Measuring International Policy Performance

Tobias Siegfried and Thomas Bernauer
ETH Zurich

Correspondence:
Thomas Bernauer
ETH Zurich
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

Abstract

We develop a methodology for estimating the performance (or effectiveness) of international policies (or regimes), building on previous work by Underdal, Sprinz, Helm, and Hovi. Our policy 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 problem solving efforts with respect to the Naryn / Syr Darya, a major international river basin 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 exchanges of water releases for upstream hydropower-production in winter and water releases for downstream irrigation in spring to autumn. We observe that the international regime in place since 1998 is generally characterized by low average performance and high variability. The summer and winter months are contributing to this high variability, whereas in spring and autumn performance is close to optimal. We end with some observations on how the regime could be improved.

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

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 Federal Institute of Hydrology (BfG) in Koblenz, Germany, for providing the pre-1991 flow data.

1. Introduction

Most political science research on the determinants of international cooperation operates with simple notions of the outcome to be explained – most commonly, the existence of agreements, treaties, or international regimes (Bernauer 1995). Substantive assessment of the contents of cooperative arrangements and their performance in terms of solving problems that motivate international cooperation is usually left to qualitative case study research. Recent work on the effectiveness of international environmental cooperation suggests that a quantitative approach is feasible (Underdal 1992; Helm and Sprinz 2000; Young 2001). Such an approach would help in systematically measuring and comparing success or failure in international cooperation over time and across cases. Hence it would provide a more substantive basis also for explaining variation in success or failure of international cooperation. Moreover, it would be of practical relevance for policy evaluation.

The existing literature offers only very limited concepts for measuring international policy performance. Questions about international policy performance are usually answered either with reference to non-causal criteria, for example by describing the development of a particular problem (e.g. pollution) over time without systematic analysis of how the problem has been affected by international cooperation per se. Or they are answered with reference to widely shared views among experts about the effectiveness of cooperation. Moreover, particularly in the tradition of welfare economics, performance is defined chiefly in terms of efficiency (in a cost-benefit sense) rather than effectiveness. Policy performance in the local or national context (e.g. (Bennear and Coglianese 2005)) is usually assessed through quasi-experimental research designs and statistical analysis of differences among "treatment" and "non-treatment" groups (see US clean air study). But such studies require a wealth of data that often does not exist in the international context. In addition, the statistical approach to performance measurement is usually not based on a clear notion of what outcomes would be desirable.

In this paper we develop a new methodology for estimating the performance (or effectiveness) of international policies (or regimes), building on previous work by (Underdal 1992; Helm and Sprinz 2000; Sprinz and Helm 2000; Hovi, Sprinz et al. 2003). Our policy performance metric (we call it PER) 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 fourfold: 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 large amounts of high quality 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 problem solving efforts with respect to the Naryn / Syr Darya, a major international river basin 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 exchanges of water releases for upstream hydropower-production in winter and water releases for downstream irrigation in spring to autumn. We observe that the international regime in place since 1998 is generally characterized by low average performance and high variability. The summer and winter months are contributing to this high variability, whereas in spring and autumn performance is close to optimal.

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 of this concept. In section 3 we develop a new concept that solves the problems discussed in the preceding section. In section 4 we apply this concept to the Naryn / Syr Darya case. In section 4 we summarize the results and end with some observations on how the current regime could be improved.

2. Basic Measurement Concept

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

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

where *AP*: actual performance, *CP*: counterfactual performance, *OP*: optimal performance.[1] This approach to measuring the performance (i.e., effectiveness) of international policies (or international 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 assessed*. In international water management, for example, such criteria may relate to hydropower production, irrigation water provision, and water quality. Generally, *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 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 and minimizing cases (see below) it sets a lower and upper bound and (with some exceptions) standardizes PER_i values between 0 and 1.

---

[1] The parameter names we use differ from the original.

In the remainder of this section we highlight the most important problems of the basic measurement concept. The first problem is that the basic concept has, so far, not distinguished minimizing (*min*) and maximizing (*max*) cases. This can potentially lead to wrong results and, with that, misleading effectiveness scores.

The problems stems from the fact that, surprisingly, the limiting behavior of PER_i has not been systematically examined in other work to date. Let us address this omission quickly. We assume that $\{AP, CP, OP\} \in [-\infty, \infty]$ as well as $AP = CP + \Delta$ with $\Delta \in [-\infty, \infty]$. Thus, the hypothetical limiting case can be assessed by

$$\lim_{\Delta \to \infty} \frac{\Delta}{OP - CP} = \pm\infty \tag{2}$$

Note that in Equation (2), the limiting behavior depends on the sign of the difference $OP - CP$. Similarly,

$$\lim_{\Delta \to -\infty} \frac{\Delta}{OP - CP} = \mp\infty \tag{3}$$

Such limiting behavior forces us to distinguish between two cases. In what follows, we denote the case where $OP > CP$ as the *max* case and where $OP < CP$ as the *min* case. In the *max* case, a policy is designed to maximize the value of a given outcome-variable, e.g. percentage or absolute reduction of some form of pollution or water provision for irrigation. Converse to that, in the *min* case, a policy is meant to minimize the value of a given outcome-variable, e.g. concentrations of some form of pollution. This means that PER_i as given by Equation (1) is a strictly increasing or decreasing function respectively. Figure 1 illustrates these two types of cases ($PER_i\big|_{CP<OP}$ and $PER_i\big|_{CP>OP}$).

According to the definition of PER_i and the actual value of *AP*, the following performance intervals can be identified both for the *min* and *max* cases. If $CP \geq AP > OP$ in the *min* case or $CP \leq AP < OP$ in the *max* case, then $PER_i \in [0,1]$[2]. More precisely,

$$\lim_{AP \to OP} \frac{AP - CP}{OP - CP} = 1 \tag{4}$$

which indicates perfect policy performance. Converse to that,

---

[2] Note that the above definition implies $CP > OP$ in the *min* case and $CP < OP$ in the *max* case. We therefore exclude cases where $CP = OP$. At such level and circumstances, policy-makers would probably not initiate a new policy since any deviation from the status quo would affect the performance measure negatively.

$$\lim_{AP \to CP} \frac{AP - CP}{OP - CP} = 0 \qquad (5)$$

which indicates that policy performance is nil. These results hold for both the maximization as well as the minimization case.

We can, however, think of situations where policies produce outcomes that are less favorable compared to the counterfactual performance. Therefore, in the *min* case, we get

$$\lim_{\Delta \to +\infty} \frac{\Delta}{OP - CP} = -\infty \qquad (6)$$

since $OP - CP < 0$. Very similarly, with $OP - CP > 0$ in the *max* case,

$$\lim_{\Delta \to -\infty} \frac{\Delta}{OP - CP} = -\infty \qquad (7)$$

Therefore, in such “management made things worse“ – situations, as given by Equations (6) and (7), $PER_i \in [0, -\infty]$.

A conceptual problem with the definition of PER_i arises because the basic measurement concept is not symmetric around *OP* (see Figure 1 and Figure 3). A simple example demonstrates why this is of relevance. Imagine, for example, that *PER* is assessed with regard to demand coverage. Let us assume that OP is equivalent to freshwater demand of a particular economic sector. Furthermore, we assume that $AP = OP + \Delta$. By using Equation (1), it follows immediately that $PER = 1 \pm |\Delta| / (OP - CP)$. Obviously, if $\Delta < 0$ the allocated water is somewhat suboptimal and thus $PER < 1$. This corresponds to $PER_i(AP) = a$ in Figure 1. Conversely, if $\Delta > 0$, too much water is allocated to a particular sector and hence wasted ($PER_i(AP) = c$ in Figure 1). Yet, for the latter case, we calculate $PER > 1$, which would suggest that wasting resources in allocating ‘too much’ is preferable over the allocation of ‘too little’. However, both conditions are clearly undesirable, if only from an economic point of view (see Figure 1). 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, Daebel et al. 2006)). However, the latter approach introduces additional uncertainty by the ad-hoc assignment and scaling of the parameter values.

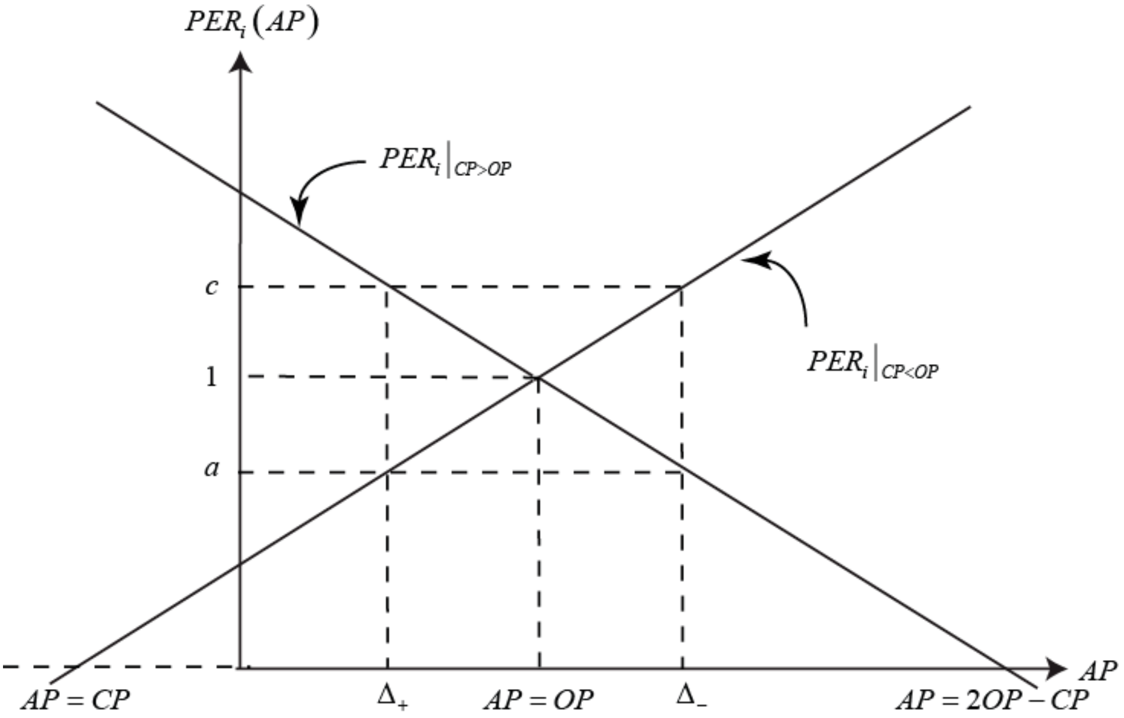

Figure 1: Conceptual difference of *max* and *min* cases in estimating PER_i. Given CP and OP at a specific time *t*, PER is simply an increasing ($CP < OP$) / decreasing ($CP > OP$) function of AP.

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 2 and 3. The estimation of PER_i at time $t = t_1$ leads to the value b as highlighted in Figure 3. If PER_i is assessed at time $t = t_2$, performance c is obtained, which clearly differs from the performance value b. Policy performance, however, 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 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 to soil as well as deltaic systems, however, accumulate in time and gradually start to show after only some decades after which related services to society may have completely vanished. If performance is viewed as a measure related to demand coverage, initial hydropower demand may have been fully met (*PER* = 1). But subsequent demand coverage in respect to downstream environmental services would have experienced a dramatic decline. Any assessment of *PER* at a certain time would therefore only provide a partial picture of performance.

In other words, measurement of *PER* must pay attention to time dependence (see the water engineering literature for other performance criteria that account for time (Kjeldsen and Rosbjerg 2004)). This critique is also put forth by Young (2001) who states 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 ongoing academic research in the respective field (Hovi, Sprinz et al. 2003; Hovi, Sprinz et al. 2003; Young 2003; Sprinz 2005).

In Section 3, we address all of the aforementioned problems in greater detail.

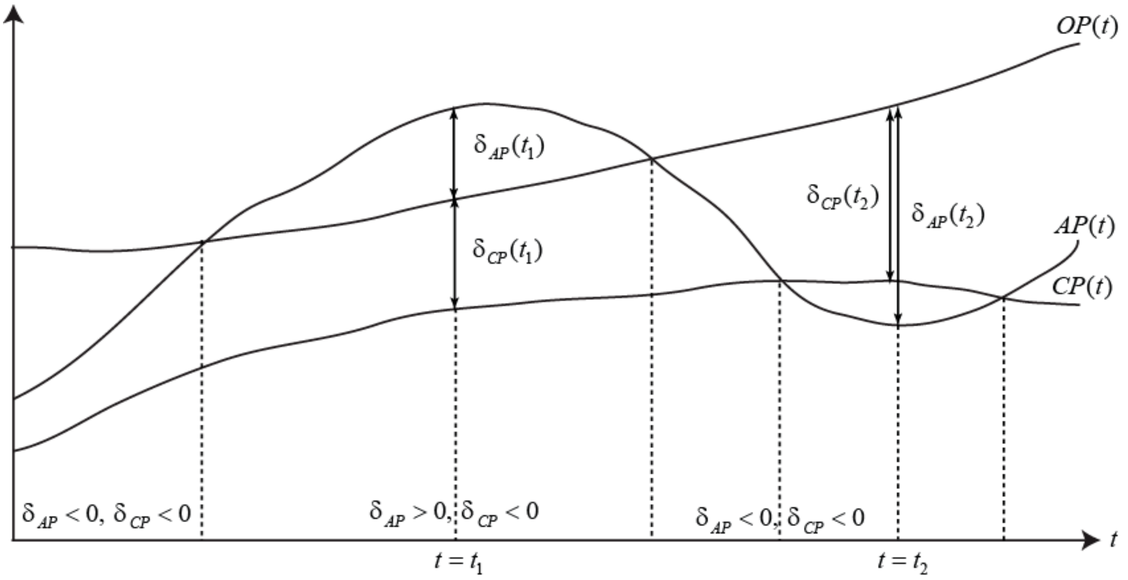

Figure 2: Stylized development of $AP(t)$, $CP(t)$ and $OP(t)$ over time. δ_{AP} and δ_{CP} as defined in Equation (9) are shown at different times t_1 and t_2.

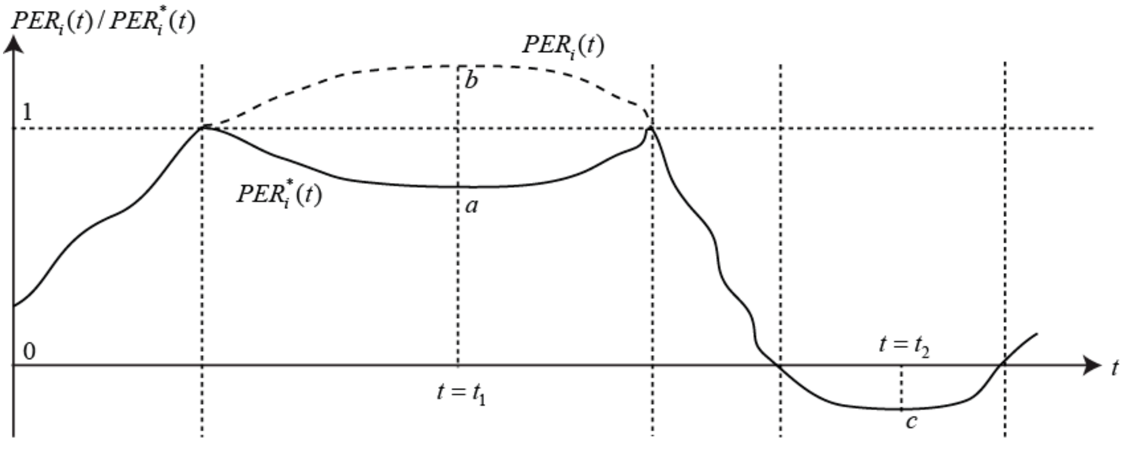

Figure 3: Stylized development of PER_i and PER_i^* as time dependent function of the stochastic processes as depicted in Figure 2. Clearly, $PER_i(t_1) \neq PER_i(t_2)$. Note that $PER_i(t) > 1$ during a certain time interval, which would lead us to assume falsely that during such wasteful allocation the performance of the investigated measure is highest.

3. Upgraded Policy Performance Concept

3.1 Definition

We propose the definition of a performance measure, given by

$$PER_i^*(t) = 1 - \left| \frac{AP(t) - OP(t)}{CP(t) - OP(t)} \right| \tag{8}$$

where $PER_i^*(t)$ is a measure of management performance at a certain time t. If we use the notation $\delta_{AP}(t) = AP(t) - OP(t)$ and $\delta_{CP}(t) = CP(t) - OP(t)$, then Equation (8) becomes

$$PER_i^*(t) = 1 - \sqrt{\frac{\delta_{AP}^2(t)}{\delta_{CP}^2(t)}} \tag{9}$$

by the definition of the absolute value. Since we assume $\{AP(t), CP(t), OP(t)\} \in [-\infty, \infty]$, it follows immediately $\{\delta_{AP}(t), \delta_{CP}(t)\} \in [-\infty, \infty]$.

According to Equation (9), $PER_i^*(t)$ is defined as long as $\delta_{CP} \neq 0$. Depending on the signs of $\delta_{AP}(t)$ and $\delta_{CP}(t)$, Equation (9) is equivalent to

$$PER_i^*(t) = \begin{cases} 1 + \dfrac{\delta_{AP}(t)}{\delta_{CP}(t)}, & \text{if } \operatorname{sign}(\delta_{AP}(t))\operatorname{sign}(\delta_{CP}(t)) < 0 \\ 1 - \dfrac{\delta_{AP}(t)}{\delta_{CP}(t)}, & \text{if } \operatorname{sign}(\delta_{AP}(t))\operatorname{sign}(\delta_{CP}(t)) > 0 \end{cases} \tag{10}$$

If $\operatorname{sign}(\delta_{AP}(t))\operatorname{sign}(\delta_{CP}(t)) < 0$, either $\delta_{AP}(t) < 0$ and $\delta_{CP}(t) > 0$ or $\delta_{AP}(t) > 0$ and $\delta_{CP}(t) < 0$ holds. Similarly, $\delta_{AP}(t) < 0$ and $\delta_{CP}(t) < 0$ or $\delta_{AP}(t) > 0$ and $\delta_{CP}(t) > 0$ so that $\operatorname{sign}(\delta_{AP}(t))\operatorname{sign}(\delta_{CP}(t)) > 0$. Converse to PER_i, $PER_i^*(t)$ measures performance relative to optimal performance *OP* at a specific observation time t. Note that $PER_i^*(t)$ is symmetric around *OP* (see Figure 4).

3.2 Limiting Behavior

The definition of $PER_i^*(t)$ involves the absolute values of the differences between *AP(t)* and *OP(t)* as well as *CP(t)* and *OP(t)*. The conceptual difference that was necessary to account for while dealing with PER vanishes, i.e. the cases of $CP \geq AP > OP$ and $CP \leq AP < OP$ can be treated mathematically in a similar way. To see this, we explore the limiting behavior of $PER_i^*(t)$.

For optimal management, i.e. $AP = OP$, we obtain

$$\lim_{\delta_{AP}(t)\to 0}\left(1-\sqrt{\frac{\delta_{AP}^2(t)}{\delta_{CP}^2(t)}}\right)=1 \tag{11}$$

If performance is nil, i.e. $AP = CP$,

$$\lim_{\delta_{AP}(t)\to \delta_{CP}(t)}\left(1-\sqrt{\frac{\delta_{AP}^2(t)}{\delta_{CP}^2(t)}}\right)=0 \tag{12}$$

Finally, the hypothetical worst case scenario is defined by

$$\lim_{\delta_{AP}\to \pm\infty}\left(1-\sqrt{\frac{\delta_{AP}^2}{\delta_{CP}^2}}\right)=-\infty \tag{13}$$

Hence, and compared to *PER*, the use of $PER_i^*(t)$ does not force us to take into account conceptual differences between maximization and minimization cases. Furthermore, wasteful management, i.e. situations where $AP(t) > OP(t)$ in the *min* case and $AP(t) > OP(t)$ in the *max* case are no longer rewarded.

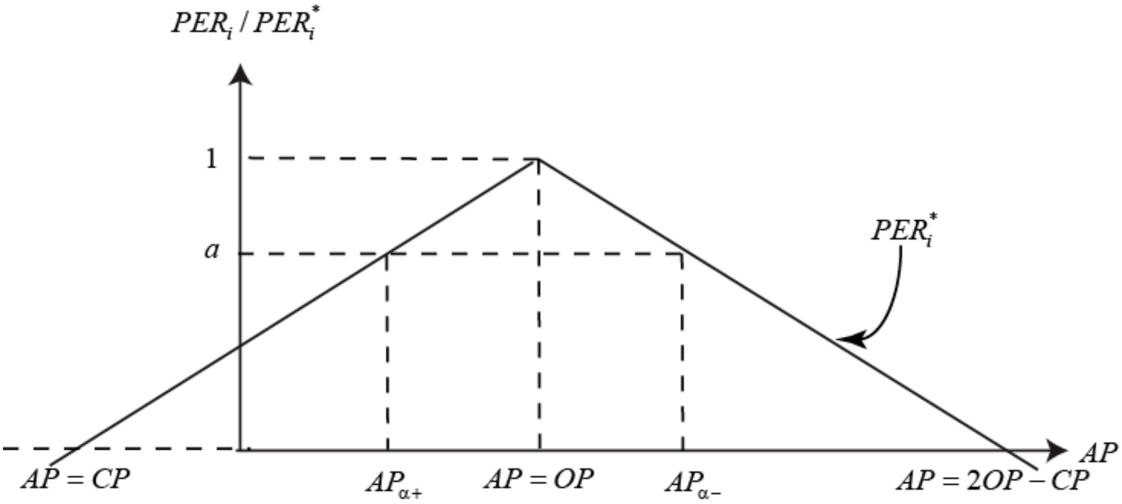

Figure 4: The maximum of PER_i^* occurs at $AP = OP$. Suboptimal performance, i.e. either too much or too less of AP, leads to $PER_i^*(t) = a$ with $a < 1$. Note that time subscripts have been omitted in the Figure for clarity.

3.3 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. Variations in time can for example be caused by seasonal variations, 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 runoff processes (for an example, see Figure 6) 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 the underlying stochastic processes. In the subsequent analysis, we restrict our attention to stationary processes[3]. Our goal is to provide a general, yet easy approach to the characterization of our performance measure over a certain period of time by making use of basic concepts and definitions of probability theory and statistics. In doing so, we neither assume knowledge of the underlying probability distribution functions nor of the stochastic processes that eventually produce the realizations $\delta_{AP}(t)$ and $\delta_{CP}(t)$.

---

[3] 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.

Let us use a first-order Taylor approximation to linearize Equation (9) 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 get

$$PER_i^*(t) = 1 + \sqrt{\frac{\delta_{AP}^2(t)}{\mu_{\delta_{CP}}^2}} - \frac{\left(\delta_{CP}(t) - \mu_{\delta_{CP}}\right)\sqrt{\frac{\delta_{AP}^2(t)}{\mu_{\delta_{CP}}^2}}}{\mu_{\delta_{CP}}} + O\left[\delta_{CP}^{Abs}(t) - \mu_{\delta_{CP}^{Abs}}\right]^2 \qquad (14)$$

We define two new random variables, $\delta_{AP}^{Abs}(t) = \left|\delta_{AP}(t)\right|$ and $\delta_{CP}^{Abs}(t) = \left|\delta_{CP}(t)\right|$ with $\delta_{AP,CP}^{Abs}(t) \in [0, \infty]$ so that Equation (14) can be simplified to

$$PER_i^*(t) \approx 1 - \frac{2\delta_{AP}^{Abs}(t)}{\mu_{\delta_{CP}^{Abs}}} + \frac{\delta_{AP}^{Abs}(t)\delta_{CP}^{Abs}(t)}{\mu_{\delta_{CP}^{Abs}}^2} \qquad (15)$$

Note that the second order terms $O\left[\delta_{CP}^{Abs}(t) - \mu_{\delta_{CP}^{Abs}}\right]^2$ have been dropped in this approximation of Equation (14). Hence, for an approximation of the expected value of $PER_i^*(t)$ we get

$$\left\langle PER_i^* \right\rangle = 1 - \frac{\mu_{\delta_{AP}^{Abs}}}{\mu_{\delta_{CP}^{Abs}}} + \frac{1}{\mu_{\delta_{CP}^{Abs}}^2}\mathrm{Cov}(\delta_{AP}^{Abs}, \delta_{CP}^{Abs}) \qquad (16)$$

where $\mathrm{Cov}(\delta_{AP}^{Abs}, \delta_{CP}^{Abs})$ denotes the covariance of the time series δ_{AP}^{Abs} and δ_{CP}^{Abs}[4]. Taking the covariance into account is relevant in many cases. Imagine for example pre- and post-impoundment runoff in a river. Depending on the management of the constructed dam, pre and post flow regimes are still correlated to variable degree[5]. The magnitude of such covariance depends on the variances $\sigma_{\delta_{AP}^{Abs}}^2$ and $\sigma_{\delta_{CP}^{Abs}}^2$. If δ_{AP}^{Abs} and δ_{CP}^{Abs} are entirely uncorrelated, then $\mathrm{Cov}(\delta_{AP}^{Abs}, \delta_{CP}^{Abs}) = 0$. Note that $\left\langle PER_i^* \right\rangle < 1$ since $\mu_{\delta_{AP}^{Abs}} / \mu_{\delta_{CP}^{Abs}} > 0$. Furthermore, $\left\langle PER_i^* \right\rangle$ is not defined for $\mu_{\delta_{CP}^{Abs}} = 0$ which would again correspond to the situation described in Footnote 1. Finally, in the case of optimality, i.e. *AP(t)=OP(t)*, $\left\langle PER_i^* \right\rangle = 1$ since $\delta_{AP}^{Abs}(t) = 0$ for all *t* and hence $\mu_{\delta_{AP}^{Abs}} = 0$ and therefore $\mathrm{Cov}(\delta_{AP}^{Abs}, \delta_{CP}^{Abs}) = 0$[6].

---

[4] The derivation of the expectation values is shown in Appendix A.

[5] See also Section 4 for a real world example of pre- and post-impoundment flow correlation.

[6] The latter can be easily shown by noting that

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

Similarly, the variance[7] of $PER_i^*(t)$ is approximated by

$$\sigma^2_{PER_i^*} = \mu^{-4}_{\delta^{Abs}_{CP}} \left(4\mu^2_{\delta^{Abs}_{CP}}\sigma^2_{\delta^{Abs}_{AP}} - \mu^2_{\delta^{Abs}_{AP}}\sigma^2_{\delta^{Abs}_{CP}} - \mathrm{Cov}(\delta^{Abs}_{AP},\delta^{Abs}_{CP})\left(\mathrm{Cov}(\delta^{Abs}_{AP},\delta^{Abs}_{CP}) + 2\mu_{\delta^{Abs}_{AP}}\mu_{\delta^{Abs}_{CP}}\right)\right) \tag{17}$$

(Young 2001) states that procedures involving counterfactual analysis to assess international regime effectiveness (i.e. international policy performance) 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 present the results of an ex post assessment of international policy 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^2, its population around 40 million – i.e. population density is rather low. Figure 5 provides an overview. The economies of the Naryn / Syr Darya's riparian countries (Kazakhstan, Kyrgyzstan, Uzbekistan, Tadjikistan, Turkmenistan) are heavily dependent on irrigated agriculture (with shares of 40 – 50 % of GDP in 1960-1990, and around 20-30% thereafter). Farming employs about 60 % of the rural population and 25-60% of the total labor force (World Bank 1996, Dukhovny and Sokolov). Most water for irrigation is abstracted from the two Daryas. While some upstream parts of the basin are mountenous and humid, the mid- and downstream areas are arid (low and irregular precipitation, large daily and seasonal temperature differences, high solar radiation, low humidity). Over the past 40 years, excessive water withdrawals have led to a drastic shrinkage of the Aral Sea; the latter receives all 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).

The Syr Darya river originates as the Naryn river in the mountains of Kyrgyzstan (see Figure 5). It then flows through Uzbekistan and Tadjikistan and ends in the Aral Sea in Kazakhstan (total length around 2800 km). In total, approximately 20 million people

---

[7] The derivation of the variance is given in Appendix B.

inhabit this river catchment which covers an area of ca. 250’000 km^2. The river is mainly fed by snowmelt and water from glaciers. The natural runoff regime, with a mean annual flow of around 23.5–51 km^3 (around 40 km^3 in the past few years) is characterized by a spring / summer flood. It usually starts in April and peaks in June. Around 93% of the Syr Darya’s mean annual flow is regulated by storage reservoirs. Around 75 % of the run-off stems from Kyrgyzstan (Dukhovny and Sokolov). 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 6 shows the time series of the Naryn / Syr Darya river flow over the last 72 years as measured at Uch Kurgan gauge station, Uzbekistan.

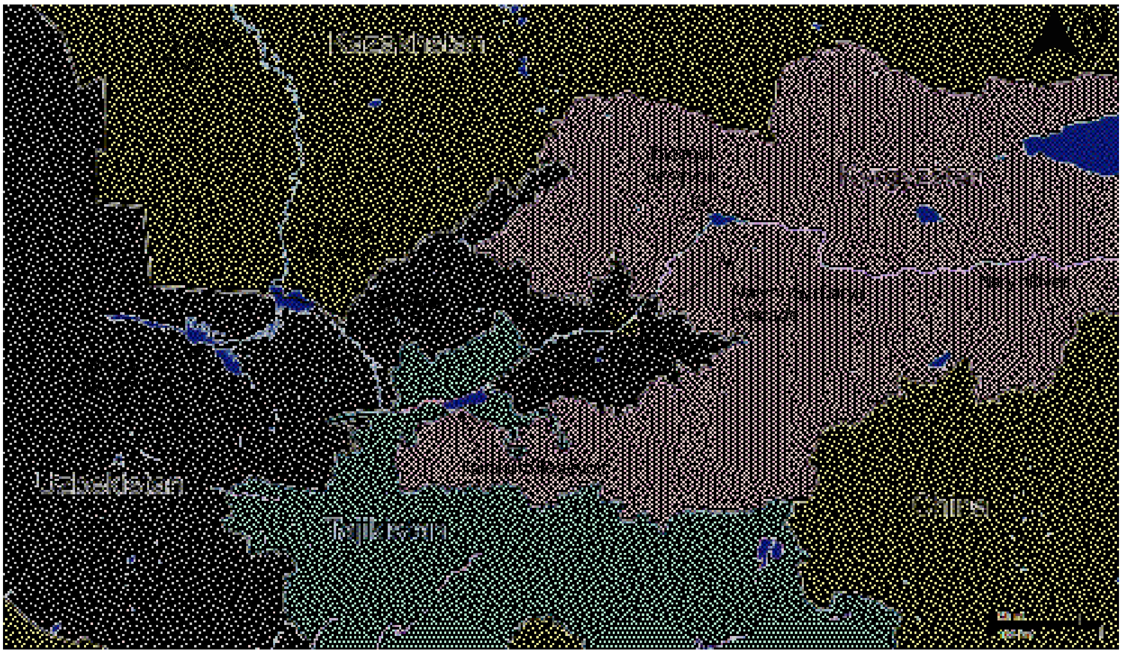

Figure 5: This map shows the part of the Naryn and Syr Darya catchment that is of most interest in this paper. The Uch Kurgan gauge station is located in the center of the map.

As highlighted in Figure 6 the run-off regime of the Naryn / Syr Darya, as measured at the foot of the Naryn / Syr Darya cascade right after the river enters Uzbekistan from Kyrgyzstan, varies strongly over time. It is marked by four distinct periods. The first substantial change in flow patterns came with the Toktogul reservoir in 1976 (though some smaller reservoirs downstream, notably the Kairakkum and Chardara reservoirs, had been put in place earlier). The Toktogul reservoir is by far the largest storage facility in the Aral Sea basin. It came into operation in 1976 after a 14 year construction phase. It has 14 km^3 effective capacity, 8.7 km^3 firm yield and a full capacity of ca. 19.5 km^3. The reservoir area is around 280 km^2, its length around 65 km.[8] Hydropower capacity of the Toktogul power plant is 1’200 MW, i.e. the second biggest in the Aral Sea basin (see (Antipova,E., A. Zyryanov, et a. 2002)).

---

[8] Total usable reservoir capacity in the Syr Darya basin is around 27 km^3.

The time-period 1976–1991 was characterized by centralized management of the river system by the former Soviet Union. This management system was oriented primarily towards adequate water provision for irrigated agriculture (above all, cotton production) in Uzbekistan and Kazakhstan. 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^3/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. The also had the authority to increase or reduce allocations to each country 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 her increased energy demand during the colder months. The fossil fuel was used primarily in the thermal power plants in Bishkek and Osh. (Cai, McKinney et al. 2002).

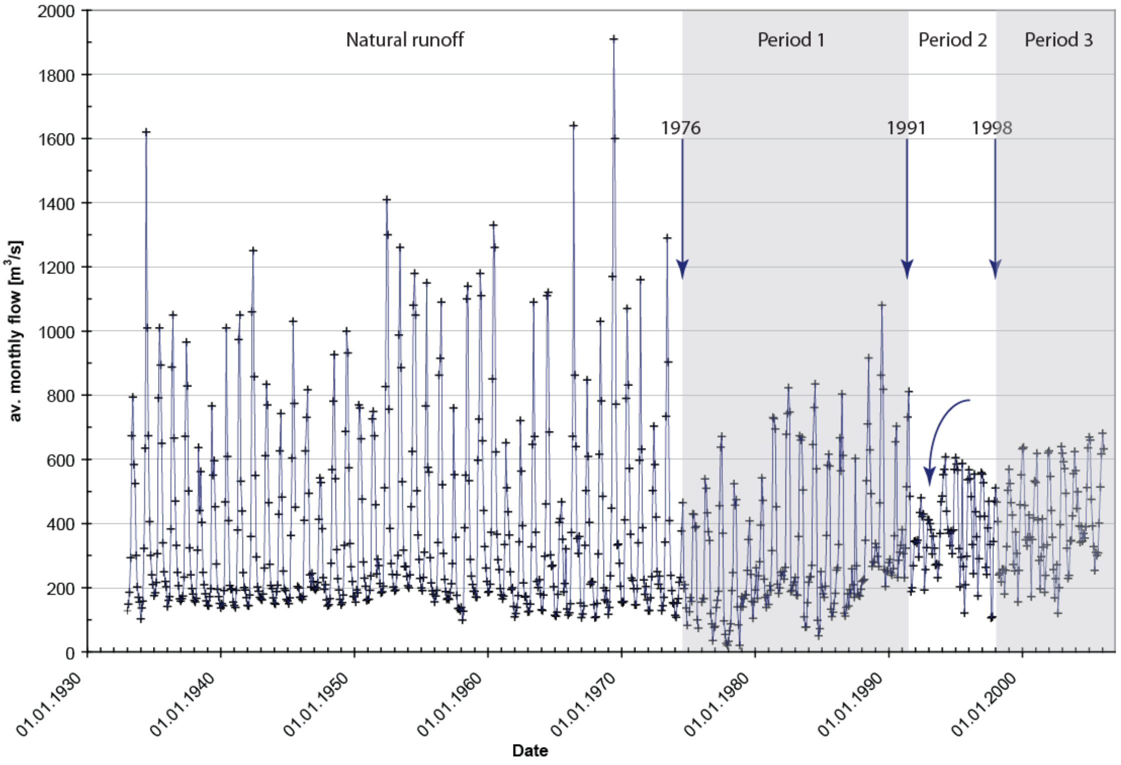

Figure 6: Mean monthly flow of Naryn / Syr Darya River at Uch Kurgan gauge from January 1933 to February 2006. The four different flow regimes, i.e. pre-Toktogul (1933 – 1975), USSR Naryn – Syr Darya cascade management (1976 – 1991), post-USSR operation (1992 – 1998) and ICWC agreement regime

(1998 – today) are clearly distinguishable in the time-series. Data Sources: Global Runoff Data Center (GRDC) and Andrey Yakovlev, Head of the Department of Operational Hydrology of the Uzbek Hydrometeorological Service, Uzbekistan.

The collapse of the Soviet Union in 1991 led to the breakdown of centralized water resources management and water-energy tradeoff arrangements, causing serious disputes between the states over water allocation issues (see Figure 7 for a timeline of key events). Coal, oil, natural gas, and electricity supplies to Kyrgyzstan declined dramatically between 1991 and 1998, and so did the thermal and electric power output of Kyrgyz thermal power plants (TPP).[9] Consumers thus turned to electricity, which increased winter demand by more than 100%. Purchases of energy from abroad are difficult because the government has been (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 power production mode. As of winter 2003/2004 the flow peaks no longer occur in summer but rather in winter, as indicated by the bent arrow in Figure 6. Since 1992 winter spills from the river into the desert have damaged infrastructure and land resources in Uzbekistan. They have also deprived the Syr Darya delta and the northern part of the Aral Sea of water, and they have reduced the potential for water releases for irrigation during the vegetation perid. Ever since 1991 the riparian countries have been struggling to re-establish an effective management scheme (Savoskul, Chevnina et al. 2003).

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 (see above). Kyrgyzstan also views electricity production as a potential export commodity. Kyrgyzstan is thus eager to store water in spring to autumn and release it in winter to spring for energy production. Conversly, 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 reduced water flow capacity (Savoskul, Chevnina 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 tradeoffs between consumptive water use for downstream irrigation purposes and non-consumptive use for upstream energy production in Kyrgyzstan.

---

[9] Thermal power output in Kyrgyz TPPs between 1991 and 1998 declined from 5.8 Gcal. to 2.1 Gcal. Electric power output decreased from 3.9 to 1.0 MkWh.

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.[10] 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 hydrolic 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 largely transferred to their national energy agencies).

In summary, the time-period of 1991 – 1998 (see below) is marked by a collapse of the formerly centralized basin management system and, prima facie, very little success in establishing an effective new international management system that would allow for exchanges of resources among Kyrgyzstan (which is rich in water but poor in fossil fuels) and downstream countries (which are poor in water but richer in fossil fuels). The BWOs lost much of their authority and operational capacity.

A series of declarations by the riparian countries and attempts by European and North-American governmental 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 information system and a water use and farm management system were set up. However, 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. In 1999 Tajikistan joined this agreement.[11]

The 1998 agreement sets the following water release schedule for the Toktogul reservoir in 1998:

January	495 m3/sec
February	490 m3/sec
March	300 m3/sec
April	230 m3/sec
May	270 m3/sec
June	500 m3/sec
July	650 m3/sec
August	600 m3/sec
September	190 m3/sec

No indications are given for the time-period of October to December. However, the aim of the parties is to also prevent flooding of areas in the mid- and downstream Syr Darya sections. So we may presume that water releases of no more than 200 m3/sec in that period would seem reasonable.

---

[10] http://www.icwc-aral.uz/

[11] http://ocid.nacse.org/cgi-bin/qml/tfdd/treaties.qml

The agreement (consisting of two separate formal treaties) is set up as a more general framework agreement and a specific barter agreement on energy-water exchanges in 1998. The specific agreement holds that i the growing season (April 1 – October 1), Kyrgyzstan agrees to supply 2.2bn kWh of electricity to Kazakhstan and Uzbekistan (1.1bn 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[12], 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. In 2003 the agreement was automatically extended for another five years.

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^3 of water during the vegetation period and transfer around 2.2 M kWh of electricity to Uzbekistan and Kazakhstan.

4.2 Assessment of Performance

Based on the methodology developed in section 3 we now assess the performance of the international water management system introduced in 1998. Figure 6 reveals four distinct periods of management. These periods are characterized by differing flow regimes that are associated with the timeline of political events as portrayed in Figure 7.

During the phase of natural runoff (1933–1975), mean flow was 388 m^3/s, with a high variability in summer (for the latter, see σ(natural runoff regime) in Appendix D, Table 3). In this period, the high variability is entirely determined by climatic variability (see also Figure 6). In the period of centralized water resources management under USSR rule (period 1, 1976–1990), the mean flow was reduced to 311 m^3/s mainly due to the filling of the Toktogul reservoir[13]. The characteristics of the yearly averages do not substantially differ from the undisturbed regime, with a summer discharge peak and winter low flow.

---

[12] http://ocid.nacse.org/cgi-bin/qml/tfdd/treaties.qml

[13] If we assume an average of 14 km^3 dam storage volume to be filled at a rate of 70 m^3/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.

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 μ(P3) in Figure 6. 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 σ(Period 3) in Appendix D, Table 3) characterizes the unilateral upstream management of the Syr Darya runoff. Finally, after the implementation of the 1998 agreement, monthly flows appear to reflect the tradeoffs made in that agreement. During this time-period, average flow is 396 m^3/s, with a considerable decline in monthly variability.

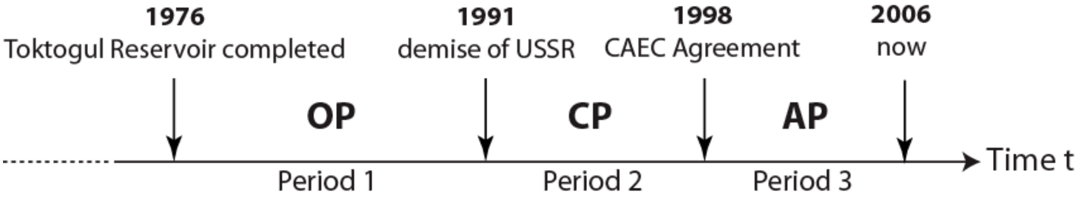

Figure 7: Timeline of events. OP: period of optimal performance (16 years); CP: period of counterfactual performance (7 years); AP: actual regime performance (8 years).

In the following, we start with the assumption that the centralized management approach during Soviet times amounts to optimal performance $OP_S(t)$, for at that time diverging upstream and downstream interests were successfully addressed. Clearly, from the perspective of the Aral Sea problem, this period has hardly been optimal[14]. Consequently, we employ a second notion of optimality, which emphasizes sustainability of natural resources management on the basin scale (including soil, surface and subsurface water resources, see (McKinney, Cai et al. 1999; Cai, McKinney et al. 2003)). Note that μ(optim.) in Figure 8 is not actually observed but rather the result of a simulation-optimization approach that we denote as $OP_C(t)$. The period of breakdown of the centralized management system in 1992–1998 is defined as counterfactual performance, i.e. *CP(t)*. Finally, the current flow regime is denoted as actual performance *AP(t)*.

---

[14] Young (2001) argues that the agreed upon notion of what is the optimum with reference to which performance is assessed must not necessarily be based on an objective notion, but rather depends on an understanding of the nature of the problem and the options available for solving the problem.

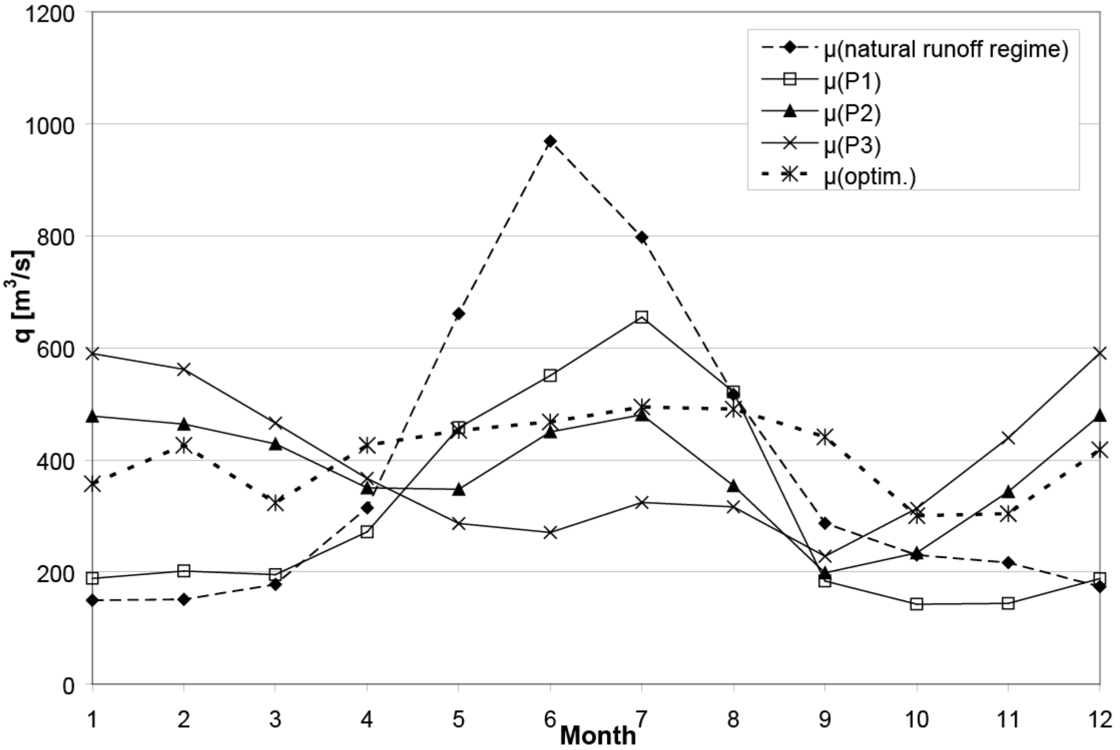

Figure 8: Monthly long-term average flows at the Uch Kurgan gauge (based on data from GRDC and Andrey Yakovlev). The data on flow variability for the corresponding months and regimes is plotted in Figure 9, the numeric data can be found in Appendix D, Table 3. See text for further explanation. The monthly data μ(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 (Cai, McKinney et al. 2003).

To calculate $\left\langle PER_i^* \right\rangle$ and $\sigma^2_{PER_i^*}$, as discussed in Section 3.3, we need to estimate the sample means $\widehat{\mu}_{\delta_{AP}^{Abs}(\bullet)}$, $\widehat{\mu}_{\delta_{CP}^{Abs}(\bullet)}$, the variances $\widehat{\sigma}^2_{\delta_{AP}^{Abs}(\bullet)}$, $\widehat{\sigma}^2_{\delta_{CP}^{Abs}(\bullet)}$ as well as the covariances $\hat{\gamma}(\delta_{AP}^{Abs}(\bullet), \delta_{CP}^{Abs}(\bullet))$ (see Appendix C, Equations (23), (24), (25) and (26)). Note that in the above notation, $(\bullet)$ is a placeholder for both, *OP_S(t)* and *OP_C(t)*. The values of $\delta_{AP}^{Abs}(S)$, $\delta_{AP}^{Abs}(S)$, $\delta_{AP}^{Abs}(S)$ and $\delta_{AP}^{Abs}(S)$ are provided in Appendix D, Tables 3 – 6.

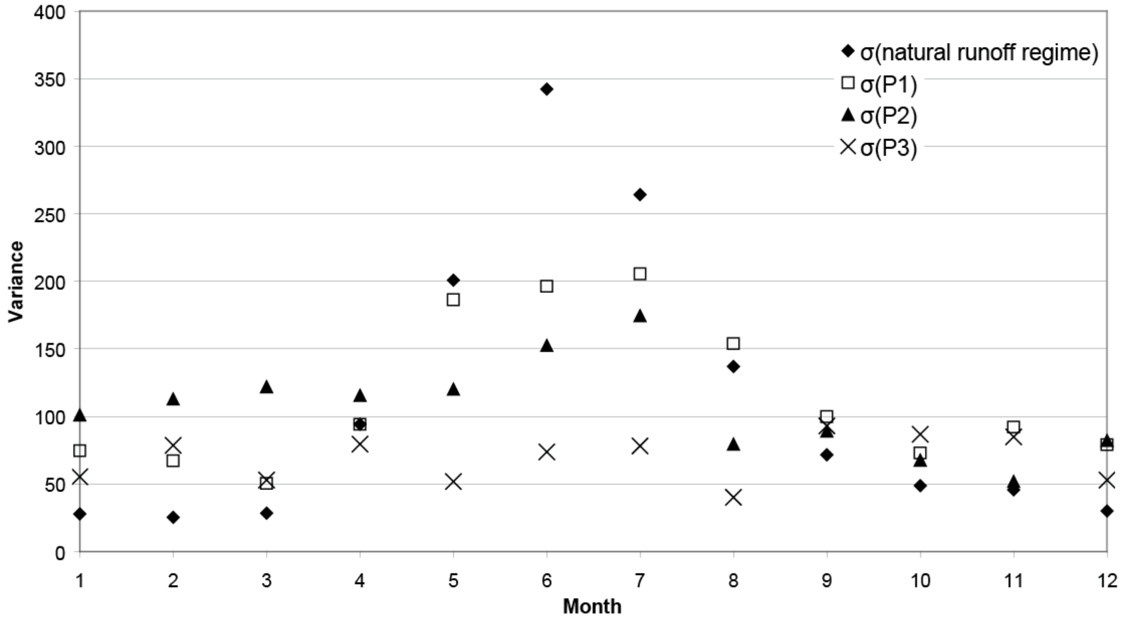

Figure 9: Monthly flow variation is calculated over the respective regime period lengths. Generally, human river regulation has led to an overall decline of monthly flow variability. This decline is most pronounced in the undisturbed summer months, i.e. June – August. See also Figure 8 for monthly mean flows.

The fact that we are dealing with times series of unequal length forces us to choose a maximal management period interval for our analysis[15]. Period 2, i.e. *CP(t)*, lasted for 7 years and is the shortest management period identified. Hence, *AP(t)* and $OP_S(t)$ are truncated accordingly. We choose the interval 1998-2004 for AP and 1984 – 1990 for $OP_S(t)$[16]. The estimated values for the mean and variance are shown below in Table 1. For the covariances, we obtain $\hat{\gamma}(\delta_{AP}^{Abs}(S), \delta_{CP}^{Abs}(S)) = 11854.6$ and $\hat{\gamma}(\delta_{AP}^{Abs}(C), \delta_{CP}^{Abs}(C)) = 1149.0$ correspondingly.

	$\delta_{AP}^{Abs}(S)$	$\delta_{CP}^{Abs}(S)$	$\delta_{AP}^{Abs}(C)$	$\delta_{CP}^{Abs}(C)$
μ	260.3	198.5	155.4	120.4
σ^2	26505.9	17085.8	6034.9	7118.4

Table 1: 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[17].

---

[15] An alternative approach would be to calculate monthly averaged fluxes for OP_S, *CP* and *AP* as they are shown in Figure 8. However, the loss of temporal information introduces an estimation error into the sample values of mean, variance and covariance.

[16] This ensures also that we remove the trend effect in OP_S that is due to the filling of the Toktogul reservoir (see also Footnote 13).

[17] 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 simply assume that the monthly values of OP_C do not change over the period of assessment (7 years).

Finally, we calculate the regime performance and its variance. The results are displayed in Table 2.

	$\langle PER_i^* \rangle$	$\sigma^2_{PER_i^*}$
OP_S	-0.01	1.07
OP_C	-0.21	0.64

Table 2: Average regime performance and variance with reference to OP_S and OP_C respectively. The calculations are based on the values presented in Table 1.

5. Conclusion

In a recent review of existing approaches to the measurement of international regime effectiveness, (Sprinz 2005) identifies several issues that should be addressed in future research in this area. One of the key issues is inter-temporal comparison and assessment of performance. As discussed in Section 2 and illustrated in our case study, non-regime counterfactuals (counterfactual performance) as well as optima change over time and over variable scales (monthly to decadal variations). Sprinz addresses this problem by setting absolute upper and lower bounds between which *AP*, *OP* and *CP* may vary in time. In our view, this approach does not solve the problem.

First, absolute lower and upper bounds depend on the policy problem at hand and are hard to identify in a reliable fashion. One example is an environmental bad, where OP=0 will inevitably lead to the highest welfare. Yet, in many other cases, these bounds are unclear and/or contested – examples include upper bounds of carbon dioxide concentrations in the atmosphere or the identification of an optimal surface water runoff level that obviously varies seasonally and according to downstream consumptive use. Sprinz (2005) notes this problem, but simply refers to sensitivity analysis to assess the robustness of the calculated performance level. This is clearly not enough. Second, such an approach does not solve any of the other problems mentioned in Section 2.

The methodology proposed in this paper addresses these gaps more systematically. It deals in a transparent and tractable way with the fact that *AP*, *CP* and *OP* are time dependent variables that relate to a particular international policy (or regime) and particular realizations of underlying stochastic processes. To the extent that times-series data of reasonable quality for policy outcomes is available, our methodology can be applied to virtually an international (and also national or local) policy or international regime to study its performance (or effectiveness).

To illustrate the empirical relevance of the methodology, we carried out an ex post performance assessment of the international regime for managing the Naryn / Syr Darya river basin (with a focus on the Toktogul reservoir in Kyrgyzstan). The results show that this regime is generally characterized by low average performance and high variability (see $\langle PER_i^* \rangle$ and $\sigma^2_{PER_i^*}$ in Table 2). In looking at monthly averages runoff graphs shown in Figure 8, we note that the summer and winter months are contributing to this high variability, whereas in spring and autumn performance is close to optimal.

Low average performance and high variability are certainly a major problem in the Naryn / Syr Darya regime. But this does not mean that the 1998 agreement per se is the wrong approach or obsolete. Performance could undoubtedly be improved with reference to the optimal water release schedule $OP_C(t)$. This could be achieved by adjusting AP(t) closer to OP(t) – see the definition of $\left\langle PER_i^* \right\rangle$ and Equation (16). To that end, either $\mu_{\delta_{AP}^{Abs}}$ can be reduced and/or $\mathrm{Cov}(\delta_{AP}^{Abs}, \delta_{CP}^{Abs})$ can be increased. However, in our specific case of international river management through reservoir operation average discharge $\mu_{\delta_{AP}^{Abs}}$ is much harder to control since this quantity cannot be increased or decreased significantly unless new dams are constructed or existing ones are decommissioned (see also below). It is likely to be easier to reduce temporal flow variability.

In practical terms, international efforts sponsored by the World Bank, USAID, the EU, and other actors have focused on three types of problem-solving strategies. First, technical aid to the riparian countries has focused on forecasting of seasonal runoff based on precipitation estimates for the upstream parts of the basin (e.g. Schär et al. 2004) as well as decision support and operational planning tools for reservoir management and water-energy exchanges among the riparian states[18]. Better predictions of water availability in the upstream catchment and of water demand by Uzbekistan and Kazakhstan well ahead of the growing season[19] could be helpful in designing fixed operating rules for the Toktogul reservoir and well-structured and transparent exchanges of water and energy. That is, these tools could lower the transaction costs that, under the international regime currently in place, inhibit multi-year planning and effective implementation of international commitments.

Second, tensions among the riparians could be alleviated if irrigation efficiency downstream and energy efficiency upstream were increased – this would reduce the inter-temporal divergence of up- and downstream interests in respect to water releases from the Toktogul reservoir. Return flows of 13.5 to 5.5 km^3 per year suggest that only around 40-50% of water withdrawals downstream (mainly for irrigation) are fully consumptive. This suggests a lot of room for improving the efficiency of water consumption through well-known irrigation technologies. This would clearly reduce net irrigation abstraction. It would thus allow for reduced water releases from the Toktogul reservoir in the growing season (which would save water for electricity production in winter) and could help in reducing the pollution problem associated with return flows. Alternatively, if at higher levels of irrigation efficiency downstream water releases in the growing-season were not reduced this would provide more water for the Aral Sea. As to energy efficiency upstream, the Kyrgyz energy system is highly inefficient, with losses of 40% and more. This is partly a technical problem, but partly also a problem of government

---

[18] http://www.usaid.gov/locations/europe_eurasia/car/briefers/transboundary_water.html

[19] One of the complicating factors is that, since the demise of the USSR, crop patterns in the downstream catchment are changing (mostly away from cotton and towards cereals and other crops).

mismanagement, corruption, and the general economic crisis in Kyrgyzstan (which is harder to deal with). Moreover, increasing energy efficiency will not automatically lead to less electricity production by Kyrgyzstan in winter and therefore less water releases from the Toktogul reservoir in the non-growing season (the preferred outcome from downstream countries' perspective). Kyrgyzstan may simply wish to export the energy surplus thus obtained in winter in order to generate foreign earnings. In other words, increasing irrigation and energy efficiency may create unintended, perverse incentives that need to be dealt with.

Third, structural changes to the current hydraulic system, notably construction of the socalled Kambarata 1 and 2 projects upstream of the Toktogul, could allow Kyrgyzstan to increase hydropower production while maintaining capacity in the Toktogul reservoir for water releases in the growing season. Such a solution would, therefore, be beneficial for up- and downstream countries. However, plans for the Kambarata project were developed already under Soviet rule, but have so far floundered because of great uncertainty over the financial viability of such a project.

Appendices

Appendix A – Derivation of Expected Value of PER_i^*

The expected value as denoted by Equation (16) is easily obtained in the following way:

$$\left\langle PER_i^* \right\rangle^+ = 1 - \frac{1}{\mu_{CP}^2}\left\langle \delta_{AP}^{Abs}\delta_{CP}^{Abs}\right\rangle + \frac{2}{\mu_{CP}}\left\langle \delta_{AP}^{Abs}\right\rangle =$$

$$1 + 2\frac{\mu_{AP}}{\mu_{CP}} - \frac{1}{\mu_{CP}^2}\left(\left\langle \delta_{AP}^{Abs}\right\rangle\left\langle \delta_{AP}^{Abs}\right\rangle + Cov(\delta_{AP}^{Abs},\delta_{CP}^{Abs})\right) = \tag{18}$$

$$1 + \frac{\mu_{AP}}{\mu_{CP}} - \frac{1}{\mu_{CP}^2}Cov(\delta_{AP}^{Abs},\delta_{CP}^{Abs})$$

Appendix B – Derivation of Variance of PER_i^*

According to standard textbook definition, we have

$$\sigma_{PER_i^*}^2 = \left\langle PER_i^{*2}\right\rangle - \left\langle PER_i^*\right\rangle^2 \tag{19}$$

In the case of similar signs of both, δ_{AP} and δ_{CP}, and by utilizing the result from Equation (16), we get

$$\sigma_{PER_i^*}^{2\ \ +} = \frac{4(\sigma_{\delta_{AP}^{Abs}} + 3\mu_{\delta_{AP}^{Abs}}^2)}{\mu_{\delta_{CP}}^2} + \frac{2\,\mathrm{Cov}(\delta_{AP}^{Abs},\delta_{CP}^{Abs})\mu_{\delta_{AP}^{Abs}}}{\mu_{\delta_{CP}^{Abs}}^3} - \frac{\mathrm{Cov}(\delta_{AP}^{Abs},\delta_{CP}^{Abs})^2}{\mu_{\delta_{CP}^{Abs}}^4} + \frac{\left\langle \delta_{AP}^{Abs2}\delta_{CP}^{Abs2}\right\rangle}{\mu_{\delta_{CP}^{Abs}}^4} - \frac{4\left\langle \delta_{AP}^{Abs2}\delta_{CP}^{Abs}\right\rangle}{\mu_{\delta_{CP}^{Abs}}^3} \tag{20}$$

Unfortunately, the variance of $PER_i^{'}$ cannot be determined without knowledge of the underlying probability distribution functions of AP, CP and OP since third and fourth order moments have to be determined (last two terms of Equation (20)). However, we can again linearize these terms. By doing so, after a somewhat tedious calculation, we obtain for the individual higher order terms

$$\frac{\left\langle \delta_{AP}^{Abs2}\delta_{CP}^{Abs2}\right\rangle}{\mu_{\delta_{CP}^{Abs}}^4} \approx \frac{\mu_{\delta_{AP}^{Abs}}\left(4\,\mathrm{Cov}\left(\mu_{\delta_{AP}^{Abs}},\mu_{\delta_{CP}^{Abs}}\right) + \mu_{\delta_{AP}^{Abs}}\mu_{\delta_{CP}^{Abs}}\right)}{\mu_{\delta_{CP}^{Abs}}^3} \tag{21}$$

and

$$\frac{4\left\langle \delta_{AP}^{Abs2}\delta_{CP}^{Abs}\right\rangle}{\mu^{3}_{\delta_{CP}^{Abs}}} \approx \frac{\mu_{\delta_{AP}^{Abs}}\left(2\,\mathrm{Cov}\left(\mu_{\delta_{AP}^{Abs}},\mu_{\delta_{CP}^{Abs}}\right)+\mu_{\delta_{AP}^{Abs}}\mu_{\delta_{CP}^{Abs}}\right)}{\mu^{3}_{\delta_{CP}^{Abs}}} \tag{22}$$

Plugging these results into Equation (20) and simplifying leads to Equation (17). Note that $\sigma^{2}_{PER_i^{*}}{}^{+} = \sigma^{2}_{PER_i^{*}}{}^{-}$.

Appendix C – Sample Values Estimation

Note that $\mu_{\delta_{AP}^{Abs}}$, $\mu_{\delta_{CP}^{Abs}}$, $\sigma^{2}_{\delta_{AP}^{Abs}}$, $\sigma^{2}_{\delta_{CP}^{Abs}}$ and $\mathrm{Cov}(\delta_{AP}^{Abs},\delta_{CP}^{Abs})$ have to be empirically estimated from available data. When the underlying probability distribution functions are not known but a set of observations $\{\{\delta_{AP}^{Abs}(1),\delta_{CP}(1)\},\{\delta_{AP}^{Abs}(2),\delta_{CP}(2)\},...,\{\delta_{AP}^{Abs}(n),\delta_{CP}(n)\}\}$ is available in time, the moments of the distributions of δ_{AP}^{Abs} and δ_{CP}^{Abs} can be estimated by the estimated sample values

$$\hat{\mu}_{\delta_{AP}^{Abs}} = \frac{1}{n}\sum_{t=1}^{n}\delta_{AP}^{Abs}(t) = \frac{1}{n}\sum_{t=1}^{n}\left|\delta_{AP}(t)\right| \tag{23}$$

and

$$\hat{\mu}_{\delta_{CP}^{Abs}} = \frac{1}{n}\sum_{t=1}^{n}\delta_{CP}^{Abs}(t) = \frac{1}{n}\sum_{t=1}^{n}\left|\delta_{CP}(t)\right| \tag{24}$$

respectively for the mean. The estimation of the sample variances is carried out in the following way

$$\hat{\sigma}^{2}_{\delta_{AP}^{Abs}} = \frac{1}{n}\sum_{t=1}^{n}\left(\delta_{AP}^{Abs}(t)-\hat{\mu}_{\delta_{AP}^{Abs}}\right)^{2} = \frac{1}{n}\sum_{t=1}^{n}\left(\left|\delta_{AP}(t)\right|-\hat{\mu}_{\delta_{AP}^{Abs}}\right)^{2} \tag{25}$$

for the variance of either, δ_{AP}^{Abs} and δ_{CP}^{Abs}. In Equations (23) to (25), *n* denotes the number of observations at hand.

As already stated, temporal random variables are functions whose values change with time and are observed as a particular time series of a stochastic process. In other words, observations can be positively correlated. This would increase the sample estimate of the variance as given by Equation (25) which, in fact, is an approximation and not taking into account this autocorrelation[20].

---

[20] The sample estimate of $\hat{\sigma}^{2}_{\delta_{AP,CP}}$ in case of δ_{AP} and δ_{CP} being observations resulting from Markov processes and taking into account autocorrelation is given in (Loucks, Stedinger et al. 1981).

Similarly, $\mathrm{Cov}(\delta_{AP},\delta_{CP})$ can be estimated by

$$\hat{\gamma}\left(\delta_{AP}^{Abs},\delta_{CP}^{Abs}\right)=\frac{1}{n}\sum_{t=1}^{n}\left(\delta_{AP}^{Abs}(t)-\hat{\mu}_{\delta_{AP}}\right)\left(\delta_{CP}(t)-\hat{\mu}_{\delta_{CP}}\right) \tag{26}$$

as shown in (Loucks, Stedinger et al. 1981). If we plug in the sample estimates of the mean $\hat{\mu}_{\delta_{AP,CP}}$, the variance $\hat{\sigma}^2_{\delta_{AP,CP}}$ and the covariance $\hat{\gamma}\left(\delta_{AP}^{Abs},\delta_{CP}\right)$ into Equations (16) and Equation (17), we obtain an estimated mean and variance of PER_i^* over the period of assessment.

Appendix D – Toktogul Data Sets

	natural runoff regime		Period 1		Period 2		Period 3		Optim.
month	**µ**	**ó**							
1	150.0	27.9	188.6	74.7	478.5	101.1	590.0	55.3	357.7
2	151.1	25.5	202.1	67.2	464.2	113.0	561.8	78.6	426.2
3	178.2	28.5	195.5	50.4	428.9	122.1	465.8	52.9	323.4
4	314.7	94.3	271.9	94.2	350.2	115.6	367.0	79.6	426.2
5	661.4	200.8	457.8	186.4	348.0	120.2	286.8	52.0	452.8
6	969.3	342.2	550.9	196.3	450.1	152.6	270.6	73.8	468.0
7	797.6	264.1	654.8	205.5	481.0	174.5	324.3	78.2	494.7
8	516.9	137.0	521.7	153.9	354.1	79.5	316.6	40.3	490.9
9	287.1	71.7	184.1	99.8	198.5	89.2	228.1	93.0	441.4
10	230.4	48.8	142.6	73.0	234.5	67.7	313.7	86.8	300.6
11	217.0	45.7	144.3	92.1	343.5	51.9	439.4	84.9	304.4
12	174.1	30.1	188.5	79.2	479.7	82.3	590.6	53.0	418.6
Overall	388	307	311	215	384	139	396	141	409

Table 3: Mean and standard deviation of monthly flows given management regime. The bottom row displays overall mean and standard deviation for the duration of the regime periods. Units are m^3/s for μ and σ. The last column shows data from (Cai, McKinney et al. 2003).

Month	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
1	293.7	368.0	451.0	385.3	407.0	281.0	286.0
2	233.7	344.0	382.7	424.7	323.7	266.0	159.3
3	172.0	295.3	194.7	229.3	214.7	274.3	253.3
4	30.0	221.7	14.0	25.3	9.3	228.0	153.7
5	400.3	90.7	337.0	12.7	276.3	140.3	81.7
6	544.0	329.7	207.0	126.7	542.0	633.0	28.3
7	577.7	242.3	377.7	188.3	688.0	841.3	309.3
8	313.0	272.7	254.3	58.3	359.0	472.7	334.0
9	82.0	23.0	52.3	5.0	372.3	90.0	122.7
10	204.0	143.7	237.7	62.0	67.3	220.0	103.7
11	256.0	334.3	278.3	159.7	115.7	303.0	201.3
12	253.7	400.3	351.7	425.3	361.7	372.0	304.3

Table 4: $\delta_{AP}^{Abs}\left(S\right)$ for the 7 year management period under investigation. Units are m^3/s.

Month	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
1	163.7	145.0	229.0	342.7	386.3	229.0	275.0
2	73.0	158.7	145.0	358.3	362.0	213.7	238.7
3	2.7	178.0	119.0	386.7	340.3	187.3	176.7
4	53.3	90.7	29.7	348.0	26.3	68.7	24.7
5	131.3	71.3	306.7	229.0	330.3	220.3	18.0
6	30.7	102.3	240.7	210.7	414.7	308.0	24.3
7	23.7	194.0	533.7	233.3	328.7	643.0	186.0
8	85.0	149.7	338.3	63.3	364.0	459.3	368.3
9	89.7	60.0	111.0	195.0	371.7	82.3	126.0
10	150.0	146.3	158.7	129.7	58.0	86.3	176.7
11	196.0	200.7	229.7	205.7	17.3	169.7	32.7
12	94.3	260.7	287.0	374.0	258.0	307.0	138.0

Table 5: $\delta_{CP}^{Abs}(S)$ for the 7 year management period under investigation. Units are m^3/s.

Month	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
1	258.2	316.5	380.5	275.8	373.5	366.5	312.5
2	165.8	224.2	336.8	317.8	245.8	291.2	146.5
3	177.7	237.0	228.4	183.0	212.4	343.0	271.0
4	60.8	125.8	48.2	16.5	56.5	193.2	91.8
5	74.0	47.3	10.4	123.3	63.0	4.0	22.0
6	112.4	78.0	26.6	33.7	162.4	101.4	60.0
7	91.8	24.5	77.2	65.5	121.2	110.5	3.5
8	89.5	40.2	11.2	20.2	75.5	1.2	22.5
9	131.6	155.6	139.2	125.6	190.9	35.4	43.1
10	41.8	41.5	137.5	25.8	12.5	211.8	177.5
11	113.1	243.5	202.5	118.8	181.8	340.1	297.5
12	208.2	255.9	237.2	324.9	344.2	328.5	340.9

Table 6: $\delta_{AP}^{Abs}(C)$ for the 7 year management period under investigation. Units are m^3/s.

Month	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
1	128.2	93.5	158.5	233.2	352.8	314.5	301.5
2	5.2	38.8	99.2	251.5	284.2	238.8	225.8
3	3.0	119.7	152.7	340.4	338.0	256.0	194.4
4	22.5	5.2	4.5	306.2	20.8	33.8	37.2
5	195.0	114.7	40.7	118.4	117.0	76.0	77.6
6	401.0	149.3	7.0	50.3	35.0	223.6	56.0
7	462.2	72.8	78.8	20.5	238.2	87.8	119.8
8	138.5	82.8	72.8	15.2	80.5	12.2	11.8
9	123.9	118.6	80.6	64.4	190.2	136.9	205.5
10	12.2	44.1	58.5	93.5	3.2	78.1	102.9
11	53.1	109.8	153.8	164.8	83.5	206.8	128.8
12	48.9	116.2	172.5	273.5	240.5	263.5	174.5

Table 7: $\delta_{CP}^{Abs}(C)$ for the 7 year management period under investigation. Units are m^3/s.

Bibliography

Bennear, L. S. and C. Coglianese (2005). "Measuring progress: Program evaluation of environmental policies." Environment **47**(2): 22-39.

Bernauer, T. (1995). "The Effect of International Environmental Institutions - How We Might Learn More." International Organization **49**(2): 351-377.

Cai, X. M., D. McKinney, et al. (2003). "Integrated Hydrologic-Agronomic-Economic Model for River Basin Management." Journal of Water Resources Planning and Management **129**(1): 4-17.

Cai, X. M., D. C. McKinney, et al. (2002). "A framework for sustainability analysis in water resources management and application to the Syr Darya Basin." Water Resources Research **38**(6): -.

Dukhovny, V. and V. Sokolov Integrated Water Resources Management in the Aral Sea Basin, Inter-State Commission for Water Coordination in the Aral Sea Basin**:** 27.

Helm, C. and D. Sprinz (2000). "Measuring the Effectiveness of International Environmental Regimes." Journal of Conflict Resolution **45**(5): 630-652.

Hovi, J., D. Sprinz, et al. (2003). "Regime Effectiveness and the Oslo-Potsdam Solution: A Rejoinder to Oran Young." Global Environmental Politics **3**(3): 105-107.

Hovi, J., D. F. Sprinz, et al. (2003). "The Oslo-Potsdam Solution to Measuring Regime Effectiveness: Critique, Response, and the Road Ahead." Global Environmental Politics **3**(3): 74-96.

Kjeldsen, T. R. and D. Rosbjerg (2004). "Choice of reliability, resilience and vulnerability estimators for risk assessments of water resources systems." Hydrological Sciences Journal-Journal Des Sciences Hydrologiques **49**(5): 755-767.

Loucks, D. P., J. R. Stedinger, et al. (1981). Water Resource Systems Planning and Analysis. Englewood Cliffs, New Jersey, Prentice-Hall, Inc.

McKinney, D., X. M. Cai, et al. (1999). Integrated Water Resources Management Model for the Syr Darya Basin. E. P. a. I. f. C. A. E. Program. Almaty, Kazhakhstan, US Agency for International Development

Rieckermann, J., H. Daebel, et al. (2006). "Assessing the Performance of International Water Management at Lake Titicaca." accepted for publication in Aquatic Sciences.

Savoskul, O., E. Chevnina, et al. (2003). Water, Climate, Food, and Environment in the Syr Darya Basin. Adaptation strategies to changing environments**:** 62.

Schär, C. et al. (2004). Seasonal Runoff Forecasting Using Precipitation from Meteorological Data Assimilation Systems. Journal of Hydrometeorology 5: 959973

Sprinz, D. (2005). Regime Effectiveness: The Next Wave of Research. Conference on the Human Dimensions of Global Environmental Change, Berlin.

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

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

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

Young, O. (2001). "Inferences and Indicies: Evaluating the Effectiveness of International Environmental Regimes." Global Environmental Politics **1**(1): 99-121.

Young, O. (2003). "Determining Regime Effctiveness: A commentary on the Oslo-Potsdam Solution." Global Environmental Politics **3**(3): 97-104.