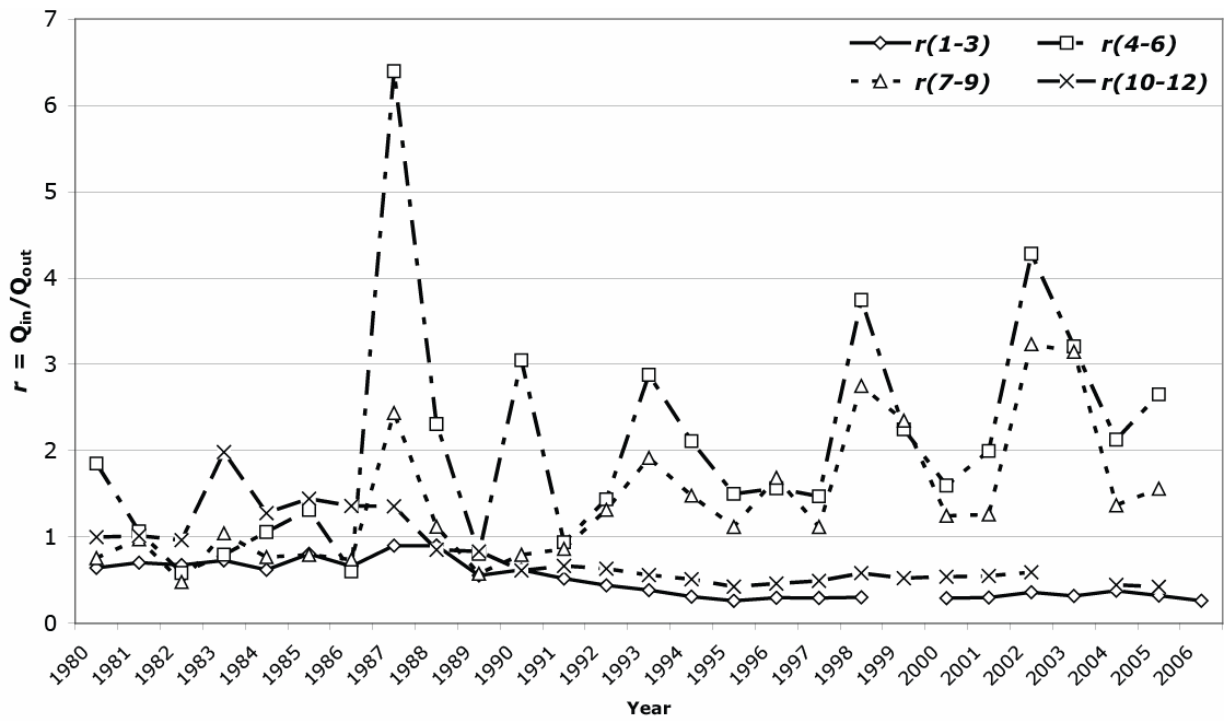
---

<sup>8</sup> Thermal power output from Kyrgyz TPPs in 1991 - 1997 declined from  $5.8 \cdot 10^6$  Gcal to  $2.8 \cdot 10^6$ , electric power output from 3.9 to 1.6 M kWh.

because the government was (for political and administrative reasons) unable to raise and collect appropriate energy tariffs. Moreover, financial contributions from Moscow and the former republics in the basin for the maintenance of the reservoir ceased. In response to the sharp drop in thermal power output and rising winter demand for electricity, Kyrgyzstan switched the operation of the Toktogul reservoir from irrigation to electric power production mode. Since winter 1993, the flow peaks no longer occur in summer but rather in winter (see Figure 6). This has led to the opening up of a gap between the summer inflow/outflow ratios  $r$  and their winter counterparts (see Figure 5).

The main problem is that upstream interests deriving from temporal water demands are diametrical to downstream water demands and interests. Kyrgyzstan uses very little water consumptively, i.e. for irrigation. But it is interested in producing hydro-electricity at the Toktogul electric power plant, particularly in winter when energy demand is higher (Kyrgyzstan has no fossil fuel sources of its own). This interest has become ever stronger as the downstream countries have cut back on energy supplies to Kyrgyzstan. Kyrgyzstan also views electricity production as a potential export commodity. It is thus eager to store water in spring to autumn and release it in winter to spring for energy production. Conversely, downstream Uzbekistan and Kazakhstan, by far the largest consumers of irrigation water in the river basin, are interested in obtaining much more water during the growing season (April to September) than in the non-growing season (October to March). They are also interested in electricity produced upstream through water release during the growing season for operating irrigation pumps. Moreover, from the perspective of downstream countries, water releases in winter should be rather low, for high flows may cause floods because ice in the river bed reduces water flow capacity (Savoskul et al.

2003). The principal problem to be solved thus pertains to coordinating the management of the Naryn / Syr Darya cascade of reservoirs that are located entirely in Kyrgyzstan, and in particular the handling of trade-offs between consumptive water use for downstream irrigation purposes and non-consumptive use for upstream energy production in Kyrgyzstan.



**Figure 5:** The figure shows the ratio  $r$  of inflow to outflow averaged over 3 months. The switch from cooperative to non-cooperative water resources management is characterized by the opening up of a gap between winter (months 1-3 and 10-12) and summer (months 4-6 and 7-9) inflow/outflow ratios from 1991 onwards. Note that the pronounced peaks in the summer month ratios characterize years of above average summer runoff (compare with inflow data in Figure 6).

As a consequence of the non-cooperative management during the second period, the high winter spills from the river have damaged infrastructure and land resources in downstream Uzbekistan. Additionally, they have reduced the potential for water releases for irrigation during the vegetation period. Ever since 1991, the riparian countries have been struggling to re-establish an effective management scheme (Savoskul et al. 2003).

However, during that period, international talks focusing on the management of the Toktogul reservoir continued. In February 1992 the five newly independent states set up the Inter-State Commission for Water Coordination (ICWC). This Commission has four bodies: its secretariat, the two BWOs for the Aral Sea basin, and the Scientific Information Center. In 1993, the International Fund for Saving the Aral Sea was added to the ICWC<sup>9</sup>. The five countries agreed to keep the water allocation principles of the former USSR system in place until a new system could be established, albeit without the funding for the infrastructure that had formerly come from Moscow. The most important hydraulic structures, and in particular the biggest reservoirs in the basin (including the Toktogul), were not put under the control of the BWOs (i.e. they were de facto nationalized by the newly independent countries and mostly transferred to their national energy agencies). As a consequence, the BWOs lost much of their authority and operational capacity.

Several declarations by the riparian countries and attempts by European and North-American government agencies to help in the problem-solving effort produced only minimal progress. In 1995, for example, sponsored by the European Union, a water resources management

---

<sup>9</sup> <http://www.icwc-aral.uz/>

information system and a water use and farm management system were set up. Only in March 1998, under the aegis of the Executive Committee of the Central Asian Economic Community and assisted by USAID, Kazakhstan, Kyrgyzstan, and Uzbekistan signed a formal agreement that marks the beginning of management Period 3. In 1999 Tajikistan joined this agreement.<sup>10</sup> The agreed release schedule is shown in Table 5, Appendix D.

The 1998 agreement (consisting of two separate treaties) is set up as a general framework agreement plus specific barter agreement on energy-water exchanges in 1998. The specific agreement holds that in the growing season (April 1 – October 1), Kyrgyzstan agrees to supply 2.2 M kWh of electricity to Kazakhstan and Uzbekistan (1.1M kWh each). Kazakhstan and Uzbekistan, in turn, agree to deliver specific amounts of electricity, natural gas, fuel oil, and coal to Kyrgyzstan in specific months under conditions set forth in bilateral agreements concluded already in 1997. Compensation can also be carried out in the form of “other products” (labor and services are mentioned) or money. Possible adjustments to the barter deal can be performed by the BWO Syr Darya and UDC Energia in agreement with the interested countries. Kyrgyzstan agreed to cut its energy consumption by 10% against 1997 levels. The framework agreement, also concluded in March 1998<sup>11</sup>, holds that these exchanges will subsequently be defined annually through negotiations. It installs the BWO Syr Darya and UDC Energia as the implementing agencies for the release schedules and energy transfers, pending the establishment of a new International Water and Energy Consortium.

---

<sup>10</sup> <http://ocid.nacse.org/cgi-bin/qml/tfdd/treaties.qml>

<sup>11</sup> <http://ocid.nacse.org/cgi-bin/qml/tfdd/treaties.qml>

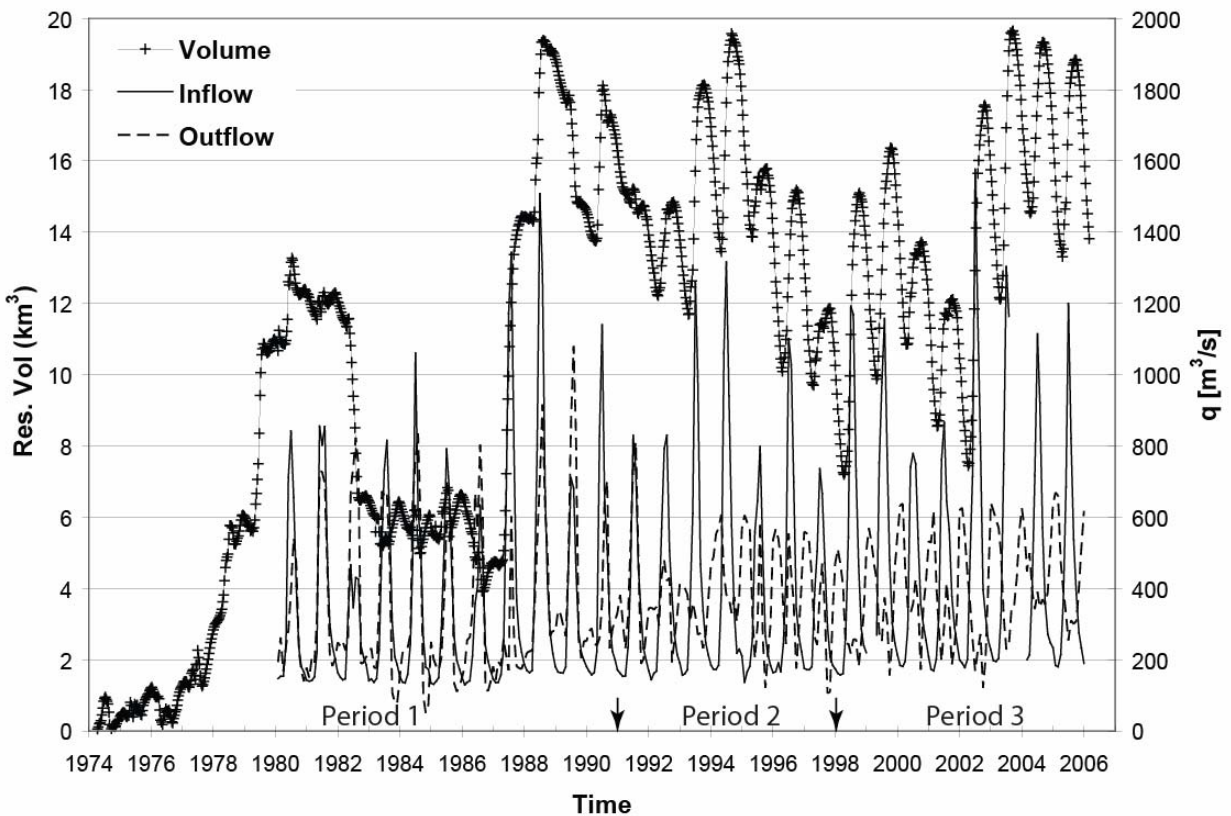


In other words, the water management system put in place in 1998 holds that during the vegetation period Kyrgyzstan releases more water than it needs for its own hydro-power demands, and that the energy surplus is distributed to Kazakhstan and Uzbekistan. In the non-growing period (October 1 – April 1) Uzbekistan and Kazakhstan supply Kyrgyzstan with energy resources in amounts that are approximately equivalent to the electricity they receive from Kyrgyzstan during the growing season. The exact amounts of water and energy are defined annually through negotiations among the countries. Typically, Kyrgyzstan has been scheduled to release around 6.5 km<sup>3</sup> of water during the vegetation period and transfer around 2.2 M kWh of electricity to Uzbekistan and Kazakhstan.

In the subsequent section and based on the methodology developed in Section 3, we now assess the performance of the international water management system introduced in 1998.

## 4.2 Assessment of Performance

The three management periods (Period 1 – 3) that are of interest to our assessment can be seen in Figure 6. These periods are characterized by differing flow patterns that are associated with the timeline of political events (see above).

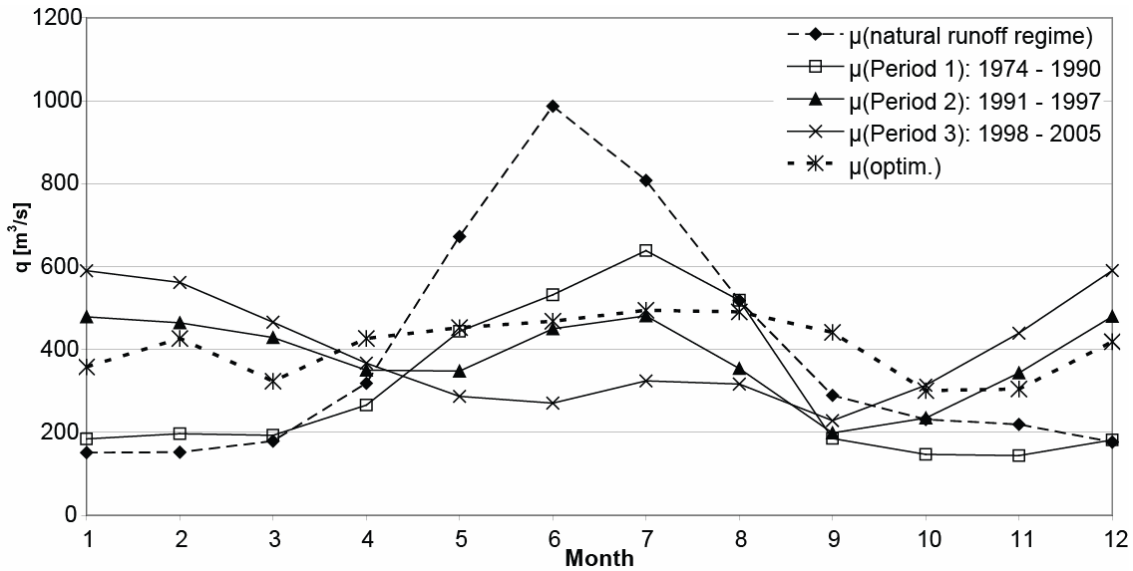


**Figure 6:** In- and outflow of the Toktogul reservoir and reservoir volume since dam closure. Note that the full reservoir capacity was not reached until the beginning of high inflow years (starting in July 1987). The switch from irrigation to power production mode is clearly visible after 1992 (see also Figure 5). Reservoir in- and outflows are only shown from 1980. Data Source: Andrey Yakovlev.

In the period of centralized water resources management under USSR rule (Period 1, 1974 - 1990), mean flow was reduced to 311 m<sup>3</sup>/s mainly due to the filling of the Toktogul reservoir<sup>12</sup>. The characteristics of the yearly averages do not differ substantially from the natural flow, with a summer discharge peak and winter low flow. Yet, due to the filling of the reservoir, the summer peak is less pronounced. This characteristic flow pattern changes after the breakdown of central governance as can be seen by looking at the curve  $\mu$ (Period 3) in Figure 7. As discussed above, the increased hydropower demand in upstream Kyrgyzstan led to a pronounced increase of reservoir water releases in the winter months. The somewhat reduced monthly variability in flow (see  $\sigma$ (Period 3) in Appendix D, Table 4) characterizes the unilateral upstream management of the Syr Darya run-off. Finally, with the implementation of the 1998 agreement in Period 3, monthly flows appear to reflect the trade-offs made in that agreement. Average flow is 396 m<sup>3</sup>/s, with a considerable decline in monthly variability compared to the prior period.

---

<sup>12</sup> If we assume an average of 14 km<sup>3</sup> dam storage volume to be filled at a rate of 70 m<sup>3</sup>/s (which is the difference in mean flow between the undisturbed regime and management period 1), we obtain an approximate filling time of 6.3 years as observed in Figure 6.



**Figure 7:** Monthly long-term average flows at the Uch Kurgan gauge (based on data from GRDC and Andrey Yakovlev). The data on flow variability for the corresponding months and periods is provided in Appendix D, Table 4. The monthly data  $\mu(\text{optim.})$  are calculated optimal releases from the Naryn / Syr Darya cascade. Optimization was carried out with a coupled hydrologic-agronomic-economic model on the basin scale by (Cai, McKinney, and Lasdon 2003).

For the performance assessment, we start with the assumption that centralized management in Soviet times (Period 1, 1980-1990<sup>13</sup>) was optimal ( $OP_S(t)$ ) because up- and downstream interests were successfully addressed through an integrated water-energy exchange system. From the perspective of the long-term Aral Sea problem and local economic and environmental interests

<sup>13</sup> We start with 1980 because 1974-1979 were years of reservoir filling (see fn 12 and Figure 6). We also expect a strong trend effect in reservoir outflows during the latter years - this would introduce errors in our performance estimation.

there, Period 1 was certainly not optimal<sup>14</sup>. We thus employ a second notion of optimality,  $OP_C(t)$ , which emphasizes sustainability of natural resources management on the basin scale (Cai, McKinney, and Lasdon 2003; McKinney, Cai, and Lasdon 1999).  $\mu(\text{optim.})$  in Figure 6 is not observed but is the result of a simulation-optimization approach that we denote as  $OP_C(t)$  (see also Table 4, Appendix D). We use the following notation to distinguish the scaling of  $PER^*(t)$ :  $PER^*(t)|_{OP_S}$  is calculated with respect to  $OP_S(t)$  and  $PER^*(t)|_{OP_C}$  with respect to  $OP_C(t)$ .

The period of breakdown of the centralized management system in 1991–1997 (Period 2), i.e. the period where there is no international agreement, is defined as counterfactual performance, i.e.  $CP(t)$ .<sup>15</sup> The current flow regime (Period 3, 1998-today) is defined as actual performance  $AP(t)$ .

To compute the performance  $PER^*(t)$  of the international regime installed in 1998 we use monthly averaged flow values for  $OP_S(t)$  and  $CP(t)$  (see Table 1). This is necessary for two reasons. First, we cannot compare individual hydrological years with differing resource

---

<sup>14</sup> Young (2001) argues that definitions of the optimum with reference to which performance is assessed must not necessarily be based on objective notions, but can depend on understandings of the nature of the problem and the options available for solving the problem.

<sup>15</sup> Another approach to measuring CP could be to assume unconstrained (by the downstream countries or actors from outside the basin) maximization by Kyrgyzstan of hydropower production to cover domestic energy needs and export excess energy to obtain foreign currency, and to carry out a simulation-optimization (from the Kyrgyz perspective) on that basis. Discussions with experts on the region led us to the conclusion that such a scenario would have been very unlikely, and that the scenario of CP in terms of Kyrgyz behavior along the lines observed in 1991-97 would have been more likely.

endowments (i.e. inflow as well as reservoir levels) and demand (electricity as well as irrigation water). Doing so would lead to an arbitrary comparison of reservoir outflows between years that are not necessarily comparable with respect to the above mentioned state variables. Second, the individual periods have different lengths. Hence, they cannot be compared directly.

Month	1	2	3	4	5	6	7	8	9	10	11	12
$\mu(OP_S(t))$ [ $m^3/s$ ]	245	251	231	281	457	593	790	583	213	176	203	248
$\mu(CP(t))$ [ $m^3/s$ ]	497	487	454	338	310	377	402	322	197	242	352	502

**Table 1:** Mean monthly flows for Period 1 (1980-1990,  $\mu(OP_S(t))$ ), and Period 2 (1991-1997),  $\mu(CP(t))$ .

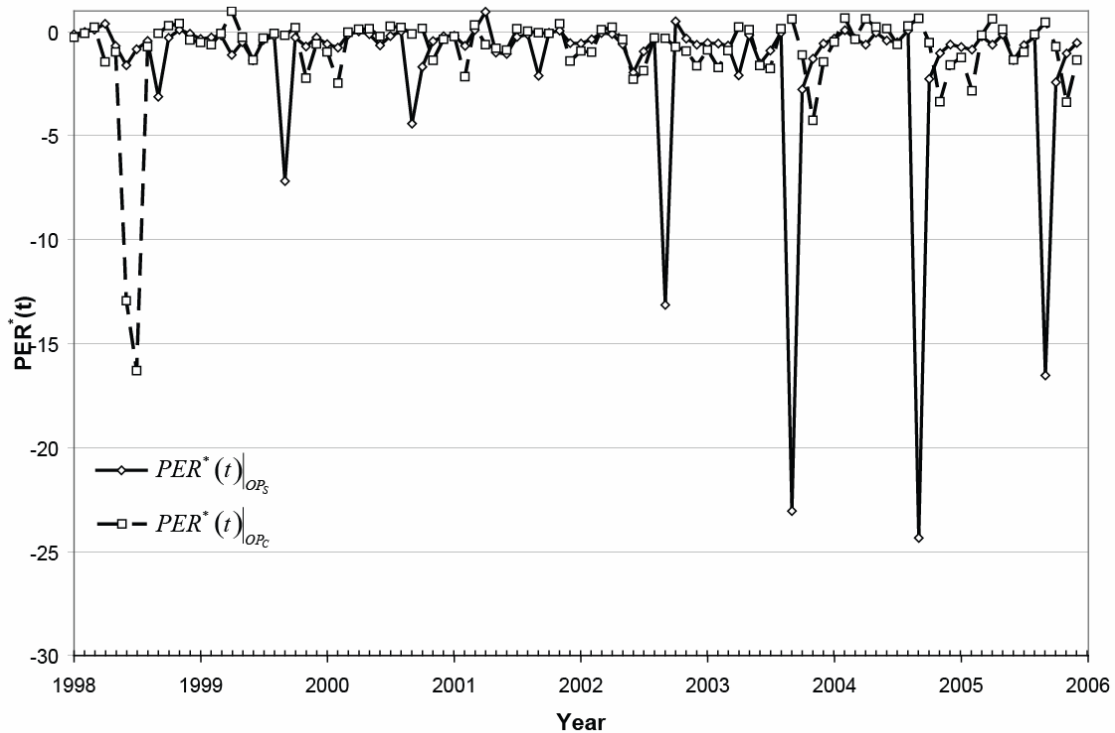
Note that for  $\mu(OP_S(t))$ , the mean values do not correspond to the ones shown in Figure 7 since we do not take into account the initial years of reservoir filling (1974-1979).  $\mu(OP_C(t))$  can be found in Table 4, Annex D.

The calculation of  $PER^*(t)$  based on  $OP_S(t)$  may be problematic. The underlying assumption is that demand for irrigation water and hydroelectric power has not changed in 1980-2006. This assumption is to some extent violated by the fact that in Uzbekistan, for example, the irrigated area grew from  $3.5 \cdot 10^6$  ha<sup>16</sup> in 1980 to  $4.4 \cdot 10^6$  ha in 1998, i.e. more than 25 %. Demand in Kyrgyzstan for hydroelectric power has grown substantially as well.<sup>17</sup> We could have addressed this problem by scaling  $\mu(OP_S(t))$  according to changes in demand for irrigation water and hydroelectric power. We did not do so because there would be a high degree of arbitrariness in such approach. In particular, the very notion of optimality may loose sense after such scaling

<sup>16</sup> Data from <http://www.fao.org/ag/AGL/aglw/aquastat/countries/uzbekistan/index.stm>.

<sup>17</sup> A similar argument applies to  $\mu(CP(t))$ .

since the latter, for example, does not take into account inter-seasonal shifts of optimal water allocation. In other words, optimal allocation is not a linear function of the quantity of water available. Note that such problem does not apply to  $PER^*(t)|_{OP_c}$  because this measure of OP reflects recent up- and downstream demand constraints.



**Figure 8:** Development of  $PER^*(t)$  during Period 3 with respect to the two notions of optimality.

Generally, low performance with a declining trend over time is observed.

The temporal development of  $PER^*(t)|_{OP_S}$  and  $PER^*(t)|_{OP_C}$  is shown in Figure 8. With respect to both notions of optimality, performance of the 1998 regime has been poor. In particular, the

figure shows that extremely negative values of  $PER^*(t)|_{OP_S}$  start to occur from 2002 onwards, usually in September. This can be explained by the fact that in this month,  $|\mu(CP(t)) - \mu(OP_S(t))|$ , i.e., the denominator of  $PER^*(t)|_{OP_S}$  is small and the difference between actual performance and the monthly averaged performance of Period 1, i.e.  $|AP(t) - \mu(OP_S(t))|$ , is large<sup>18</sup>.

To calculate  $\langle PER^* \rangle$  and  $\sigma_{PER^*}^2$ , we need to estimate the sample means  $\hat{\mu}_{\delta_{AP}}|_{(\bullet)}$ ,  $\hat{\mu}_{\delta_{CP}}|_{(\bullet)}$ , the variances  $\hat{\sigma}_{\delta_{AP}}^2|_{(\bullet)}$ ,  $\hat{\sigma}_{\delta_{CP}}^2|_{(\bullet)}$  and the covariances  $\widehat{Cov}(\delta_{AP}, \delta_{CP})|_{(\bullet)}$ .  $(\bullet)$  is a placeholder for  $OP_S(t)$  and  $OP_C(t)$ . The estimated values for the mean and variance are shown in Table 2. For the covariances, we obtain  $\widehat{Cov}(\delta_{AP}, \delta_{CP})|_S = 9640.5 \text{ m}^6/\text{s}^2$  with respect to  $OP_S(t)$  and  $\widehat{Cov}(\delta_{AP}, \delta_{CP})|_C = 1250.1 \text{ m}^6/\text{s}^2$  with respect to  $OP_C(t)$ .

	$\delta_{AP}^{Abs}(S)$	$\delta_{CP}^{Abs}(S)$	$\delta_{AP}^{Abs}(C)$	$\delta_{CP}^{Abs}(C)$	unit
$\mu$	259.9	171.3	157.7	85.2	$[\text{m}^3/\text{s}]$
$\sigma^2$	15314.4	6932.9	5900.2	3755.7	$[\text{m}^6/\text{s}^2]$

<sup>18</sup> The *PER* measure suggests performance values of  $PER = \{-13.7, 26.6, 28.0, 19.7\}$  for September 2002 - September 2005. The positive *PER* values are clearly nonsensical and thus indicate the inherent problem associated with the utilization of *PER* as proposed by Helm and Sprinz (2000) (see also section 2).



**Table 2:** Estimated sample mean and variance. The times series  $AP(t)$  and  $OP_S(t)$  have been truncated to 7 years for the sample estimations of the mean, variance and covariance values<sup>19</sup>.

We now calculate the regime performance and its variance. The results are shown in Table 3.

	$\langle PER^* \rangle$	$\sigma_{PER^*}^2$
$OP_S$	-0.24	0.64
$OP_C$	-0.71	0.92

**Table 3:** Average regime performance and variance with reference to  $OP_S$  and  $OP_C$ . The calculations are based on Equations (4) and (5) and Table 2.

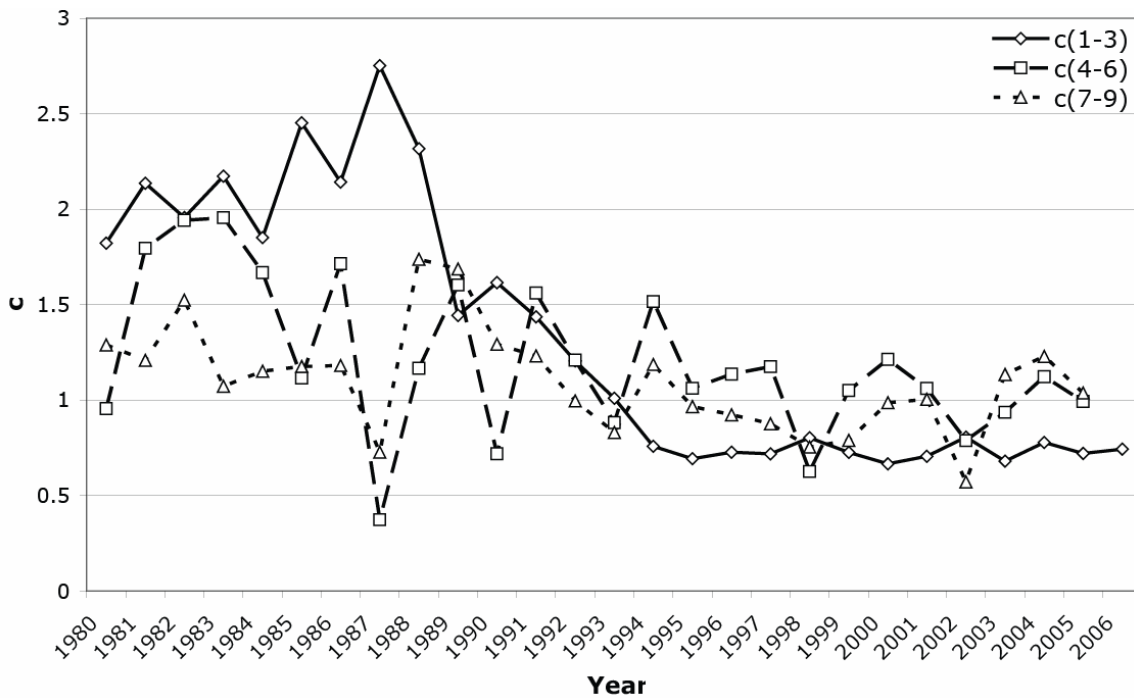
As concluded already from visual inspection of  $PER^*(t)$  in Figure 8, overall performance of the 1998 regime is poor indeed.

Finally, we compare the above results with the results from a compliance-oriented assessment approach. To this end, we compute ratios of actual water releases from the Toktogul reservoir (three month averages) and the targets for the respective months as defined in the 1998 agreement. Figure 9 shows that this assessment produces much more positive performance scores than the more sophisticated estimation.

---

<sup>19</sup>  $OP_C$  as given in (Cai, McKinney et al. 2003) is provided as monthly averaged series of values. In the calculations based on this computed optimum, we assume that the monthly values of  $OP_C$  do not change over the period of assessment.

This finding exposes an important analytical problem associated with the compliance-oriented assessment approach. As noted by Downs, Rocke, and Barsoom (1996), generally good compliance in international regulatory regimes can be misleading because of endogeneity and selection problems. They note that states often define treaty commitments that require little or no effort beyond what the states concerned would do in the absence of the respective treaty. The empirical application of our measurement concept, which uses *OP* and *CP* rather than the 1998 treaty targets as benchmarks, demonstrates that good news about compliance is in the Syr Darya case is certainly not good news about cooperation.



**Figure 9:** Compliance  $c$  in months 1-3 is defined as the 1998 target divided by the actual water release; compliance in months 4-6 and 7-9 is defined as the actual water release divided by the 1998 target. These definitions are based on the assumption that exceeding the target in the growing season is better for

downstream countries than exceeding the target in winter. We also show the results for the years before 1998 to provide an idea of the general trend, though the 1998 agreement was, of course, not in force before 1998. The average compliance scores (1 = perfect compliance) were 1.6 in 1980-90, 1.1 in 1991-97, and 0.9 in 1998-2006.

## 5. Conclusion

The methodology proposed in this paper addresses several deficiencies in extant concepts for estimating the performance of international regulatory regimes. Notably, it deals in a transparent and tractable way with the fact that actual performance, optimal performance, and counterfactual performance are time dependent variables that relate to particular realizations of underlying stochastic processes. To the extent times-series data of reasonable quality for policy outcomes is available, our methodology can be applied to virtually any international (and also national or local) policy or regulatory regime to study its performance. If data is more limited, the methodology also lends itself to snapshot assessments of performance at particular points in time, and to assessments based on ordinal scaled data – e.g. data obtained through expert interviews based on the Delphi method or other approaches.

To demonstrate the empirical relevance of the methodology, we carried out a performance assessment of the international regulatory regime for the Naryn / Syr Darya river basin (with a focus on the Toktogul reservoir in Kyrgyzstan). The results show that this regime is characterized by low average performance and high variability (see  $\langle PER^* \rangle$  and  $\sigma_{PER^*}^2$  in Table 3).

A comparison of these results with results from a conventional compliance assessment revealed that the more sophisticated method produced much more negative performance estimates. Thus it highlights the analytical problems (selection effects and endogeneity) with compliance- or policy-output-oriented performance estimations. Recent work by Brochmann and Gleditsch (2006), for example, finds that cooperation in the form of international river management treaties is more likely in upstream-downstream settings than in other settings. This result is indeed surprising because it suggests that upstream-downstream asymmetries can be overcome through Coasian deals (i.e., side-payments, issue-linkages) at reasonably low transactions costs. Yet, the Syr Darya case, a clear upstream-downstream case where we observe a treaty and good compliance with its obligations, shows that such an assumption is probably too optimistic.

Our empirical findings also have important policy-implications. They suggest that the riparian countries of the Syr Darya (and stakeholders from outside the river basin) should not be misled by good compliance. Our results show that the institutional solution to the problem, as put in place in 1998, is performing very poorly. Conflicts over water allocation among the riparian countries have in the past few years been muted by high levels of precipitation upstream. As soon as an extended period of low precipitation sets in (due to climate change or for other reasons) the conflict is likely to heat up. The obvious recommendation is to repair the regime before this happens.

## Appendices

### ***Appendix A – Limiting Behavior of PER and PER\****

If the performance of an international regulatory regime is optimal, i.e.  $AP = OP$ , the limiting behavior of  $PER$  is given by

$$\lim_{AP \rightarrow OP} \frac{AP - CP}{OP - CP} = 1 \quad (6)$$

Similarly for  $PER^*$ , if at a certain point in time  $AP(t) = OP(t)$ , and by using Equation (3) we obtain

$$\lim_{\delta_{AP(t)} \rightarrow 0} \left( 1 - \frac{\delta_{AP}(t)}{\delta_{CP}(t)} \right) = 1 \quad (7)$$

If performance is nil, i.e.  $AP = CP$  for  $PER$  and  $AP(t) = CP(t)$  for  $PER^*$ , the limits are simply

$$\lim_{AP \rightarrow CP} \frac{AP - CP}{OP - CP} = 0$$

for  $PER$  and

$$\lim_{\delta_{AP}(t) \rightarrow \delta_{CP}(t)} \left( 1 - \frac{\delta_{AP}(t)}{\delta_{CP}(t)} \right) = 0 \quad (8)$$

for  $PER^*$ . Finally, fictitious worst case scenarios for  $PER$  can be defined by  $AP \rightarrow -\infty$  in maximizing cases, i.e. where  $CP \geq AP \geq OP$ , and by  $AP \rightarrow +\infty$  in minimizing cases, i.e.  $CP \leq AP \leq OP$ . Hence, for the former we get

$$\lim_{AP \rightarrow -\infty} \frac{AP - CP}{OP - CP} = \mp \infty \quad (9)$$

depending on the sign of the difference  $OP - CP$ . Similarly,

$$\lim_{AP \rightarrow +\infty} \frac{AP - CP}{OP - CP} = \pm \infty \quad (10)$$

for the minimizing case. In other words and as explained in Section 2,  $PER$  is a strictly increasing or decreasing function depending on the sign of the denominator. This complication no longer occurs in the case of  $PER^*$  since the worst case can be simply defined by

$$\lim_{\delta_{AP}(t) \rightarrow +\infty} \left( 1 - \frac{\delta_{AP}(t)}{\delta_{CP}(t)} \right) = -\infty \quad (11)$$

## Appendix B – Derivation of Expected Value of $PER^*$

We use a first-order Taylor approximation to linearize Equation (3) around the mean  $\mu_{\delta_{CP}}$  of  $\delta_{CP}(t)$ , assuming that  $\delta_{CP}(t)$  is sufficiently well behaved in the neighborhood of  $\mu_{\delta_{CP}}$ . Hence, we obtain

$$PER^*(t) \approx 1 - \frac{\delta_{CP}(t)}{\mu_{\delta_{CP}}} + \frac{\delta_{AP}(t)(\delta_{CP}(t) - \mu_{\delta_{CP}})}{\mu_{\delta_{CP}}^2} + O[\delta_{CP}(t) - \mu_{\delta_{CP}}]^2 \quad (12)$$

If we drop the second and higher order terms  $O[\delta_{CP}(t) - \mu_{\delta_{CP}}]^2$  in Equation (12), the expected value as denoted by Equation (4) is obtained in the following way:

$$\begin{aligned} \langle PER^* \rangle &= 1 - \frac{2}{\mu_{CP}} \langle \delta_{AP} \rangle + \frac{1}{\mu_{CP}^2} \langle \delta_{AP} \delta_{CP} \rangle = \\ &= 1 - 2 \frac{\mu_{AP}}{\mu_{CP}} + \frac{1}{\mu_{CP}^2} (\langle \delta_{AP} \rangle \langle \delta_{CP} \rangle + Cov(\delta_{AP}, \delta_{CP})) = 1 - \frac{\mu_{AP}}{\mu_{CP}} + \frac{1}{\mu_{CP}^2} Cov(\delta_{AP}, \delta_{CP}) \end{aligned} \quad (13)$$

Note that in Equation (13), we dropped the time subscripts for notational convenience.  $\langle PER^* \rangle$  is not defined for  $\mu_{\delta_{CP}} = 0$ . At such level and circumstances, i.e.  $CP=OP$ , policy-makers would probably not initiate a new policy since any deviation from the status quo would affect the performance measure negatively. Again, in the case of optimality, i.e.  $AP(t)=OP(t)$ ,  $\langle PER^* \rangle = 1$  since  $\delta_{AP}(t) = 0$  for all  $t$  and hence  $\mu_{\delta_{AP}} = 0$ . Therefore  $Cov(\delta_{AP}, \delta_{CP}) = 0$  which follows from

$$Cov(\delta_{AP}, \delta_{CP}) = \langle \delta_{AP}(t) \cdot \delta_{CP}(t) \rangle - \langle \delta_{AP}(t) \rangle \langle \delta_{CP}(t) \rangle = \langle 0 \cdot \delta_{CP}(t) \rangle - 0 \cdot \mu_{CP} = 0.$$

## Appendix C – Derivation of Variance of PER\*

According to standard textbook definition, we start with

$$\sigma_{PER^*}^2 = \langle PER^*(t)^2 \rangle - \langle PER^*(t) \rangle^2 \quad (14)$$

By using Equation (4), the second term on the right-hand side of Equation (14) is

$$\begin{aligned} \langle PER^*(t) \rangle^2 &= \left( 1 - \frac{\mu_{\delta_{AP}}}{\mu_{\delta_{CP}}} + \frac{1}{\mu_{\delta_{CP}}^2} \text{Cov}(\delta_{AP}, \delta_{CP}) \right)^2 = \\ &= \frac{\left( \text{Cov}(\delta_{AP}, \delta_{CP}) + \mu_{\delta_{CP}} (\mu_{\delta_{CP}} - \mu_{\delta_{AP}}) \right)^2}{\mu_{\delta_{CP}}^4} \end{aligned} \quad (15)$$

Similarly, by using Equation (3), the first term on the right-hand side is

$$\langle PER^*(t)^2 \rangle = 1 + \frac{2 \text{Cov}(\delta_{AP}, \delta_{CP}) + 4\mu_{\delta_{AP}}^2 - 2\mu_{\delta_{AP}}\mu_{\delta_{CP}} + 4\sigma_{\delta_{AP}}^2}{\mu_{\delta_{CP}}^2} + \frac{\delta_{AP}^2 \delta_{CP}^2}{\mu_{\delta_{CP}}^4} - \frac{4\delta_{AP}^2 \delta_{CP}}{\mu_{\delta_{CP}}^3} \quad (16)$$

$\langle PER^*(t)^2 \rangle$  cannot be calculated without knowledge of the underlying probability distribution functions of  $AP(t)$ ,  $CP(t)$  and  $OP(t)$  since third and fourth order moments have to be determined (last two terms in Equation (16)). However, we can again linearize these terms. By doing so, after a somewhat tedious calculation, we obtain for the individual higher order terms



$$\frac{\langle \delta_{AP}^2 \delta_{CP}^2 \rangle}{\mu_{\delta_{CP}}^4} \approx \frac{\mu_{\delta_{AP}} \left( 4 \text{Cov}(\mu_{\delta_{AP}}, \mu_{\delta_{CP}}) + \mu_{\delta_{AP}} \mu_{\delta_{CP}} \right)}{\mu_{\delta_{CP}}^3} \quad (17)$$

and

$$\frac{4 \langle \delta_{AP}^2 \delta_{CP} \rangle}{\mu_{\delta_{CP}}^3} \approx \frac{\mu_{\delta_{AP}} \left( 2 \text{Cov}(\mu_{\delta_{AP}}, \mu_{\delta_{CP}}) + \mu_{\delta_{AP}} \mu_{\delta_{CP}} \right)}{\mu_{\delta_{CP}}^3} \quad (18)$$

Subtracting the right-hand side of Equation (15) from the one in Equation (16) we obtain the result in Equation (5).

## Appendix D – Data

month	natural runoff regime		Period 1		Period 2		Period 3		Optim.
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$
1	150.9	27.7	183.9	74.5	478.5	101.1	590.0	55.3	357.7
2	152.2	24.8	196.5	68.9	464.2	113.0	561.8	78.6	426.2
3	178.8	28.6	192.9	49.9	428.9	122.1	465.8	52.9	323.4
4	318.4	92.5	265.7	94.7	350.2	115.6	367.0	79.6	426.2
5	672.2	190.5	443.6	189.7	348.0	120.2	286.8	52.0	452.8
6	987.3	325.8	532.1	205.2	450.1	152.6	270.6	73.8	468.0
7	807.9	258.8	638.5	210.1	481.0	174.5	324.3	78.2	494.7
8	518.1	138.4	518.4	149.6	354.1	79.5	316.6	40.3	490.9
9	289.3	71.2	184.9	96.7	198.5	89.2	228.1	93.0	441.4
10	231.0	49.3	146.5	72.4	234.5	67.7	313.7	86.8	300.6
11	219.0	44.4	143.8	89.0	343.5	51.9	439.4	84.9	304.4
12	176.3	26.8	181.9	80.9	479.7	82.3	590.6	53.0	418.6
<b>Overall</b>	388	307	311	215	384	139	396	141	409

**Table 4:** Means and standard deviations of monthly flows under management systems. The bottom row displays overall means and standard deviations for the duration of the management periods. Units are m<sup>3</sup>/s for  $\mu$  and  $\sigma$ . The last column shows data from (Cai, McKinney et al. 2003).

Month	1	2	3	4	5	6	7	8	9
q [m <sup>3</sup> /s]	495	490	300	230	270	500	650	600	190

**Table 5:** Release schedule of Toktogul reservoir as established in the 1998 treaty. No values were defined for the months of October to December.

## 1 **Bibliography**

- 2 Bennear, L. S., and C. Coglianese. 2005. Measuring progress: Program evaluation of  
3 environmental policies. *Environment* 47 (2):22-39.
- 4 Bernauer, T. 1995. The Effect of International Environmental Institutions - How We  
5 Might Learn More. *International Organization* 49 (2):351-377.
- 6 Brochmann, M., and N. P. Gleditsch. 2006. Conflict, Cooperation, and Good Governance  
7 in International River Basins. Paper read at Governance and the Global Water  
8 System: Institutions, Actors, Scales of Water Governance. Facing the Challenges  
9 of Global Change, 20-23 June 2006, at Bonn, Germany.
- 10 Cai, X. M., D. C. McKinney, and L. S. Lasdon. 2002. A framework for sustainability  
11 analysis in water resources management and application to the Syr Darya Basin.  
12 *Water Resources Research* 38 (6):-.
- 13 Cai, X. M., D. McKinney, and L. S. Lasdon. 2003. Integrated Hydrologic-Agronomic-  
14 Economic Model for River Basin Management. *Journal of Water Resources  
15 Planning and Management* 129 (1):4-17.
- 16 Dukhovny, Victor, and Vadim Sokolov. 2005. Integrated Water Resources Management,  
17 Experience and Lessons Learned form Central Asia - -towards the Fourth World  
18 Water Forum. Tashkent: Inter-State Commission for Water Coordination in the  
19 Aral Sea Basin.
- 20 Greenstone, M. 2004. Did the Clean Air Act Cause the Remarkable Decline in Sulfur  
21 Dioxide Concentrations? *Journal of Environmental Economics and Management*  
22 47:585-611.

- 23 Helm, C., and D. Sprinz. 2000. Measuring the Effectiveness of International  
24 Environmental Regimes. *Journal of Conflict Resolution* 45 (5):630-652.
- 25 Hovi, Jon, D. Sprinz, and Arild Underdal. 2003. Regime Effectiveness and the Oslo-  
26 Potsdam Solution: A Rejoinder to Oran Young. *Global Environmental Politics* 3  
27 (3):105-107.
- 28 Hovi, Jon, Detlef F. Sprinz, and Underdal Arlid. 2003. The Oslo-Potsdam Solution to  
29 Measuring Regime Effectiveness: Critique, Response, and the Road Ahead.  
30 *Global Environmental Politics* 3 (3):74-96.
- 31 Kjeldsen, T. R., and D. Rosbjerg. 2004. Choice of reliability, resilience and vulnerability  
32 estimators for risk assessments of water resources systems. *Hydrological Sciences*  
33 *Journal-Journal Des Sciences Hydrologiques* 49 (5):755-767.
- 34 Loucks, Daniel P., Jerry R. Stedinger, and Douglas A. Haith. 1981. *Water Resource*  
35 *Systems Planning and Analysis*. Edited by I. Prentice-Hall. Englewood Cliffs,  
36 New Jersey: Prentice-Hall, Inc.
- 37 McKinney, D., X. M. Cai, and L. S. Lasdon. 1999. Integrated Water Resources  
38 Management Model for the Syr Darya Basin, edited by EPIC. Almaty,  
39 Kazakhstan: US Agency for International Development
- 40 Neumayer, E. 2002a. Do Democracies Exhibit Stronger International Environmental  
41 Commitment? A Cross-Country Analysis. *Journal of Peace Research* 39:139-164.
- 42 Neumayer, E. 2002b. Does Trade Openness Promote Multilateral Environmental  
43 Cooperation? *The World Economy* 25:815-832.

44 Nurmi, Pertti. 2003. Recommendations on the verification of local weather forecasts. In  
45 *ECMWF technical Memoranda*, edited by E. C. f. M. R. W. Forecasts. Reading:  
46 Finnish Meteorological Institute.

47 Rieckermann, J., H. Daebel, M. Ronteltap, and T. Bernauer. 2006. Assessing the  
48 Performance of International Water Management at Lake Titicaca. *accepted for*  
49 *publication in Aquatic Sciences*.

50 Savoskul, Oxana, Elena Chevnina, Felix Perziger, Ludmila Vasilina, Viacheslav Baburin,  
51 Alexander I. Danshin, Bahtiyar Matyakubov, and Ruslan Murakaev. 2003. Water,  
52 Climate, Food, and Environment in the Syr Darya Basin. In *Adaptation strategies*  
53 *to changing environments*.

54 Sprinz, D. 2005. Regime Effectiveness: The Next Wave of Research. Paper read at  
55 Conference on the Human Dimensions of Global Environmental Change, at  
56 Berlin.

57 Sprinz, D., and C. Helm. 2000. Measuring the Effectiveness of International  
58 Environmental Regimes. *Journal of Conflict Resolution* 44 (5):630-652.

59 Underdal, Arild. 1992. The Concept of Regime "Effectiveness". *Cooperation and*  
60 *Conflict* 27 (3):227-240.

61 World Bank. 1996. Developing a Regional Water Management Strategy: Issues and  
62 Work Plan. *Aral Sea Basin Program Technical Paper Series* 1.

63 Young, Oran. 2001. Inferences and Indices: Evaluating the Effectiveness of International  
64 Environmental Regimes. *Global Environmental Politics* 1 (1):99-121.

65 Young, Oran. 2003. Determining Regime Effectiveness: A commentary on the Oslo-  
66 Potsdam Solution. *Global Environmental Politics* 3 (3):97-104.

67

68